

RESEARCH

Open Access



Generalization of intensity distribution of medical images using GANs

Dong-Ho Lee, Yan Li and Byeong-Seok Shin*

*Correspondence:
bsshin@inha.ac.kr
Dept of Computer
Engineering, INHA University,
Incheon, South Korea

Abstract

The performance of a CNN based medical-image classification network depends on the intensities of the trained images. Therefore, it is necessary to generalize medical images of various intensities against degradation of performance. For lesion classification, features of generalized images should be carefully maintained. To maintain the performance of the medical image classification network and minimize the loss of features, we propose a method using a generative adversarial network (GAN) as a generator to adapt the arbitrary intensity distribution to the specific intensity distribution of the training set. We also select CycleGAN and UNIT to train unpaired medical image data sets. The following was done to evaluate each method's performance: the similarities between the generalized image and the original were measured via the structural similarity index (SSIM) and histogram, and the original domain data set was passed to a classifier that trained only the original domain images for accuracy comparisons. The results show that the performance evaluation of the generalized images is better than that of the originals, confirming that our proposed method is a simple but powerful solution to the performance degradation of a classification network.

Keywords: Generative adversarial network, Intensity distribution, Medical image, Machine learning

Introduction

Computer-aided diagnosis (CAD) based on deep learning has already been studied extensively [1]. In particular, many successful studies have applied convolutional neural networks (CNNs) [2, 3] to medical image processing. Studies of the classification of pathology [4–6], lesion segmentation [7–9] and body detection [10–13] using CNNs have been carried out with good performance. However, CNNs learn the intensity of the images. If test an image with a completely different intensity from the learned image, the performance of the CNN greatly degrades. This problem is even more prominent in the medical imaging domain. Unlike the domain of real-world images, medical images are grayscale, and features of lesions in the images can be very detailed and complex. Unfortunately, medical images show different intensity distributions depending on the characteristics of the imaging equipment and the operating methods of the radiologist, and it is very difficult to extract the features of lesions that are valid for all intensity. Therefore, CNNs for medical imaging suffer from poor

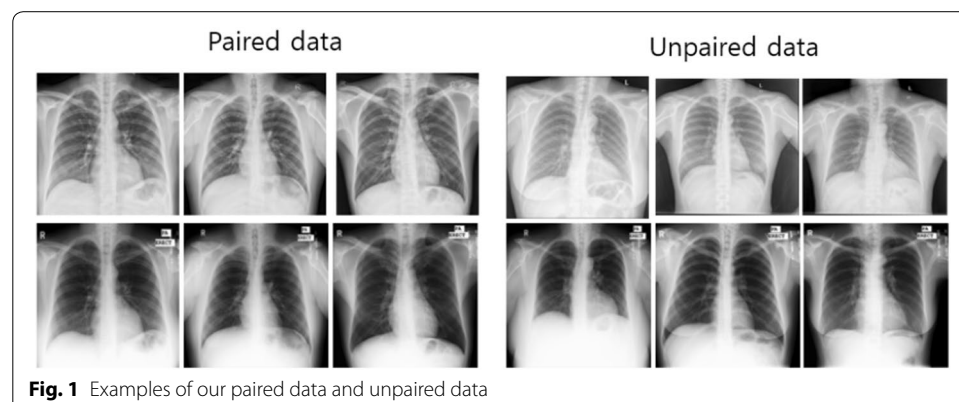
performance for new medical images with intensity distributions that are completely different from training data sets. This unstable performance change makes it impossible to commercialize that CNNs for medical domain, because it is impossible to obtain a data set that considers all conditions in the image-shooting environment. Also, because of the infinite number of new data sets with a variety of intensities, training new ones to the network each time is a very expensive task. Therefore, in order to solve this problem practically, generalization of new data sets needs to be considered.

Traditionally, histogram processing, such as histogram equalization and histogram matching, was used to adjust the similarity of intensity distribution. However, it is very difficult to adjust the intensity distribution of all input images to the distribution of a training data set with these methods.

The task of transforming image data from an arbitrary domain into a target domain is known as image-to-image translation. This is a kind of domain adaptation. Image-to-image translation has been actively researched using generative adversarial networks (GANs) [14–16] and variational auto-encoders (VAEs) [17, 18].

The Pix2Pix network [19] performs image-to-image translation using a paired data set. For each image in the original domain, the paired data set contains an image converted to the target domain. It is not easy to get a paired data set like this, but CycleGAN [20] and UNIT [21] solved this limitation and proposed that a GAN can learn with an unpaired data set. In the medical imaging domain, it is practically impossible to obtain a true paired data set. Therefore, much research has been done through GANs that can be trained with unpaired data sets. Figure 1 shows example of paired and unpaired image dataset. The paired data shows what multiple chest X-rays of a single person taken on several machines might look like. In reality, this data is virtually unattainable, and the paired data above shows fake images that we created. The unpaired data shows X-rays of people taken on different machines. This is usually the data set we are dealing with. The intensities of the two sets are very different.

GANs have been applied to medical imaging in earnest since 2017 [22]. In particular, many studies show data augmentation using GANs for image synthesis [23, 24], and most of them were conducted using magnetic resonance (MR) and computed tomography (CT) data. Data augmentation is useful for training the network, but it is not a good method for maintaining existing network performance; MR and CT



can obtain high-quality images but are less accessible than X-rays. Therefore, image synthesis needs to be considered to maintain a network's performance of the X-ray target.

In this paper, we propose a generalization framework that adjusts a set of X-ray images with arbitrary intensity distribution to match the intensity distribution of a training set, using CycleGAN and UNIT as generalizers to maintain the accuracy of a medical-image classification network (Fig. 2).

