



Semi-supervised CycleGAN for domain transformation of chest CT images and its application to opacity classification of diffuse lung diseases

Shingo Mabu¹ · Masashi Miyake¹ · Takashi Kuremoto² · Shoji Kido³

Received: 19 April 2021 / Accepted: 31 August 2021 / Published online: 16 October 2021
© CARS 2021

Abstract

Purpose The performance of deep learning may fluctuate depending on the imaging devices and settings. Although domain transformation such as CycleGAN for normalizing images is useful, CycleGAN does not use information on the disease classes. Therefore, we propose a semi-supervised CycleGAN with an additional classification loss to transform images suitable for the diagnosis. The method is evaluated by opacity classification of chest CT.

Methods (1) CT images taken at two hospitals (source and target domains) are used. (2) A classifier is trained on the target domain. (3) Class labels are given to a small number of source domain images for semi-supervised learning. (4) The source domain images are transformed to the target domain. (5) A classification loss of the transformed images with class labels is calculated.

Results The proposed method showed an F-measure of 0.727 in the domain transformation from hospital A to B, and 0.745 in that from hospital B to A, where significant differences are between the proposed method and the other three methods.

Conclusions The proposed method not only transforms the appearance of the images but also retains the features being important to classify opacities, and shows the best precision, recall, and F-measure.

Keywords CycleGAN · Domain transformation · Semi-supervised learning · Classification · CT · Diffuse lung diseases

Introduction

Deep learning (DL) has been applied to image classifiers in computer-aided diagnosis (CAD) [17]; however, DL requires many annotated data. Also, the accuracy of CAD may fluctuate when the imaging devices are different. For example, since different CT devices and settings show the different pixel values, CAD showing good performance in a certain hospital does not always show the same performance in other

hospitals. In this case, the classifier needs to be retrained, which requires many training data again.

One of the solutions to normalize the image styles is domain transformation using CycleGAN [26]. For example, CycleGAN has been applied to classifying opacities in chest CT images [16]. However, since CycleGAN does not use labeled data, the transformation is not always suitable for the opacity classification. Hence, we propose a semi-supervised CycleGAN combined with a classifier trained to classify lung opacities in another hospital. In detail, (1) We use CT images taken at two hospitals (source and target domains). (2) A ResNet-based classifier [6] is trained on the target domain. (3) Class labels are given to a small number of source domain images. (4) Both labeled and unlabeled images in the source domain are transformed to the target domain. (5) A classification loss of the transformed images with class labels is used as an additional loss of CycleGAN. (4) and (5) are repeated to make the domain transformation suitable for the opacity classification.

There have been many studies on domain transformation. In [7], cycle-consistent domain adaptation is proposed,

✉ Shingo Mabu
mabu@yamaguchi-u.ac.jp

¹ Graduate School of Sciences and Technology for Innovation, Yamaguchi University, 2-16-1, Tokiwadai, Ube, Yamaguchi 755-8611, Japan

² Department of Information Technology and Media Design, Nippon Institute of Technology, 4-1 Gakuendai, Miyashiro-machi, Minamisaitama-gun, Saitama 345-8501, Japan

³ Graduate School of Medicine, Osaka University, 2-2, Yamadaoka, Suita, Osaka 565-0871, Japan

which adapts between domains using both generative image space alignment and latent representation space alignment. In [3], domain adaptation in person re-identification that finds the relevant images to the query is proposed. Similarity preserving GAN is used and two types of unsupervised dissimilarities are incorporated. Bak et al. [1] solved the person re-identification problem when drastic variations in illumination across surveillance cameras occur. A synthetic dataset with various illumination conditions and a domain adaptation technique are designed. Unsupervised speech domain adaptation is proposed in [8], where multiple discriminators on the power spectrogram are designed to deal with different frequency bands. Xie et al. discussed content distortion in image-to-image translation [21] and a GAN with a self-supervised module is designed to enforce the image content consistency without extra annotations. In [11], leveraging synthetic data with pixel-level labels for segmentation is described. To reduce the gap between synthetic and real domains, considering the difference between the domains as a texture, a method to adapts to the target domain's texture is proposed. In [5], emotion recognition from audio data is considered, where publicly available facial image datasets are used for audio emotion recognition by transforming the images to audio spectrograms by an adversarial network.

Modality adaptation such as between CT and MRI is actively studied research. Yang et al. achieved cross-modality domain adaptation [23], where semantic feature-level information is preserved by finding a shared content space instead of a direct pixel-wise transformation. In [10], an adversarial domain adaptation from CT to MRI is studied for tumor segmentation on MRI. In [18], a deformation invariant cycle-consistency model that can filter out the domain-specific deformation is proposed and evaluated on multi-sequence brain MR data and multi-modality abdominal CT and MR data. A domain adaptation for medical image segmentation is proposed in [2], where the method simultaneously transforms the appearance of images across domains and enhances domain-invariance of the extracted features. In [20], CycleGAN and unsupervised image-to-image translation network [13] are evaluated in the transformation of T1- and T2-weighted MR images, and two supervised models are also compared.

CycleGAN has been also applied to improve the quality of images. In [25], a supervised learning model of CycleGAN is proposed to transform low-dose PET images to full-dose images. In [14], a model combining parallel imaging with GAN for the reconstruction of MRI is proposed. This method effectively reconstructs multi-channel MR images at a low noise level for undersampling patterns. In [15], several GAN-based methods are compared to find the best methods that reconstruct MRI for undersampling images. In [24], an undersampled MRI reconstruction method based on GAN with self-attention and the relative average discriminator is

proposed to improve the speed of MRI imaging and reduce patient suffering. In [4], Wasserstein GAN and recurrent neural networks are combined to fully utilize the relationship among sequential MRI slices, and an additional attentive unit enables the method to reconstruct more accurate anatomical structures for MRI data. In [22], a conditional GAN-based model to reconstruct compressed sensing MRI is proposed, where a refinement learning method is designed to stabilize the U-Net-based generator and reduce aliasing artifacts. In addition, frequency-domain information is incorporated to enforce similarity in both the image and frequency domains.

