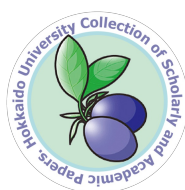




HOKKAIDO UNIVERSITY

Title	Synthetic Gastritis Image Generation via Loss Function-Based Conditional PGGAN
Author(s)	Togo, Ren; Ogawa, Takahiro; Haseyama, Miki
Citation	IEEE Access, 7, 87448-87457 https://doi.org/10.1109/ACCESS.2019.2925863
Issue Date	2019-07-01
Doc URL	https://hdl.handle.net/2115/75023
Rights(URL)	https://creativecommons.org/licenses/by/3.0/
Type	journal article
File Information	01SyntheticGastritisImageGenerationViaLossFunction-basedConditionalPGGAN.pdf



Received June 20, 2019, accepted June 27, 2019, date of publication July 1, 2019, date of current version July 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2925863

Synthetic Gastritis Image Generation via Loss Function-Based Conditional PGGAN

REN TOGO^{ID}, (Member, IEEE), TAKAHIRO OGAWA^{ID}, (Senior Member, IEEE),
AND MIKI HASEYAMA^{ID}, (Senior Member, IEEE)

Faculty of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

Corresponding author: Ren Togo (togo@lmd.ist.hokudai.ac.jp)

This work was supported in part by the JSPS KAKENHI under Grant JP17H01744 and Grant JP19J10821.

ABSTRACT In this paper, a novel synthetic gastritis image generation method based on a generative adversarial network (GAN) model is presented. Sharing medical image data is a crucial issue for realizing diagnostic supporting systems. However, it is still difficult for researchers to obtain medical image data since the data include individual information. Recently proposed GAN models can learn the distribution of training images without seeing real image data, and individual information can be completely anonymized by generated images. If generated images can be used as training images in medical image classification, promoting medical image analysis will become feasible. In this paper, we targeted gastritis, which is a risk factor for gastric cancer and can be diagnosed by gastric X-ray images. Instead of collecting a large amount of gastric X-ray image data, an image generation approach was adopted in our method. We newly propose loss function-based conditional progressive growing generative adversarial network (LC-PGGAN), a gastritis image generation method that can be used for a gastritis classification problem. The LC-PGGAN gradually learns the characteristics of gastritis in gastric X-ray images by adding new layers during the training step. Moreover, the LC-PGGAN employs loss function-based conditional adversarial learning so that generated images can be used as the gastritis classification task. We show that images generated by the LC-PGGAN are effective for gastritis classification using gastric X-ray images and have clinical characteristics of the target symptom.

INDEX TERMS Generative adversarial network, anonymization, deep learning, data sharing, medical image analysis.

I. INTRODUCTION

With the development of image recognition technologies, there have been expectations of their applications to clinical devices in the field of medicine [1]. Although recent machine learning-based image recognition techniques have shown good prospects as diagnostic supporting systems [2], the usefulness of these techniques is still limited. There have been many studies in which machine learning techniques were applied for world-renowned inspections (*e.g.*, CT, MRI, and mammography; hereafter called major inspections) [3]–[5]. In major inspections, well-equipped research environments such as large-scale annotated medical image datasets have already been established [6], [7], and we can easily access such large-scale public clinical datasets containing data obtained from all over the world [8]. However, there have

been few studies on regional inspections that are executed in a certain limited area (hereafter called minor inspections). In minor inspections, it is difficult to obtain high-quality annotated clinical data since there are different types of imaging equipment, which affect the quality of images, in medical facilities and the detailed annotation requires specialist knowledge. Moreover, new imaging methods have frequently been developed for minor inspections and even if we can construct an annotated clinical dataset from such data, updating training data becomes a fundamental function unlike that for established major inspections. It is necessary to prepare a new dataset that is adapted to the new imaging methods. Hence, minor inspections suffer from a lack of high-quality accessible data, and data-driven approaches are needed to apply machine learning techniques.

Sharing data is one of the effective data-driven approaches that can solve the problems of a lack of high-quality data and the need to update training data [9]. However, preservation

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

of patient privacy should be the top priority in the process to share clinical information [10]. In the field of medicine, it has been considered that privacy information and convenience of data have an inverse proportional relation, and that is one of the challenging problems. Clinical image data include not only privacy information but also identifiers such as social security number, gender, age, and occupation, and such data should be treated carefully. The possibility of re-identification increases when amount of data becomes small. Although the removal of identifiers has been performed for anonymization [11], attention has not been paid to anonymization of image data. In order to accelerate the use of machine learning-based support for minor inspections, simplified approaches considering privacy prevention issues are required.