Contributions

This paper presents two contributions:

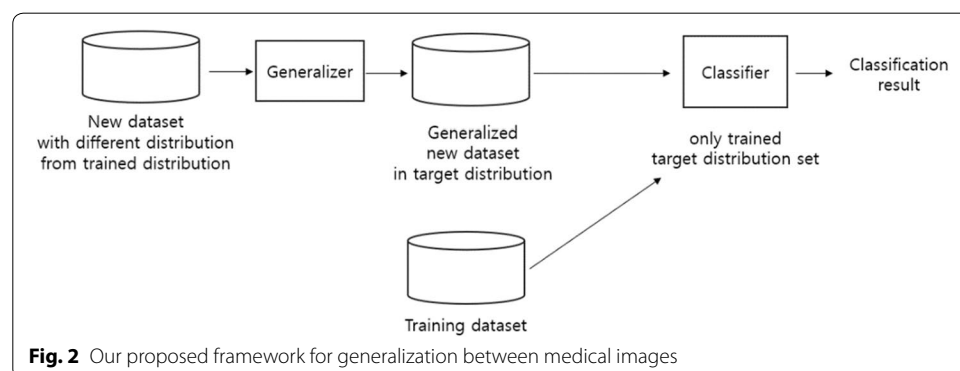
1. A solution of performance degradation for lesion classification and other tasks via the generalization of medical-image intensity distribution.
2. Data augmentation using intensity generalization in the medical image domain suffering from a lack of data.

Paper organization

The rest of the paper is organized as follows: In “[Related works](#)” section, we introduce traditional intensity generalization using histogram processing and recent research in image-to-image translation tasks using GANs, finally providing some GAN applications for the medical image domain. In “[Methods](#)” section, we detail the architecture of our proposed generalization network that adjusts the intensity of new test images to those of the original training data set. We also provide brief details of CycleGAN and UNIT as generalizers in our network. Performance comparisons of the proposed networks are in “[Experiments](#)” section. Finally, in “[Conclusion](#)” section, we present our conclusions.

Related works

To adjust the intensity of arbitrary image data sets to the intensity of a specific image data set is a difficult task. We introduce the traditional approach and new approach with generative adversarial networks to solve this problem in this section.



Traditional method

Traditionally, histogram matching was used to solve this problem. Histogram matching (or histogram specification) is the method in which an input histogram matches a target histogram using their cumulative distribution function (CDF). The cumulative histogram is calculated from each image; any value to be matched to another histogram, x_i , has a cumulative histogram value given by $G(x_i)$. This is the cumulative distribution value in the target image, namely $H(x_i)$. The input data value, x_i , is replaced by x_j , where $G(x_i)$ is equal to $H(x_j)$. However, this method cannot handle matching histograms between two different data sets and simply uses CDF matching of both histograms. Therefore, it cannot be used for challenges such as finding the probability distribution or the manifold in the training data set.

Generative adversarial networks

A GAN is a generative model that estimates the probability distribution of training data sets, $p(x)$, and generates new data, $G(x)$, similar to that distribution. This allows the GAN to find the manifold in a specific domain. Sampling in a well-approximated manifold space yields results that are similar to the original but have different details. A GAN consists of two neural networks: the generator and the discriminator. The generator learns to generate images that can fool the discriminator. The discriminator learns whether an input image is an original image or a fake image from the generator. That is, the generator and discriminator have different goals. The discriminator needs to maximize $\log(D(x))$, while the generator needs to minimize $\log(1-D(G(z)))$ where z is a random vector. Therefore, this network can be considered adversarial. Vanilla GAN loss is called adversarial loss [14], defined in Eq. 1 as follows:

$$\min_G \max_D \mathcal{L}_{GAN} = \mathbb{E}_{x \sim p(x)} [\log D(x)] + \mathbb{E}_{z \sim p(z)} [\log(1 - D(G(z)))] \quad (1)$$

The vanilla GAN generates z with random noise. Because of this, the generator at the beginning of the training always produces a completely fake image. This allows the discriminator to completely distinguish whether the input is fake or not. That is, $D(G(z))$ becomes 1, and a learning generator is impossible. Therefore, we will have maximized $\log(D(G(z)))$ instead of minimizing $\log(1 - \log D(G(z)))$.

To solve the problem of a z vector with random noise, a conditional GAN (cGAN) [25] is proposed. We can use the conditional input vector, c , to add to the random noise of z using concatenation for a better output image. In the cGAN, a new vector combining z and c becomes the input for the generator and the loss function, providing Eq. 2 below, where y is a given label vector.

$$\min_G \max_D \mathcal{L}_{cGAN} = \mathbb{E}_{x \sim p(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p(z)} [\log(1 - D(G(z|y)))] \quad (2)$$

The c vector does not have a specific type. For example, image labels can be used as the c vector [26].

There are some popular methods using cGAN in image-to-image translation. Pix2Pix proposes a modified loss function for cGAN in Eq. 3 combined with L1 regularization in Eq. 4 for denoising the generated result. Also, their L1 loss contains self-similarity

between $G(x, z)$ and label y . However, in the actual implementation of Eq. 3, we don't need random noise because our input image is sufficiently complex.

$$\mathcal{L}_{cGAN-pix2pix} = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log(1 - D(x, G(x, z)))] \quad (3)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1] \quad (4)$$

This network shows good performance but has the limitation that it is only trained by a paired data set. To solve the unpaired data set problem, CycleGAN improved Pix2Pix is proposed. They use an idea called cycle-consistency loss (or reconstructed loss) that performs bidirectional conversion between the source domain and the target domain.

In addition to the cGAN, there are many GAN variants. There are networks that use auxiliary classifiers or VAE. In addition to image-to-image translations using cGAN, UNITs that use VAE and weight sharing have been widely used.

GANs for computer-aided diagnosis

Various GAN methods have already been applied to CAD, especially in the synthesis, segmentation, reconstruction (such as enhancement or denoising), and classification fields. Most studies have focused on synthesis and segmentation [22] so that these are well suited for image-to-image translation.

For the segmentation researches, Li et al. [27] proposed cGAN combining with Pix2Pix and ACGAN [28] for MR segmentation. Dai et al. [29] proposed SCAN network, which shows that adversarial loss can be applied to organ segmentation in X-rays.

For the synthesis researches, the performance of converting between MR and CT is outstanding. Emami et al. [30] synthesized brain MRIs from CT using cGAN with a paired data set. On the other hand, Wolterink et al. [24] synthesized MRI images into CT images using CycleGAN with an unpaired data set. Dar et al. [31, 32] studied the transformation between T1- and T2-weighted MR images using CycleGAN. Mahmood et al. [33] applied adversarial training methods to depth-estimation from monocular endoscopy.