The aim of this paper is to perform a domain transformation of chest CT images taken by different CT devices in two hospitals, and we propose a semi-supervised CycleGAN with a classification loss function to achieve domain transformation with high classification accuracy. For example, when we compare our method with image generation using the GAN-based method [4,14,15,22,24] that aim at generating high-quality images from undersampled images, our method aims to transform CT images taken at a certain hospital so that they can be accurately classified by the classifier trained in another hospital. The proposed method is trained with a semi-supervised learning manner to reduce the cost of annotation by combining CycleGAN and an additional loss based on the classification accuracy.

Materials and methods

Datasets

We used 503 chest CT images taken at Yamaguchi University Hospital, Japan (Domain A, SOMATOM Sensation 64, SIEMENS) and 636 images taken at Osaka University Hospital, Japan (Domain B, Discovery CT750 HD, GE). Generally, CycleGAN works in the entire image, identifies nonlinear regions that are to be changed and others that are kept intact. The proposed method is not the nonlinear regional transformation, as in the case of putting lines to a horse to let it appear as a zebra. Since the main difference between the images of hospitals A and B are intensity range, contrast, and the reconstruction function that generates tomographic images from X-ray projection data, the proposed method aims to normalize them. For example, domain A images are slightly darker and have smoother contours, while domain B images are lighter and have sharper contours. Both domains A and B contain six opacity classes: consolidation (CON), diffuse nodular (DN), emphysema (EMP), ground-glass opacity (GGO), honeycombing (HCM), and normal (NOR). The numbers of images of each opacity are shown in Table 1 and image examples (512×512 [pixels]) of the two domains are shown in Fig. 1.

Fig. 1 Examples of CT images of domains A and B. There are some differences in the image properties such as intensity, contrast and sharpness of the opacities

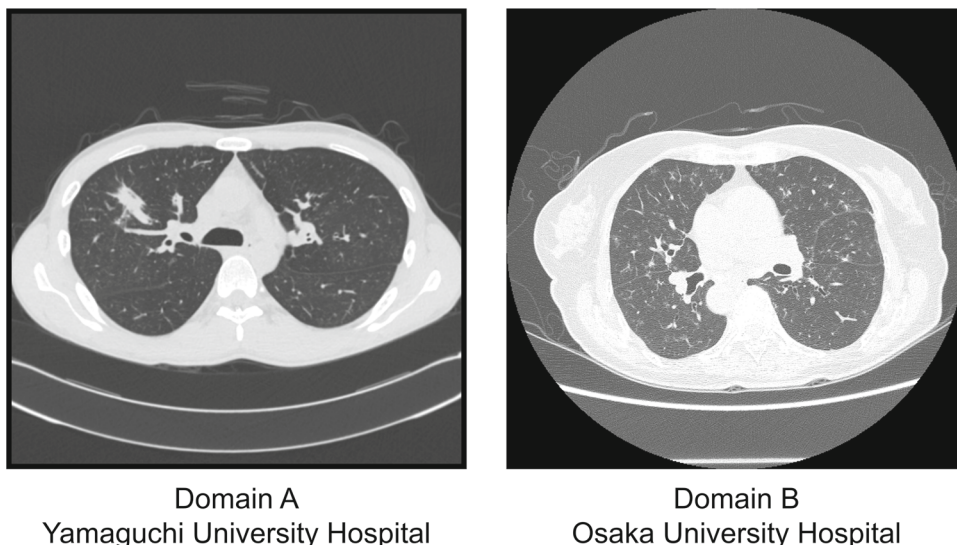


Table 1 Numbers of images of domains A and B

	Domain A	Domain B
Consolidation (CON)	109	88
Diffuse nodular (DN)	53	93
Emphysema (EMP)	112	93
Ground-Glass Opacity (GGO)	75	192
Honeycombing (HCM)	99	90
Normal (NOR)	55	90

Table 2 Numbers of ROIs and stride sizes

	Number of ROIs		Stride size [pixels]	
	Domain A	Domain B	Domain A	Domain B
CON	3071	3447	8	11
DN	3023	3311	16	14
EMP	3122	3021	24	27
GGO	3460	3273	12	18
HCM	3236	3434	13	13
NOR	3117	3035	29	32

We implemented region of interest (ROI)-based classification by dividing the CT images into 32×32 [pixels] ROIs. We chose patch-wise classification instead of pixel/voxel-wise segmentation because the number of patches for the training can be increased by extracting many patches from slices when the number of annotated CT slices is limited. The CT images have the corresponding mask images (ground truth) created by three radiologists showing the location of opacities. Figure 2 shows examples of CT images, their mask images and the extracted regions for generating ROIs. 32×32 [pixels] regions were scanned by striding from the upper left to the lower right of each CT image and the class labels were given to the regions if they contain more than 50% of the masked areas. If the stride size is the same for all the kinds of opacities, the numbers of ROIs become imbalanced. Therefore, the stride size was adjusted to extract about 3000 ROIs for each kind of opacity (Table 2).

Figure 3 shows how to split the extracted ROIs into training and testing data when ROIs of domain A are transformed to domain B. A_{train} and B_{train} are used for training, A_{test} and B_{test} are used for testing, and $A_{train_anno} \subset A_{train}$ is a small dataset with class labels. When the standard CycleGAN is trained, A_{train} and B_{train} have no class labels; however, in

this study, a small part of the training data were annotated for semi-supervised learning. In detail, CT images of five patients per opacity were annotated. In Fig. 3, the whole domain A data are split into training set A_{train} (including annotated part of domain A) and testing set A_{test} . Therefore, A_{test} is the test set for the domain A classification. Note that the testing set has been also annotated for the evaluation purpose. Actually, when the number of annotated training data is increased, the performance becomes better as we can often see in general DL. In this paper, five patients per opacity were selected by carefully considering the radiologists’ effort to make annotations and if annotation of only five patients per opacity gives positive effects on the performance, the burden on the radiologists would be reduced. Note that the ROIs extracted from the same CT image were only included in either the training data or testing data.