In recent years, medical image generation methods that enable meaningful synthetic information to be generated have attracted much attention [12], [13]. The wide availability of synthetic data may allow researchers to develop and validate more sophisticated image recognition techniques [14]. Namely, since image generation methods learn the distribution of training data without referring to real images, anonymization of individual information can be realized [15]. The use of synthetic data will contribute to image recognition tasks for minor inspections that require sharing and updating clinical data.

In this study, we targeted gastric X-ray images for the diagnosis of gastritis/non-gastritis. Gastritis is a key factor for the onset of gastric cancer, and gastritis can be diagnosed by gastric X-ray images [16], [17]. In some East Asian countries including South Korea and Japan, which have the highest gastric cancer mortality rates in the world, gastric cancer mass screening based on gastric X-ray inspections has been started [18]. Although gastric X-ray inspection is a traditional and important modality, it is still minor inspection compared to CT or MR inspections. Data-driven approaches are needed to introduce machine learning techniques in the field of gastric cancer mass screening [19].

We propose LC-PGGAN, a Loss function-based Conditional Progressive Growing Generative Adversarial Network, in this paper. The proposed method learns the distribution of the target data and realizes image generation following the distribution from a latent space. Therefore, generated synthetic images are not associated with individual patient image information, and they can easily be used by researchers to develop supporting systems. The LC-PGGAN consists of two novel points. The first one is a progressive growing network architecture. GANs easily fall into a mode of collapse in which all input noise vectors are mapped to the same output image, and optimization fails to make progress. We solve this problem by making our networks progressively learn the target distribution from low resolution to high resolution. The second novel point is loss function-based conditional adversarial training. Typical conditional GANs try to generate label-domain images by using multiple networks and one-hot vector representation. Then a discriminator plays the role

of discrimination of conditional information in addition to the task of real-fake discrimination. However, it becomes difficult to implement stable training when the class domain classification task is difficult. In medical images, there always exist abnormal and normal samples, and their differences are subtle. We have designed our method to control the conditional information based on adversarial loss functions, which is the most efficient way for performing training. Namely, efficient training is realized by adding new loss functions in the high resolution step. Although conventional one-hot vector representation approaches force the model to train a different domain classification task in the early training stage, our model does not have to perform such a difficult task in the early stage. This contributes to the realization of efficient training of the adversarial network. For improving gastritis classification performance using anonymized generated images, we mix images generated by LC-PGGAN that have conditional information and images generated by PGGAN that have rich diversity. The three main contributions of LC-PGGAN are summarized below:

- Generating anonymized synthetic medical image data
- Enabling stable training based on the progressive growing network architecture
- Controlling conditional information based on the conditional loss function for efficient training

The rest of this paper is arranged as follows. We briefly review related works in Sec. II. In Sec. III, we show the details of our synthetic image generation approach. Experimental results are provided in Sec. IV. We conclude our paper in Sec. V.

II. RELATED WORKS

In this section, we begin with an explanation of the basic concept of GANs in II-A, and we review more specific relevant works on medical data synthesis in II-B.

A. GENERATIVE ADVERSARIAL NETWORK

A GAN is an implicit generative model that competitively learns neural networks firstly proposed by Goodfellow *et al.* in 2014 [20]. A basic GAN consists of two neural network models, a generative model G that learns the unseen training data distribution and a discriminative model D that learns to classify whether samples come from the training data distribution. When given the prior distribution of the latent variable z following the latent distribution $p_z(z)$, the generator G takes z as an input vector and outputs a sample $G(z)$. On the other hand, the discriminator D takes a sample x as an input and outputs $D(x)$, which represents the probability that is real data. Those two models are trained simultaneously with a stochastic gradient descent (SGD) [21] algorithm, and their training procedures can be seen as a two players' mini-max game with the following objective function:

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

where the discriminator D tries to maximize $V(D, G)$, while the generator G tries to minimize it. In other words, the discriminator D distinguishes the images in $\mathbf{x} \sim p_{data}$ from the ones of $G(\mathbf{z})$, while the generator G generates samples to fool the discriminator D .

The concept of GANs has been applied to various tasks. In conditional image generation, supervised and unsupervised domain transformations of images have been explored. For instance, pix2pix proposed by Isola *et al.* learns an image-to-image translation task using paired data samples [22]. However, this approach requires a large number of paired samples despite the difficulty of obtaining annotated paired samples. In order to address this problem, unpaired image-to-image translation frameworks have been proposed by many researchers (*e.g.*, UNIT [23], CoGAN [24], CycleGAN [25], and DiscoGAN [26]). Moreover, recent domain classification-based GANs that control the characteristics of generated images by operating a latent distribution have shown promising results [27]. When given a large dataset for the training that is easy to access, conventional methods have already achieved the generation of high-quality images.