Although there have not been many studies, some studies have used GAN for classification. Madani et al. [34] shows that DCGAN [16] can be used for classification. They used a discriminator as a classifier and conducted data augmentation using generated images in the training process.

The research so far has the following unsatisfactory points:

1. Most studies focus on MR and CT images. There are few studies on X-ray images. X-ray images are readily available in many areas regardless of the medical infrastructure, so application to X-rays is meaningful, too.
2. There is no research on maintaining the performance of the classifier to make it robust regardless of the intensity difference. Research has only focused on data augmentation.
3. Most of the research has utilized CycleGAN and Pix2Pix for their tasks and compare their performance. This trend is evident regardless of the application. However, there is no performance comparison between UNIT based VAE and CycleGAN.

Therefore, we propose generalizing medical image intensity to maintain the performance of a network using GAN, comparing the performance of CycleGAN and UNIT.

Methods

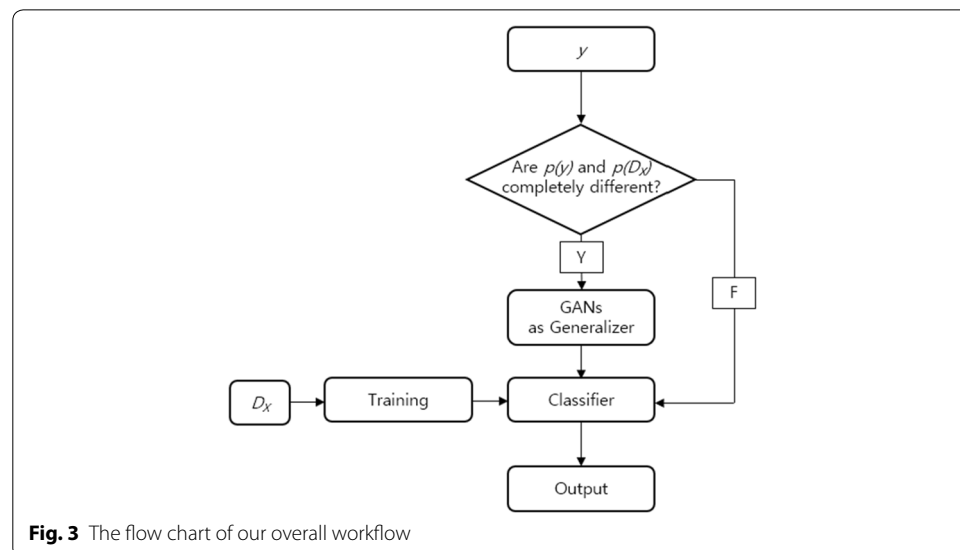
We propose a generalization method for a new data set with a different intensity distribution from the training data set to maintain the performance of good existing classification networks (Fig. 3).

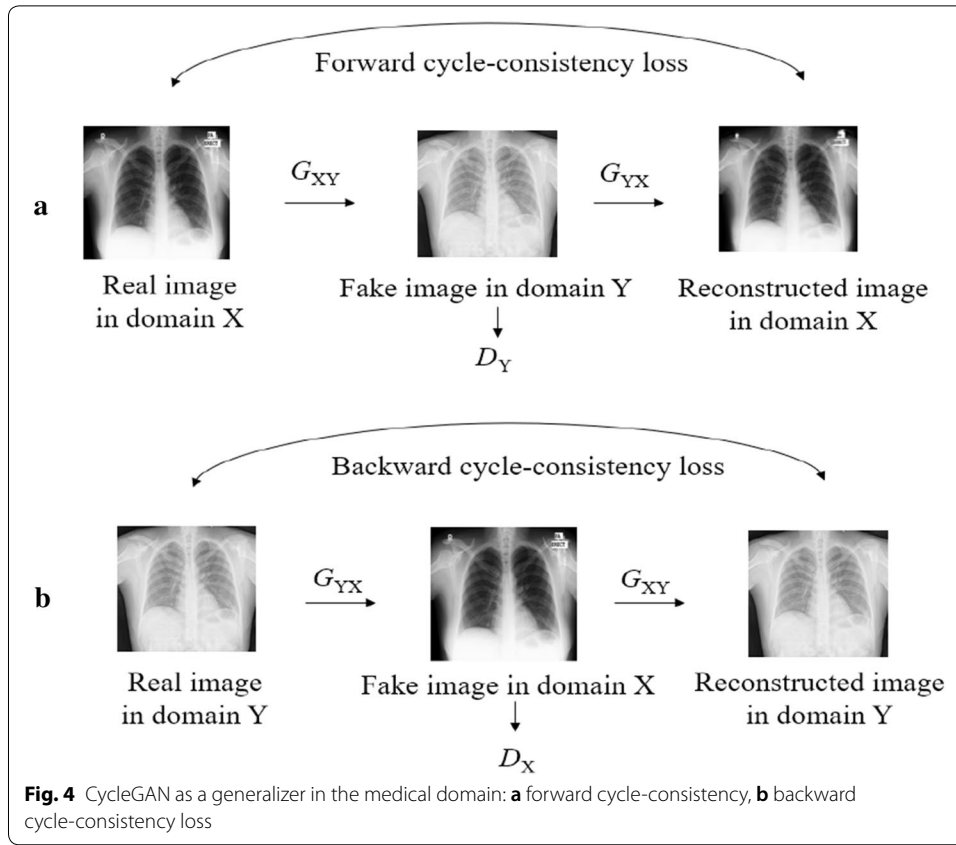
However, it is impossible to collect such a paired data set in a medical image domain. For example, it requires shooting one person on several machines at the same time to obtain a varied intensity distribution data set for one medical image. This is an impossible task and, indeed, unnecessary. Therefore, our generalizer is chosen a GAN that should be trained with an unpaired data set.

Our chosen CycleGAN and UNIT are popular image-to-image translation GANs that can be trained with unpaired data sets and work well in various domains as well as medical imaging. We introduce two networks below.