Methods

The semi-supervised CycleGAN (proposed method) consists of a standard CycleGAN and an opacity classifier. The upper part in Fig. 4 shows the classification flow with domain transformation and the lower part shows the flow without it. Here,

Fig. 2 CT images, mask (ground truth) images and extracted regions. CT images are the original slices, mask images show the annotated areas of opacities, and the extracted regions show the CT images that correspond to the masked areas

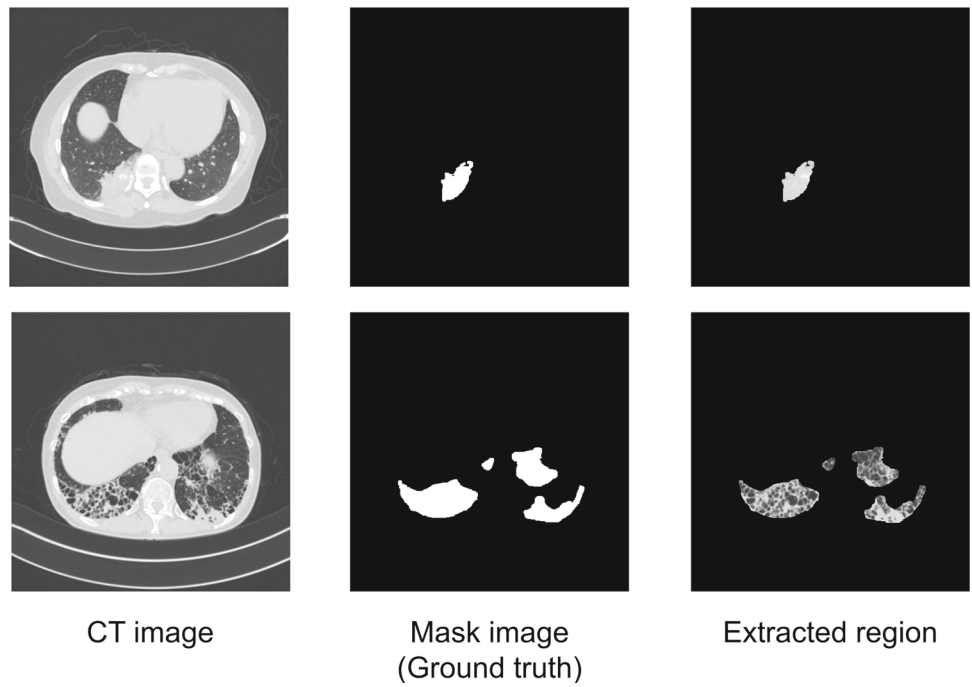
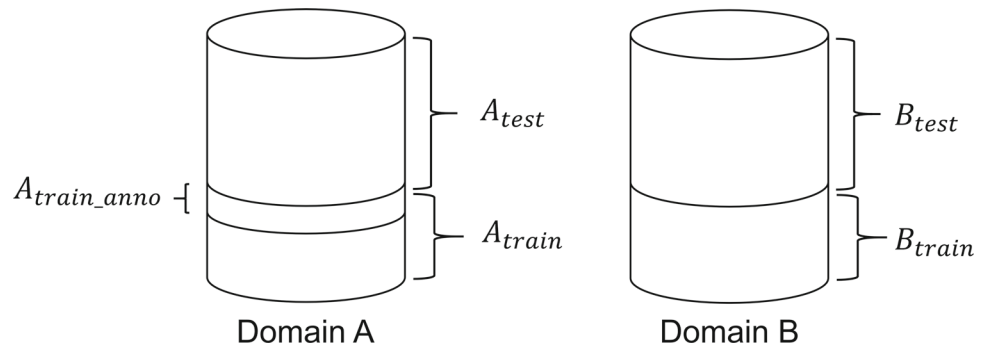


Fig. 3 Training and testing data of domains A and B when domain A data are transformed to domain B. Domain A is split into A_{train} and A_{test} , and a part of A_{train} is the training data with annotation A_{train_anno}



we suppose that a classifier (ResNet) trained on domain B is used to classify data of domain A. The proposed method transforms ROIs from domain A to B and the trained ResNet classifies the transformed ROIs. Note that the true class labels have been given to a small number of ROIs of domain A, and the loss of the ResNet is calculated when the transformed ROIs are classified. The loss is fed back to the generator that executes the transformation. This method not only adjusts the appearance of ROIs but also has the effect of clarifying the important features for opacity classification. In this paper, both “A to B” and “B to A” transformations were investigated.

Hereafter, we explain the procedure when A to B transformation is implemented. Figure 5 shows an overview of the semi-supervised CycleGAN that contains two generators G and F , two discriminators D_A and D_B , and a classifier D_{cf_B} . The training samples are $a \in A_{train}$ and $b \in B_{train}$. In the standard CycleGAN, the loss functions of Eqs. 1 through 4 are used to train G, F, D_A, D_B .

$$L_{AtoB}(G, D_B, A, B) = E_{b \sim p_{data}(b)} [\log D_B(b)] + E_{a \sim p_{data}(a)} [\log(1 - D_B(G(a)))] \tag{1}$$

$$L_{BtoA}(F, D_A, B, A) = E_{a \sim p_{data}(a)} [\log D_A(a)] + E_{b \sim p_{data}(b)} [\log(1 - D_A(F(b)))] \tag{2}$$

$$L_{cyc}(G, F) = E_{a \sim p_{data}(a)} \|F(G(a)) - a\| + E_{b \sim p_{data}(b)} \|G(F(b)) - b\| \tag{3}$$

$$L_{identity}(G, F) = E_{a \sim p_{data}(a)} \|F(a) - a\| + E_{b \sim p_{data}(b)} \|G(b) - b\| \tag{4}$$

Data distributions are denoted as $a \sim p_{data}(a)$ and $b \sim p_{data}(b)$. Generators G and F are learned by minimizing the loss of Eqs. 1 and 2, but since these loss functions alone will learn to map the same output pattern to any input images, the loss functions of Eqs. 3 and 4 are introduced [26]. Equation 3 is called cycle consistency loss, which con-

Fig. 4 Classification flow with and without domain transformation. When the domain transformation is adopted, a domain A image is inputted to the semi-supervised CycleGAN and transformed to a domain B-like image. Then, the transformed image is classified by the classifier. When the domain transformation is not adopted, a domain A image is directly inputted to the classifier

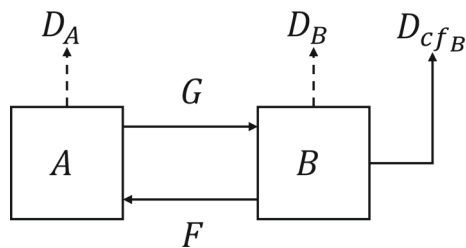
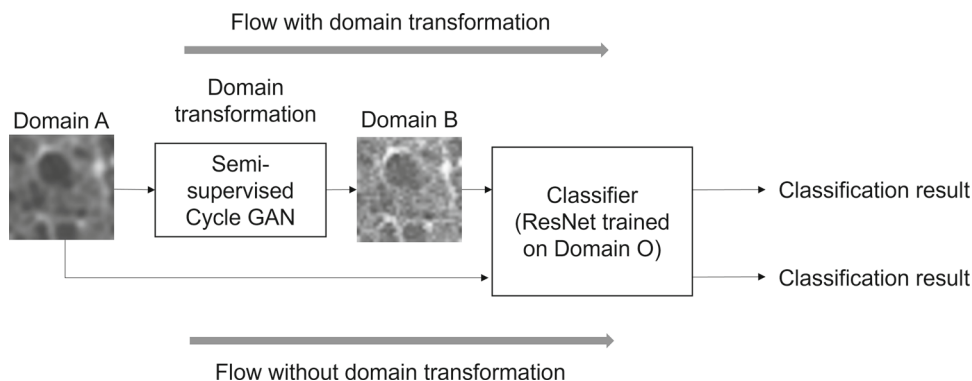


Fig. 5 Overview of a semi-supervised CycleGAN. Domain A is transformed by generator G , and domain B is transformed by generator F . D_A is a discriminator that classifies whether an inputted image is from A (real) or B (fake), and D_B classifies whether an inputted image is from B (real) or A (fake). D_{cf_B} is a classifier trained on domain B

strains the original data a and b to match the generated data $F(G(a))$ and $G(F(b))$, respectively. Equation 4 is called identity mapping loss, which constrains the generator not to convert any data that have belonged to the target domain. The structure of CycleGAN was referred to in the code provided by git repository.¹

In this paper, we designed an additional loss calculated by the ResNet. First, fake domain B data $G(a)$ are generated from domain A. Second, ResNet D_{cf_B} trained on domain B is used to classify data $G(a)$ and the loss is fed back to G to re-train. In the re-training, only the ROIs $a^{(anno)} \in A_{train_anno}$ are used and the additional loss is calculated by Eq. 5.

$$L_{resnet}(G, D_{cf_B}) = E_{a^{(anno)} \sim p_{data}(a^{(anno)})} [- \sum_{k \in C} d_k \log D_{cf_B}^{(k)}(G(a^{(anno)}))], \tag{5}$$

where C is a set of class numbers, d_k is a one-hot vector showing the correct class number, and $D_{cf_B}^{(k)}$ is an output of the ResNet for class k . Then, our full loss function is

$$L(G, F, D_A, D_B, D_{cf_B}) = L_{AtoB}(G, D_B, A, B)$$

¹ <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>.

$$+ L_{BtoA}(F, D_A, B, A) + \lambda_1 L_{cyc}(G, F) + \lambda_2 L_{identity}(G, F) + \lambda_3 L_{resnet}(G, D_{cf_B}), \tag{6}$$

where λ_1, λ_2 , and λ_3 are bias terms. λ_1 and λ_2 were set at 40, 5 and λ_3 was set at 0 from first to 100th epoch and 0.2 from 101th to 200th epoch. The proposed method uses L_{resnet} , which sometimes makes the CycleGAN destroy the original texture patterns of ROIs, thus, λ_1 and λ_2 were set at larger values than λ_3 , and λ_3 was set at a positive value after 100 epochs. In fact, we visually examined the generated ROIs in the experiments and found that the texture patterns were not destroyed. Finally, G, F, D_A , and D_B are optimized by the following objective function.

$$G^*, F^*, D_A^*, D_B^* = \arg \min_{G, F} \max_{D_A, D_B} L(G, F, D_A, D_B, D_{cf_B}), \tag{7}$$

where the weights of D_{cf_B} are fixed.

The structure of D_{cf_B} is based on ResNet34 [6] as shown in Table 3. The residual block shown in Fig. 6 is used in Conv2, Conv3, Conv4, and Conv5. For example, Conv2 uses three residual blocks with two convolution layers with kernel size 3×3 and channel size 64. After Conv5, a fully connected layer is used to output six values that correspond to the probabilities of belonging to six kinds of opacities, respectively.