Generalizer using CycleGAN

The CycleGAN is the widely used network for style transfer tasks. The results of this network tend to maintain features of the original domain, such as the shape of instance, as much as possible. Figure 4 shows the structure of a CycleGAN used to generalize the intensity distribution of medical images. The key idea is cycle consistency, that is the loss between the original domain image and the reconstructed image for the training of an unpaired data set. The reconstructed image is retransformed from the fake image to the original domain image. First, the generator G_{XY} generates a domain-transformed image, $G_{XY}(X)$, and then obtains a reconstructed image, $G_{YX}(G_{XY}(X))$, through G_{YX} that reconstructs the transformed image, $G_{XY}(X)$, into the original domain. By reducing the loss between the original domain image





and the reconstructed image, the network maintains the characteristics of the original domain. Conversely, the same method applies to converting from a target domain to the original domain.

In this case, CycleGAN uses forward–backward cycle-consistency losses. Forward cycle-consistency loss is the loss in converting from domain X to domain Y and then retransforming back to the original domain (Fig. 4a). Backward cycle-consistency loss, on the other hand, is the loss when converting from the target domain to the original domain and then back to the target domain (Fig. 4b). This cycle-consistency loss is similar to L1 loss containing self-similarity (Eq. 4), and it summarized in Eq. 5. The final objective function is constructed by adding it to the existing GAN loss.

$$\begin{aligned} \mathcal{L}_{cyc}(G_{XY}, G_{YX}) = & \mathbb{E}_{x \sim p(x)} [\| G_{YX}(G_{XY}(x)) - x \|_1] \\ & + \mathbb{E}_{y \sim p(y)} [\| G_{XY}(G_{YX}(y)) - y \|_1] \end{aligned} \quad (5)$$

For model stability and to avoid mode collapse, CycleGAN uses a least-square loss function [35] instead of vanilla GAN loss (Eq. 1) in Eq. 6.

$$\begin{aligned} \mathcal{L}_{LSGAN-cycGAN}(G_{XY}, G_{YX}) = & \mathbb{E}_{y \sim p(y)} \left[(D_Y(y) - 1)^2 \right] \\ & + \mathbb{E}_{x \sim p(x)} \left[(D_Y(G_{XY}(x)))^2 \right] \end{aligned} \quad (6)$$

Then their final loss function is in Eq. 7.

$$\begin{aligned} \mathcal{L}_{\text{CycleGAN}}(G_{XY}, G_{YX}, D_X, D_Y) = & \mathcal{L}_{\text{LSGAN-cycleGAN}}(G_{XY}, D_Y) \\ & + \mathcal{L}_{\text{LSGAN-cycleGAN}}(G_{YX}, D_X) + \lambda \mathcal{L}_{\text{cyc}}(G_{XY}, G_{YX}) \end{aligned} \quad (7)$$

Generalizer using UNIT

The UNIT consists of two structures combined with VAE and GANs as shown Fig. 5, and their core idea is a shared latent vector for inferring joint distribution between different domain data sets. The aim of the shared latent vector is to limit the space of joint distribution. A single generator is conceptually divided into an encoder, G_e , and a decoder, G_d . The two encoders, G_{eA} and G_{eB} , respectively, generate latent vectors, $z \sim q(z|x)$ through weight sharing. q is treated by random vector of $N(E(x), I)$ where I is an identity matrix. These vectors are reshared in the decoders G_{dA} and G_{dB} , and the decoders also share weights. Therefore, UNIT is quite complex, using VAE and GAN losses together.

Suppose we have different domain data, A and B, for the same x . The entire loss function is shown in Eq. (8). UNIT also learns in both directions by applying cycle-consistency, but we represent the loss of only one direction. The other loss is easily obtained by changing the domain.

$$\mathcal{L}_{\text{UNIT}_A}(G_{eA}, G_{dA}, D_A) = \mathcal{L}_{\text{VAE}_A}(G_{eA}, G_{dA}) + \mathcal{L}_{\text{GAN}_A}(G_{eB}, G_{dA}, D_A) + \mathcal{L}_{\text{cc}_A}(G_{eA}, G_{dA}, G_{eB}, G_{dB}) \quad (8)$$

Two VAE trained by minimize a variational upper bound [19], in Eq. 9.

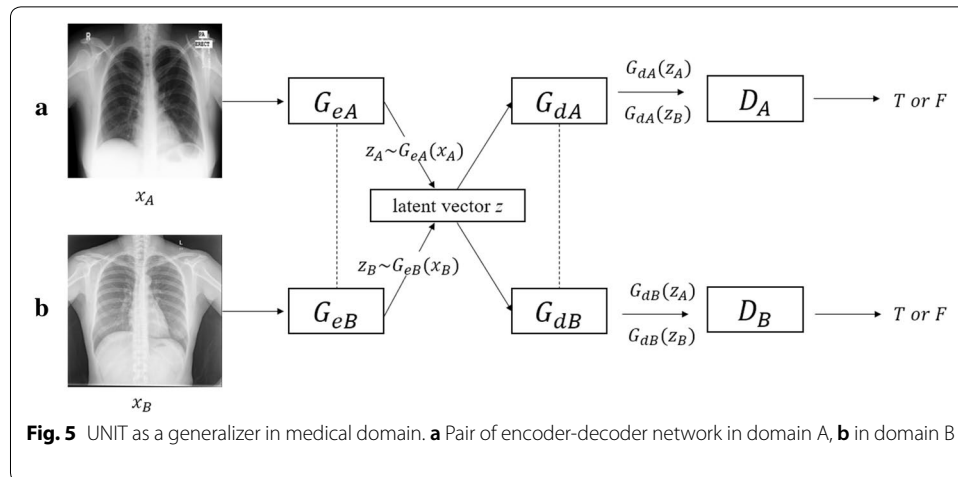
$$\mathcal{L}_{\text{VAE}_A}(G_{eA}, G_{dA}) = \lambda_1 KL(q_A(z_A|x_A) \parallel p_\eta(z)) + \lambda_2 \mathbb{E}_{z_A \sim q_A(z_A|x_A)} [\log(p_{G_{eA}}(x_A|z_A))] \quad (9)$$

GAN loss is based on cGAN loss as follows:

$$\mathcal{L}_{\text{GAN}_A}(G_{eA}, G_{dA}) = \lambda_0 \mathbb{E}_{x_A \sim P_{x_A}} [\log D_A(x_A)] + \lambda_0 \mathbb{E}_{z_B \sim q_B(z_B|x_A)} [\log(1 - D_A(G_A(z_B)))] \quad (10)$$