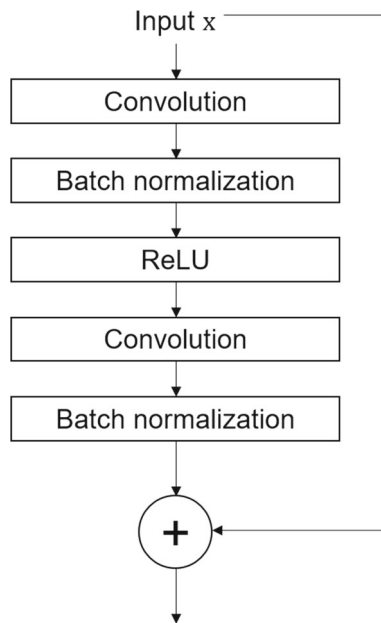
Results

Experimental setup

The numbers of ROIs are shown in Table 4, where the numbers in parentheses show the numbers of ROIs with class labels, i.e., A_{train_anno} and B_{train_anno} . Figure 7 shows four methods for comparison when A to B transformation is executed. Method 1 is the proposed method, and Method 2

Table 3 Structure of 34-layered ResNet

Layer name	Output size	Residual block type
Conv1	32×32	3×3 , stride 1
Conv2	32×32	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$
Conv3	16×16	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$
Conv4	8×8	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$
Conv5	4×4	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$
Fully connected	1×1	Average pooling 6-d fully connected

**Fig. 6** Structure of a residual block. Input x is transformed by convolution, batch normalization, and ReLU. Then, the output is the sum of the transformed x and the original input x

is based on the standard CycleGAN. Method 3 does not use domain transformation and directly inputs the ROIs of domain A to the ResNet trained on domain B. In Method 4, domain transformation is not used and the ResNet is trained on $A_{\text{train_anno}}$. The aim of this paper is to add the classification loss to CycleGAN and evaluate the effects on the classification performance when a small number of annotated data are given. Thus, if Method 1 is better than Method 2, the main objective, i.e., the effect of the additional loss is verified. In addition, to show more results for the comparison, Method 3 without domain transformation is evaluated. Also, if Method 4 is better than Method 1, the domain transformation is fundamentally meaningless, i.e., the training in the single domain is enough; thus, we conducted the comparison.

The evaluation metrics are precision, recall, and F-measure calculated by averaging the results of 20 indepen-

Table 4 Numbers of ROIs used for the training and testing. (-) shows the numbers of ROIs with annotation used to calculate the loss of the ResNet

Class	Domain A		Domain B	
	Training	Testing	Training	Testing
CON	1022 (95)	2049	1027 (104)	2420
DN	1018 (107)	2005	1020 (98)	2291
EMP	1020 (105)	2102	962 (105)	2059
GGO	989 (108)	2471	996 (108)	2277
HCM	1003 (96)	2233	1021 (96)	2413
NOR	1003 (105)	2114	1024 (101)	2011
Total	6055 (616)	12974	6050 (612)	13471

dent trials. In fact, we aimed to generate new gray-scale images that can be correctly classified by the classifiers in the target domain. In this sense, the aim of this paper is to increase the classification performance on the generated images. Therefore, precision, recall and F-measure were used, which are directly related to evaluating the classification performance.

Domain transformation from A to B

First, the ResNet was trained using all the ROIs of domain B (Table 2). The pixel values were normalized to $[-1, 1]$, the number of epochs was set at 20, the batch size was set at 16, and Adam [12] was used for training. After the training, the accuracy for the training data was 98.1%.

Next, the domain transformation was learned for 200 epochs with batch size 16. Figure 8 shows examples of the domain transformation from A to B and the reconstruction from B to A, where the ROIs of domain A are transformed to domain B-like images, and the reconstructed images still keep the textures of the original images.

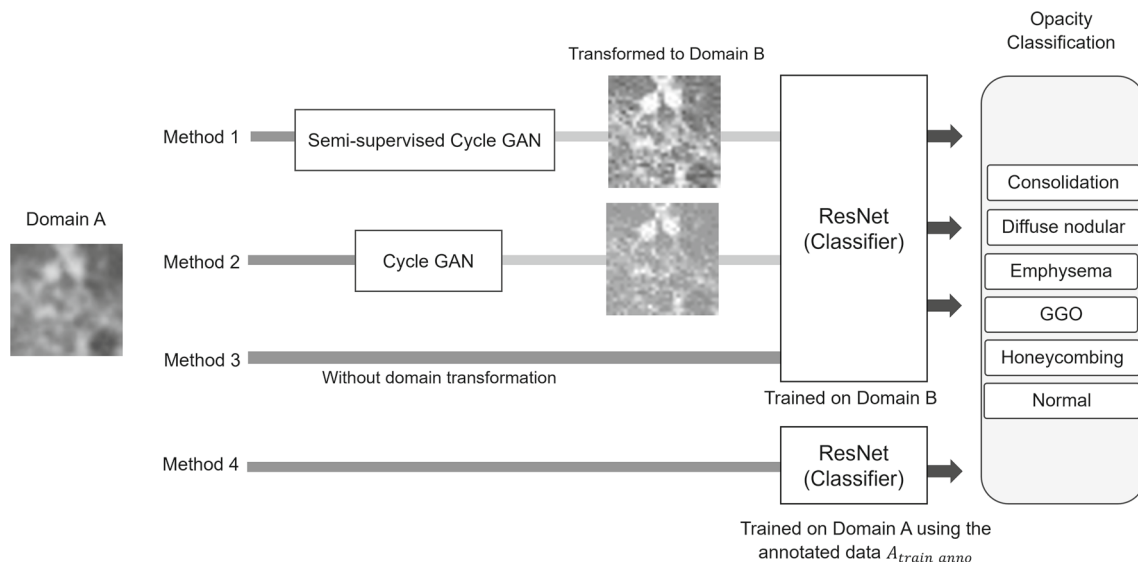


Fig. 7 Methods for comparison. Method 1 is the proposed method. Method 2 uses the standard CycleGAN. Method 3 does not use domain transformation, but directly input images of domain A to the classifier

Precision, recall and F-measure obtained by the four methods are shown in Tables 5, 6 and 7, respectively,² where Method 1 shows the best results. T-test on the mean F-measures between Method 1 and other methods shows the significant differences. The p-value between Method 1 and 2 is 4.73×10^{-7} that between Method 1 and 3 is 1.83×10^{-17} , and that between Method 1 and 4 is 3.34×10^{-6} . Since Method 1 is better than Method 2, the additional loss (Eq. 5) is effective to transform ROIs while retaining useful opacity features for classification. According to the results of Method 3, just diverting the trained ResNet does not show good performance and the image transformation is important to adapt to another domain. When comparing Method 1 and 4, although the given data with annotation are the same, Method 1 is better than Method 4. Method 4 performs worse than Method 1 because the number of training data is too small to sufficiently train the ResNet. On the other hand, Method 1 effectively makes use of the limited number of annotated data to learn the domain transformation; thus, the performance becomes better. If enough training data of the source domain can be available, Method 4 achieves better performance by sufficiently tuning the parameters.

To show the baseline of the classification performance in case where enough training data in the same domain is

trained on domain B. Method 4 does not use domain transformation, but trains the classifier using the annotated images of domain A

Table 5 Precision obtained by Method 1, 2, 3 and 4 in the domain transformation from A to B

	Method			
	1	2	3	4
CON	0.986	0.980	0.867	0.977
DN	0.720	0.702	0.268	0.391
EMP	0.555	0.439	0.131	0.584
GGO	0.772	0.660	0.486	0.884
HCM	0.775	0.703	0.408	0.790
NOR	0.627	0.599	0.014	0.538
Mean	0.740	0.679	0.365	0.701

available, a ResNet was trained on domain A, i.e., A_{train} and evaluated on domain A, i.e., A_{test} . As a result, the mean precision is 0.837, the mean recall is 0.819 and the mean F-measure is 0.819. Therefore, preparing enough training data in the same domain is important as a first step to build a classification model; however, when it is difficult, the domain transformation is effective.

Domain transformation from B to A

Next, ROIs of domain B were classified by the ResNet trained domain A. The ResNet was trained using all the ROIs of domain A (Table 2), and the accuracy for the training data was 98.8%. Then, the domain transformation was learned for 200 epochs with batch size 16. Figure 9 shows examples of the transformation from B to A, and the reconstructed images.

² In Tables 5 through 10, “Mean” values are different from those simply calculated based on the values of the six opacities in each table. “Mean” represents the mean of 20 trials, where, in each trial, a weighted average of six opacities is calculated. Also, F-measures in Tables 7 and 10 are different from those calculated based on the values in Tables 5, 6, 8, and 9, i.e., they are the average F-measures over 20 trials.

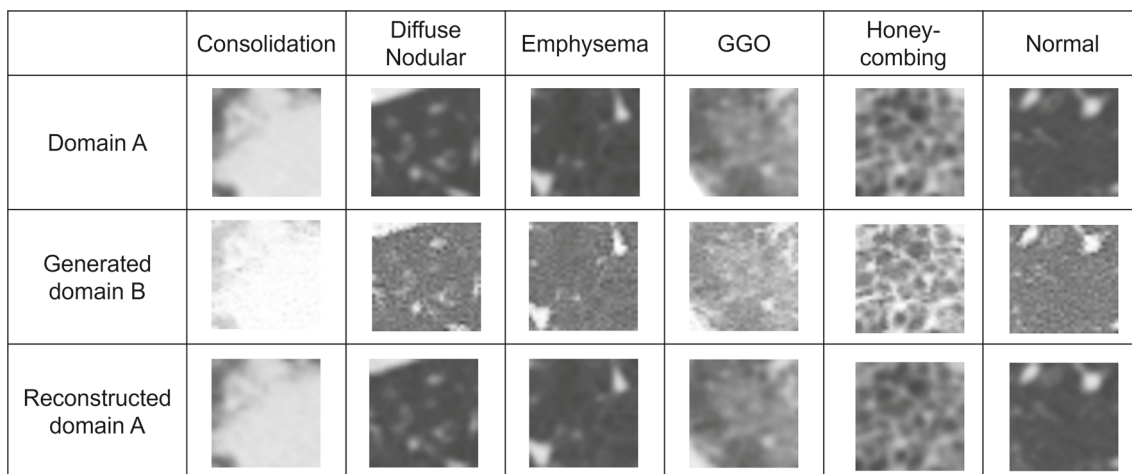


Fig. 8 Examples of the ROIs generated by the domain transformation (A to B). The row of “Domain A” shows the original ROI used as inputs. The row of “Generated domain B” shows the result of domain transfor-

mation $A \rightarrow B$. The row of “Reconstructed domain A” shows the result of domain transformation $A \rightarrow B \rightarrow A$

Table 6 Recall obtained by Method 1, 2, 3 and 4 in the domain transformation from A to B

	Method			
	1	2	3	4
CON	0.911	0.899	0.284	0.873
DN	0.563	0.497	0.063	0.408
EMP	0.623	0.526	0.549	0.517
GGO	0.716	0.730	0.352	0.587
HCM	0.869	0.807	0.400	0.840
NOR	0.670	0.469	0.042	0.719
Mean	0.727	0.658	0.286	0.658

Table 7 F-measure obtained by Method 1, 2, 3 and 4 in the domain transformation from A to B

	Method			
	1	2	3	4
CON	0.947	0.937	0.372	0.901
DN	0.626	0.571	0.051	0.376
EMP	0.582	0.473	0.208	0.513
GGO	0.742	0.687	0.374	0.700
HCM	0.818	0.739	0.308	0.805
NOR	0.643	0.517	0.021	0.593
Mean	0.727	0.655	0.228	0.652

The classification performance are shown in Tables 8, 9 and 10, where Method 1 shows the best results. T test on the mean F-measures shows significant differences between Method 1 and other methods, where the p-value between Method 1 and 2 is 6.00×10^{-3} , that between Method 1 and 3 is 5.80×10^{-42} , and that between Method 1 and 4

Table 8 Precision obtained by Method 1, 2, 3 and 4 in the domain transformation from B to A

	Method			
	1	2	3	4
CON	0.993	0.994	0.880	0.927
DN	0.697	0.668	0.213	0.577
EMP	0.639	0.629	0.053	0.786
GGO	0.834	0.827	0.177	0.816
HCM	0.800	0.788	0.627	0.729
NOR	0.566	0.556	0.037	0.538
Mean	0.764	0.752	0.350	0.734

is 2.25×10^{-13} . Method 1 shows better F-measure (0.745) in B to A transformation than A to B (0.727). However, the difference between Method 1 and 2 in A to B transformation (0.072) is larger than B to A (0.015), which shows that A to B transformation is more difficult for the standard CycleGAN because it cannot emphasize the opacity features without class label information. In B to A transformation, original ROIs of domain B may have clear features for classification, thus it is relatively easy for the standard CycleGAN to transform the domains. To clarify under what kinds of conditions the opacity features should be emphasized is a remaining problem.