To model cycle-consistency condition, the loss is given by Eq. 11.

$$\begin{aligned} \mathcal{L}_{\text{cc}_A}(G_{eA}, G_{dA}, G_{eB}, G_{dB}) = & \lambda_3 KL(q_A(z_A|x_A) \parallel p_\eta(z)) + \lambda_3 KL(q_B(z_B|G_{dB}(z_A)) \parallel p_\eta(z)) \\ & - \lambda_4 \mathbb{E}_{z_B \sim q_B(z_B|G_{dB}(z_A))} [\log p_{G_{eA}}(x_A|z_B)] \end{aligned} \quad (11)$$



We use CycleGAN and UNIT as generalizers. Both networks have very different structures depending on whether VAE is used or not. In the next section, we will compare their performance as generalizers for both networks.

Experiments

This section presents the experimental methods and evaluation result. We propose three conditions for evaluation of our generalization framework and four methods to solve these conditions. Moreover, we compare the performance of three generalizers including UNIT, GAN and histogram matching. Also, we describe our datasets in this section.

Experiment data

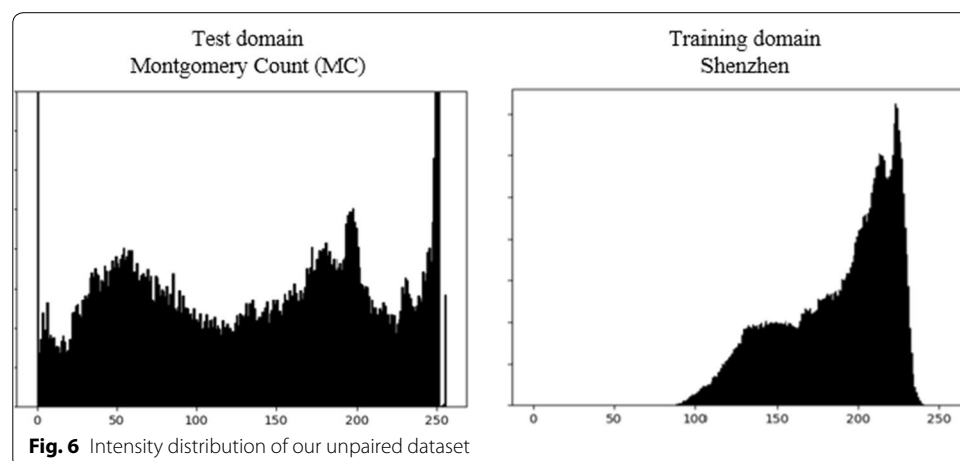
Figure 6 shows an example of the unpaired medical image data set used in the experiments of this paper. Our data set includes frontal chest X-ray image data, labeled tuberculosis (TB) or non-TB from the National Library of Medicine and National Institutes of Health (Bethesda, MD, USA) [36–38].

There were two data sets: the Shenzhen data set from Shenzhen No. 3, People's Hospital, captured with Philips DR Digital Diagnose systems, and the Montgomery County (MC) data set from the Department of Health and Human Services of Montgomery County (MD, USA), captured with a Eureka stationary X-ray machine.

The Shenzhen data set consisted of 336 cases with TB and 326 non-TB cases. The MC data set consisted of 58 cases with TB and 80 non-TB cases. Because both data sets were captured using different machines, the intensity distribution of the two data sets is completely different, so the network cannot find lesions in other test data sets, even though its original classification performance is high.

Experimental process

The experiment was divided into a generalization step and a classification step. Our test scenario assumes a situation where a new MC data set comes to the classifier that learned the Shenzhen data set. In other words, the MC data set was used as a test set for generalization performance, and the Shenzhen data set was used as a training set to train the classifier. Our classifier is based pretrained AlexNet [2] with 0.95 ± 0.02 area under



curve (AUC). The MC data sets are generalized by two generalizers, which are CycleGAN and UNIT, respectively. In the generalization step, the intensity distribution of the MC data set was adjusted to the Shenzhen data set using each generalizer.

Unlike with simple translation, the generalization of medical images requires the preservation of highly advanced features. In particular, detailed feature retention is necessary to distinguish the presence or absence of lesions.

For evaluation, we have to consider as three conditions:

1. From the MC data set M , $p_M(m_i)$ and the generalized result, $G(m_i)$, should have different distributions, where $m \sim p(M)$.
2. From the Shenzhen data set S , $p(S)$ and $G(M)$ should have similar distributions.
3. The conversion from m_i to $G(m_i)$ should minimize the loss of meaningful features.

Conditions (1) and (2) can be judged visually, but it is very difficult to confirm in condition (3). We propose the following methods to solve this problem:

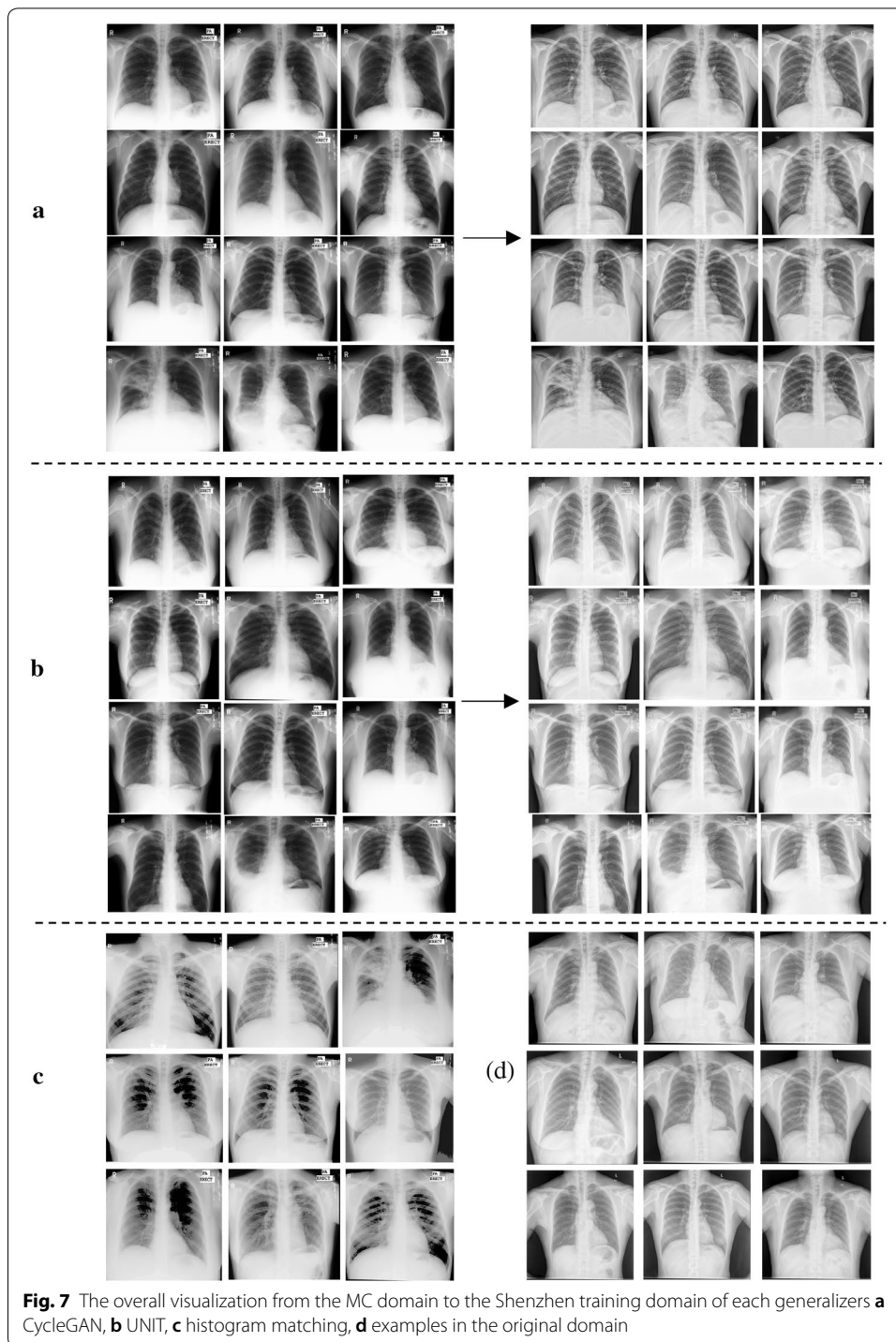
1. Show visualizations using the two generalizers.
2. Use the histogram as a simple measure for conditions (1) and (2) to identify the difference between the original image and the transformed image.
3. Use the structural similarity index (SSIM) for condition (3). $G(m_i)$ should have a structural similarity with m . Otherwise, the image may become completely different and the original lesion may be lost.
4. Compare the accuracy of the classification test to area under curve (AUC) and receiver operating characteristic (ROC) curve to assist the SSIM. Test the actual performance by testing generalized $G(M)$ on the classifier that only learned the real S .

Experimental result

Figure 7 shows the results of generalization using CycleGAN (Fig. 7a), UNIT (Fig. 7b), and histogram matching (Fig. 7c). Figure 8 shows the histogram comparison of each method. Table 1 also shows the mean SSIM score of each image and their standard deviation (std).

The difference in the results can be seen with the naked eye. The generalized image using CycleGAN is very similar to the training set, which is the target domain, and the intensity of the generalized result (red) is similar to the training domain (blue) in Fig. 8a. Compared with the distribution of the MC domain, the results are also the most distant (Fig. 8b). CycleGAN also showed good performance in the SSIM results (Table 1) as well as in visual information and histogram results. The SSIM is 0.737, a higher score than other methods. Also, the std is small, meaning CycleGAN is a stable method for the generalization task. It can be confirmed that all the features of the lung are maintained as the intensity transformation is properly performed.

Generalization with UNIT, on the other hand, has bad results. As a result of the visualization, the biggest problem was that the blurring was very severe (Fig. 7b). Therefore, UNIT did not show intensities similar to the target domain in the histogram results (yellow) in Fig. 8a, it is the furthest distribution from the target domain. This can be seen



immediately in the SSIM results. The SSIM of UNIT was 0.691, which was lower than histogram matching. However, UNIT performed better than histogram matching. As can be seen in Fig. 9, histogram matching was completely blacked out, and the features had completely disappeared. This shows that UNIT did not preserve the structural characteristics of the original domain.

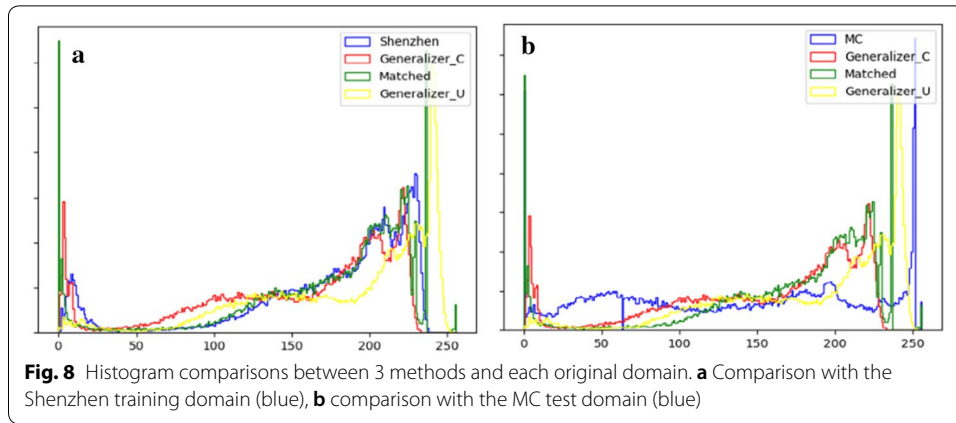
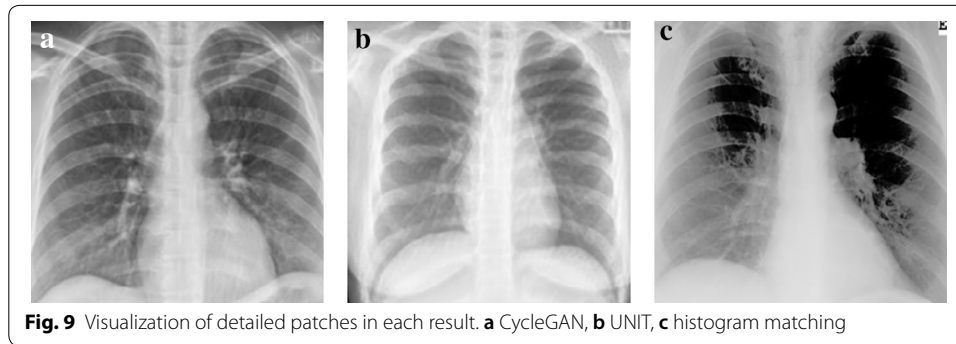