The classification performance in case where we have enough training data in the same domain is also shown. When a ResNet is trained on domain B, i.e., B_{train} , and evaluated on domain B, i.e., B_{test} , the mean precision is 0.850, the mean recall is 0.822 and the mean F-measure is 0.818.

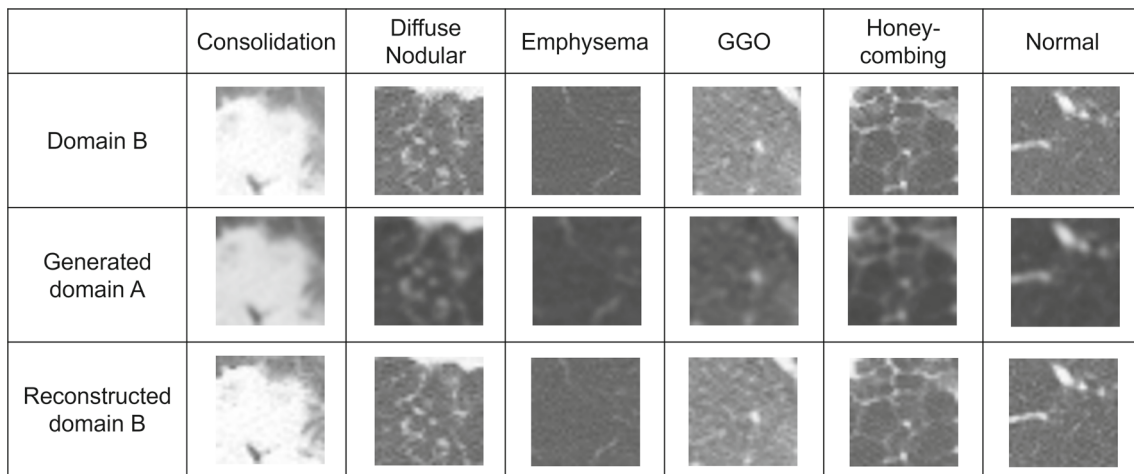


Fig. 9 Examples of the ROIs generated by the domain transformation (B to A). The row of “Domain B” shows the original ROI used as inputs. The row of “Generated domain A” shows the result of domain transfor-

mation $B \rightarrow A$. The row of “Reconstructed domain B” shows the result of domain transformation $B \rightarrow A \rightarrow B$

Table 9 Recall obtained by Method 1, 2, 3 and 4 in the domain transformation from B to A

	Method			
	1	2	3	4
CON	0.909	0.905	0.996	0.848
DN	0.761	0.773	0.022	0.396
EMP	0.482	0.445	0.000	0.741
GGO	0.659	0.633	0.568	0.659
HCM	0.884	0.877	0.489	0.762
NOR	0.741	0.722	0.024	0.796
Mean	0.747	0.734	0.370	0.699

Table 10 F-measure obtained by Method 1, 2, 3 and 4 in the domain transformation from B to A

	Method			
	1	2	3	4
CON	0.949	0.947	0.934	0.854
DN	0.723	0.715	0.037	0.448
EMP	0.541	0.503	0.000	0.709
GGO	0.733	0.713	0.269	0.717
HCM	0.839	0.827	0.541	0.740
NOR	0.637	0.622	0.028	0.629
Mean	0.745	0.730	0.319	0.687

Discussion

In this section, discussion and some remaining problems are described. First, many methods can be applied to image normalization. In this paper, we adopted one of the methods, i.e., CycleGAN, and aimed to enhance the normalization

ability of CycleGAN for the classification. Since our main proposal is the additional classification loss to CycleGAN in a semi-supervised learning manner, the effect of the method with a small number of annotated data is mainly compared to the base method, i.e., the original CycleGAN without the additional loss. Also, as our initial motivation, we supposed that it is difficult for us to judge the important global and local features to be transformed to improve the classification performance; thus, the end-to-end transformation method was considered instead of applying some image processing techniques. In addition, we would like to find the important features by directly using the classification loss because the final objective is to maximize the classification performance. In the proposed method, global features (e.g., intensity, contrast, etc.) and local features (e.g., textures) for better classification are transformed by combining CycleGAN and the additional classification loss. However, in terms of explainability, we may need to analyze the filters generated in the convolution layers in CycleGAN in the future research.

We should consider the difference in the feature distribution of labeled and unlabeled data. In the proposed method, when giving opacity labels to a small number of data, a sampling bias would occur, causing discrepancies in the empirical distribution between labeled and unlabeled data [19]. We randomly selected the annotated data, but the problem of the sampling bias has not been solved yet. If the distribution of the labeled data deviates from the actual data distribution, it may be difficult to learn an appropriate domain transformation. To reduce the bias, it is necessary to consider training data augmentation that gives class labels to the unlabeled data, where the class labels are assigned to the data for which ResNet in the semi-supervised CycleGAN shows high

classification confidence. This problem should be studied in the future.

The explainability of the classification is also discussed. Hu et al. [9] aims at not only identifying the diseases of COVID-19 but also identifying the locations using CNN, where the influences of each pixel on the neuron activation in the target maps are calculated. In our method, patch-based classification can identify disease locations to some extent, but for pixel-based segmentation or bounding box detection, it is necessary to use the activation status of neurons, as used in [9].

To further evaluate the classification ability, ideally, the test on an external dataset should be done. Currently, the experiments are executed using the CT datasets obtained by two hospitals; however, we are planning to apply the proposed method to other datasets, e.g., CT images obtained by another hospital and not only CT images but also pathological images, to confirm the performance.