Table 1 The SSIM between three methods and the original data set

Methods	SSIM (%)
Generalizer—CycleGAN	73.7 ± 5.92
Generalizer—UNIT	69.1 ± 4.84
Histogram matched dataset	73.2 ± 8.23

We calculate the SSIM for each image, average that, and get the standard deviation (std). We denote the result as mean ± std. The CycleGAN generalizer shows 0.737 of the SSIM (given in italic), higher than the SSIM of other methods



Through the above experiments of visualization, we have found that CycleGAN preserves the features of the original image best. This can be analyzed by the effect of the additional identity loss (Eq. 12) used in CycleGAN.

$$\mathcal{L}_{identity}(G_{XY}, G_{YX}) = \mathbb{E}_{x \sim p(x)} [\|G_{YX}(x) - x\|_1] + \mathbb{E}_{y \sim p(y)} [\|G_{XY}(y) - y\|_1] \quad (12)$$

This is minimized when an image in the generator's target domain is given as input. That is, when the image in the target domain comes in, the generator does nothing. This is especially good for coloring. UNIT also showed poor results in the preservation of features. However, it is difficult to be sure whether the results are preserving the features well by only evaluating the visualization results. Therefore, we use a pre-trained classifier to test the accuracy of the resulting data sets. The accuracy was confirmed by a ROC curve and AUC.

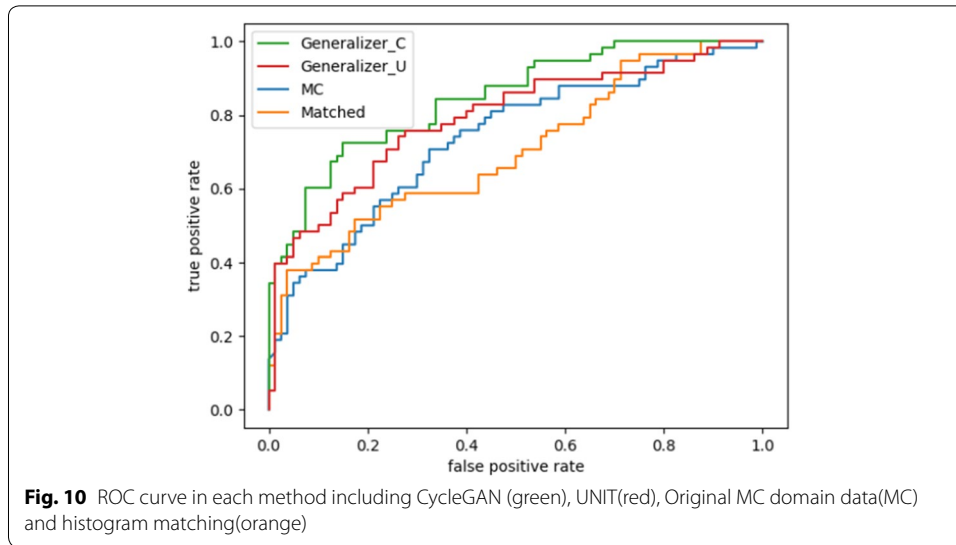


Table 2 AUC between four methods: the original data set, our two generalized data sets, and the histogram-matched data set

Methods	AUC
Generalizer—CycleGAN	0.84
Generalizer—UNIT	0.81
Original dataset	0.73
Histogram matched dataset	0.70

Figure 10 shows the ROC curve results. The difference in the curves of methods can be seen clearly. As seen in Table 2, the AUC of the CycleGAN is 0.84 and UNIT is 0.81. This is a novel score, because the AUC of the original data set is 0.73. This shows that the generalized image through the GANs performed appropriate conversion to the target domain while preserving the important features.

We have shown through experiments that the intensity generalization of medical images through GAN is effective. Generalizers using CycleGAN (given in *italic*) showed the best performance in all experiments

Conclusion

In this paper, we proposed a method to generalize the intensity of arbitrary medical images by using a GAN generalizer using CycleGAN and UNIT (based on VAE) to maintain the accuracy of a medical-image classification network. Performing generalizations without losing important features of lesions is a very sensitive task, and we evaluated the results in the following way. We created three data sets, based on two generalizers and histogram matching. We presented the detailed result images and intensity distribution of the data sets using histograms and measured the similarity of the generalized results numerically using SSIM. We also evaluated the accuracy of the proposed method and the existing method with AUC. As a result, both generalization methods using the GAN were 0.5 to 1.0 higher than the AUC of the original data set. We confirmed that the

intensity distribution of our proposed method creates images very similar to the training domain data set without significant feature loss. We have also shown that CycleGAN, which maintains the characteristics of instances, is more suitable for the generalization of medical images. These results show that our proposed generalization is an effective method to maintain performance in a classification network that suffers from performance degradation due to differences in the intensity of medical images. Recently, structure of generator that greatly improves the quality of the generated image [39, 40] and model with advanced few-shot capability [41] are proposed. As future work, Applying these methods to our generalization module would allow the robustness and accuracy of our framework.

Abbreviations

CAD: Computer-aided diagnosis; CNN: Convolutional neural network; GAN: Generative adversarial network; VAE: Variational auto-encoder; MR: Magnetic resonance; CT: Computed tomography; CDF: Cumulative distribution function; TB: Tuberculosis; MC: Montgomery County; AUC: Area under the curve; SSIM: Structural similarity index; ROC: Receiver operating characteristic.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (No. NRF-2019R1A2C1090713). This work was supported by INHA University Grant.