We used the identity mapping loss (Eq. 4) to implement the experiments in the same conditions as the original CycleGAN that has been widely used in the world. In this paper, however, the single-to-single domain transformation is executed; thus, Eq. 4 does not have effects on the transformation. Nevertheless, when we consider the domain transformation from multiple source domains to target domains in the future, Eq. 4 would be still effective.

There are many techniques to overcome the problem of the small amount of data and one of the techniques is pre-training and fine-tuning. However, in this paper, we focused on the different approach where the normalization is applied to the source domain and the well-trained classifier on the target domain is reused to reduce the annotation cost. To realize this approach, we designed the additional loss and evaluated the effects of the designed loss comparing with the method without the additional loss. In the future, it may be worthwhile to combine pre-training, fine-tuning, and domain transformation to further improve the classification performance.

Conclusions

We investigated the domain transformation of chest CT images using a semi-supervised CycleGAN so that a classifier trained at a certain hospital can be used at another hospital. The proposed method not only transforms the appearance of the images but also preserves features being important to classify lung opacities. We used the chest CT images of domain A and B and simulated the two cases where domain A is transformed to domain B, and vice versa. As a result, the effectiveness of the proposed method was confirmed. In the future, we will solve the remaining problems described in the previous sections, then apply the proposed

method to build a large-scale medical image datasets with annotation.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number 19K12120.

Declarations

Conflict of interest Shingo Mabu and Takashi Kuremoto received JSPS KAKENHI Grant Number 19K12120.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Bak S, Carr P, Lalonde JF (2018) Domain adaptation through synthesis for unsupervised person re-identification. In: Proceedings of the European conference on computer vision (ECCV), pp 189–205
2. Chen C, Dou Q, Chen H, Qin J, Heng PA (2019) Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 865–872
3. Deng W, Zheng L, Ye Q, Kang G, Yang Y, Jiao J (2018) Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)
4. Guo Y, Wang C, Zhang H, Yang G (2020) Deep attentive wasserstein generative adversarial networks for mri reconstruction with recurrent context-awareness. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 167–177
5. He G, Liu X, Fan F, You J (2020) Classification-aware semi-supervised domain adaptation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops
6. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
7. Hoffman J, Tzeng E, Park T, Zhu JY, Isola P, Saenko K, Efros A, Darrell T (2018) Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning, PMLR, pp 1989–1998
8. Hosseini-Asl E, Zhou Y, Xiong C, Socher R (2018) A multi-discriminator CycleGAN for unsupervised non-parallel speech domain adaptation. Proc Interspeech 2018:3758–3762
9. Hu S, Gao Y, Niu Z, Jiang Y, Li L, Xiao X, Wang M, Fang EF, Menpes-Smith W, Xia J, Ye H, Yang G (2020) Weakly supervised deep learning for COVID-19 infection detection and classification from CT images. IEEE Access 8:118,869–118,883
10. Jiang J, Hu YC, Tyagi N, Zhang P, Rimner A, Mageras GS, Deasy JO, Veeraraghavan H (2018) Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation. In: Medical

- image computing and computer assisted intervention—MICCAI 2018. Springer International Publishing, Cham, pp 777–785
11. Kim M, Byun H (2020) Learning texture invariant representation for domain adaptation of semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 12,975–12,984
 12. Kingma D, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
 13. Liu MY, Breuel T, Kautz J (2017) Unsupervised image-to-image translation networks. In: Advances in neural information processing systems, pp 700–708
 14. Lv J, Wang C, Yang G (2021a) PIC-GAN: A parallel imaging coupled generative adversarial network for accelerated multi-channel MRI reconstruction. *Diagnostics*. <https://doi.org/10.3390/diagnostics11010061>
 15. Lv J, Zhu J, Yang G (2021b) Which GAN? A comparative study of generative adversarial network-based fast MRI reconstruction. *Philos Trans R Soc A* 379(2200):20200203
 16. Miyake M, Mabu S, Kido S, Kuremoto T, Hirano Y (2017) Domain transformation of chest CT images using cycle GAN and its application to classification systems. In: The 38th JAMIT annual meeting, pp 108–115 (in Japanese)
 17. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 35(5):1285–1298
 18. Wang C, Yang G, Papanastasiou G, Tsaftaris SA, Newby DE, Gray C, Macnaught G, MacGillivray TJ (2021) Dicyc: Gan-based deformation invariant cross-domain information fusion for medical image synthesis. *Inf Fus* 67:147–160
 19. Wang Q, Li W, Gool LV (2019) Semi-supervised learning by augmented distribution alignment. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 1466–1475
 20. Welander P, Karlsson S, Eklund A (2018) Generative adversarial networks for image-to-image translation on multi-contrast MR images—a comparison of cyclegan and unit. arXiv preprint [arXiv:1806.07777](https://arxiv.org/abs/1806.07777)
 21. Xie X, Chen J, Li Y, Shen L, Ma K, Zheng Y (2020) Self-supervised cyclegan for object-preserving image-to-image domain adaptation. In: Vedaldi A, Bischof H, Brox T, Frahm JM (eds) Computer vision—ECCV 2020. Springer International Publishing, Cham, pp 498–513
 22. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, Liu F, Arridge S, Keegan J, Guo Y, Firmin D (2018) DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging* 37(6):1310–1321
 23. Yang J, Dvornek NC, Zhang F, Chapiro J, Lin M, Duncan JS (2019) Unsupervised domain adaptation via disentangled representations: application to cross-modality liver segmentation. In: Medical image computing and computer assisted intervention—MICCAI 2019. Springer International Publishing, Cham, pp 255–263
 24. Yuan Z, Jiang M, Wang Y, Wei B, Li Y, Wang P, Menpes-Smith W, Niu Z, Yang G (2020) SARA-GAN: Self-attention and relative average discriminator based generative adversarial networks for fast compressed sensing MRI reconstruction. *Front Neuroinform* 14:20
 25. Zhou L, Schaefferkoetter JD, Tham IW, Huang G, Yan J (2020) Supervised learning with cyclegan for low-dose FDG pet image denoising. *Med Image Anal* 65(101):770
 26. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision, pp 2223–2232

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