Authors' contributions

DHL conceptualized proposed framework for generalization of medical images, conducted all experiments and wrote the manuscript. YL and BSS supervised all experiments and advised on the manuscript. All the authors review the final manuscript for submission. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets used to support the conclusion of this study are available in the Lister Hill National Center for Biomedical Communications (LHNCBC) (<https://lhncbc.nlm.nih.gov/publication/pub9931>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 28 September 2019 Accepted: 26 March 2020

Published online: 25 April 2020

References

1. Litjens G et al (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
2. Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105
3. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. In: *Proceedings of the IEEE*, pp 2278–2324
4. Esteva A et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639):115–118
5. Sarraf S, Tofighi G, Anderson JAE. (2016) DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *bioRxiv*:070441
6. Rajpurkar P et al., (2017) CheXNet: Radiologist level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*
7. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Interv* 9351:234–241
8. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ (2017) Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging* 30(4):449–459
9. Dou Q et al (2016) Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE Trans Med Imaging* 35:1182–1195

10. Li C, Liang M, Song W, Xiao K (2018) A multi-scale parallel convolutional neural network based intelligent human identification using face information. *J Inf Process Syst* 14(6):1494–1507
11. Zhou S, Xiao S (2018) 3D face recognition: a survey. *Human Comput Inf Sci* 8(1):1
12. Sun A, Li Y, Huang YM, Li Q, Lu G (2018) Facial expression recognition using optimized active regions. *Human Comput Inf Sci* 8(1):1
13. Zhang J, Jin X, Liu Y, Sangaiah AK, Wang J (2018) Small sample face recognition algorithm based on novel Siamese network. *J Inf Process Syst* 14(6):1464–1479
14. Goodfellow I (2014) Generative adversarial nets. *Adv Neural Inf Process Syst* 27:2672–2680
15. Goodfellow I (2016) NIPS 2016 tutorial: Generative adversarial networks. *arXiv preprint* [arXiv:1701.00160](https://arxiv.org/abs/1701.00160)
16. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:1511.06434*
17. Kingma DP, Welling M (2014) Auto-encoding variational bayes. In: *International Conference on Learning Representations (ICLR)*
18. Doersch C (2016) Tutorial on variational autoencoders. *arXiv:1606.05908*
19. Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 5967–5976
20. Radford A, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp 2242–2251
21. Liu MY, Breuel T, Kautz J (2017) Unsupervised image-to-image translation networks. *Adv Neural Inf Process Syst* 30:700–708
22. Yi X, Walia E, Babyn P (2019) Generative adversarial network in medical imaging: a review. *Med Image Anal* 58:101552
23. Frid-Arari M et al (2018) GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 321:321–331
24. Wolterink JM et al. (2017) Deep MR to CT synthesis using unpaired data. *arXiv preprint* [arXiv:1708.01155](https://arxiv.org/abs/1708.01155)
25. Mirza M and Osindero S (2014) Conditional generative adversarial nets. *arXiv preprint* [arXiv:1411.1784](https://arxiv.org/abs/1411.1784)
26. Huang H, Yu PS, Wang C (2018) An introduction to image synthesis with generative adversarial nets. *arXiv preprint* [arXiv:1803.04469](https://arxiv.org/abs/1803.04469)
27. Li Y, Shen L (2018) cC-GAN: a robust transfer-learning framework for hep-2 specimen image segmentation. *IEEE Access* 6:14048–14058
28. Odena A, Olah C, Shlens J (2016) Conditional image synthesis with auxiliary classifier GANs. *arXiv preprint* [arXiv:1610.09585](https://arxiv.org/abs/1610.09585)
29. Dai W et al. (2017) Scan: Structure correcting adversarial network for chest X-rays organ segmentation. *arXiv preprint* [arXiv:1703.08770](https://arxiv.org/abs/1703.08770)
30. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst C (2018) Generating synthetic cts from magnetic resonance images using generative adversarial networks. *Med Phys* 45:3627–3636. <https://doi.org/10.1002/mp.13047>
31. Dar SUH, Yurt M, Shahdloo M, Ildiz ME, Çukur T (2018) Synergistic reconstruction and synthesis via generative adversarial networks for accelerated multi-contrast mri. *arXiv preprint* [arXiv:1805.10704](https://arxiv.org/abs/1805.10704)
32. Dar SUH, Yurt M, Shahdloo M, Ildiz ME, Çukur T (2019) Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans Med Imaging* 38(10):2375–2388. <https://doi.org/10.1109/TMI.2019.2901750>
33. Mahmood F, Chen R, Durr NJ (2018) Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE Trans Med Imaging* 37(12):2572–2581. <https://doi.org/10.1109/tmi.2018.2842767>
34. Madani A, Moradi M., Karargiris A, Syeda-Mahmood T (2018) Chest Xray generation and data augmentation for cardiovascular abnormality classification. *Medical Imaging 2018: Image Processing*:415–420
35. Mao X et al. (2017) Least squares generative adversarial networks. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp 2813–2821
36. Jaeger S et al (2014) Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg* 4(6):475–477. <https://doi.org/10.3978/j.issn.2223-4292.2014.11.20>
37. Candemir S et al (2014) Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Trans Med Imaging* 33(2):577–590. <https://doi.org/10.1109/TMI.2013.2290491>
38. Jaeger S et al (2014) Automatic tuberculosis screening using chest radiographs. *IEEE Trans Med Imaging* 33(2):233–245. <https://doi.org/10.1109/TMI.2013.2284099>
39. Karras T, Laine S, Aila T (2019) A Style-based generator architecture for generative adversarial networks. In: *2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 4401–4410
40. Karras T, Laine S, Aittala M, Hellsten J, Aila T (2019) Analyzing and improving the image quality of Style-GAN. *arXiv preprint* [arXiv:1912.04958](https://arxiv.org/abs/1912.04958)
41. Liu MY, Huang X, Mallya A, Karras T, Aila T, Lehtinen J, Kautz J (2019) Few-shot unsupervised image-to-image translation. In: *2019 IEEE International Conference on Computer Vision (ICCV)*, pp 10551–10560

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.