B. DATA SYNTHESIS IN MEDICAL IMAGES

In medical image applications, preservation of patient privacy is the top priority, and this strict regulation makes accessing and collecting clinical data much more difficult than accessing and collecting data for natural images. Data synthesis approaches have recently been used to overcome this problem. Noseong *et al.* proposed an anonymization technique for clinical table data using GANs, and their synthesized tables can be shared without any concern about information leakage [28]. The advantage of generating synthetic images by GANs is that the generated images can preserve only the characteristics of training data that are effective for image recognition tasks without having any individual information. In other words, there is no one-to-one relationship between real images and synthetic images in a GAN-based approach, which makes it difficult to re-identify the anonymized information. Therefore, it can be considered that anonymization using generative models is a useful and safe approach for realization of data sharing.

Even if we can access medical datasets, they are often highly imbalanced with a paucity of data from rare conditions. Hojjat *et al.* tried to address this problem of an imbalance by synthesized images in chest pathology classification [29]. They employed a Deep Convolutional GAN [30] architecture for generating synthetic chest X-ray images. The classification performance was improved with synthetic images for balancing the dataset, though the resolution of synthetic images was lower than that of real images. In image-to-image translation approaches, Pedro *et al.* proposed a synthetic retinal image generation method motivated by the above-mentioned assumption [14].

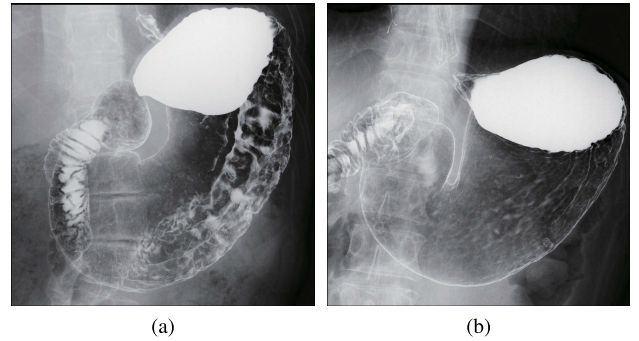


FIGURE 1. Examples of gastric X-ray images used in this study: (a) an image with gastritis and (b) an image without gastritis (non-gastritis).

III. PROPOSED IMAGE GENERATION METHOD

In this section, a gastritis image generation method, Loss function-based Conditional Progressive Growing GAN (LC-PGGAN), is presented. Firstly, in Subsec. III-A, we explain the data production for gastritis image generation using gastric X-ray images. In Subsec. III-B, the details of our progressive growing network architecture are provided. Finally, we explain how to train our model in Subsec. III-C.

A. DATA PRODUCTION

In this subsection, we propose an approach toward the realization of gastritis image generation using gastric X-ray images with consideration of clinical settings. Figure 1 shows examples of gastric X-ray images used in this study: where (a) is a sample with gastritis and (b) is a sample without gastritis (hereafter called non-gastritis). A stomach with gastritis has coarse mucosal surface patterns and non-straight folds, whereas a stomach without gastritis has uniform mucosal surface patterns and straight folds. Gastric X-ray images have a high resolution (*e.g.*, $1,024 \times 1,024$ or $2,048 \times 2,048$ pixels), for which computational costs are high. In our previous investigation, we found that a patch-division is the best approach for the gastritis classification problem using gastric X-ray images since the differences between gastritis and non-gastritis images are in local regions of the images [31]–[33]. If we use resized gastric X-ray images for gastritis classification, extracted image features cannot show the characteristics of gastritis/non-gastritis. Therefore, we use divided patch images for the generation of synthetic images in the same manner as that in our previous works.

Firstly, we divide gastric X-ray images into multiple patches. Let $F_i (i = 1, 2, \dots, I) \in \mathbb{R}^{d \times d}$ denote gastric X-ray images for the image generation, where I is the number of training images, and their class labels are denoted as $y_i \in \{1, -1\}$. Specifically, F_i is divided into $H \times W$ patches (H and W being the numbers of patches in the vertical direction and horizontal direction, respectively), and we define $X_i^{(h,w)} \in \mathbb{R}^{\hat{d} \times \hat{d}}$ ($h = 1, 2, \dots, H; w = 1, 2, \dots, W$) which represent patches extracted from F_i . Next, we classify the divided patches $X_i^{(h,w)}$ into the following three kinds of data:

- $\mathcal{A}$: data including gastritis patches,

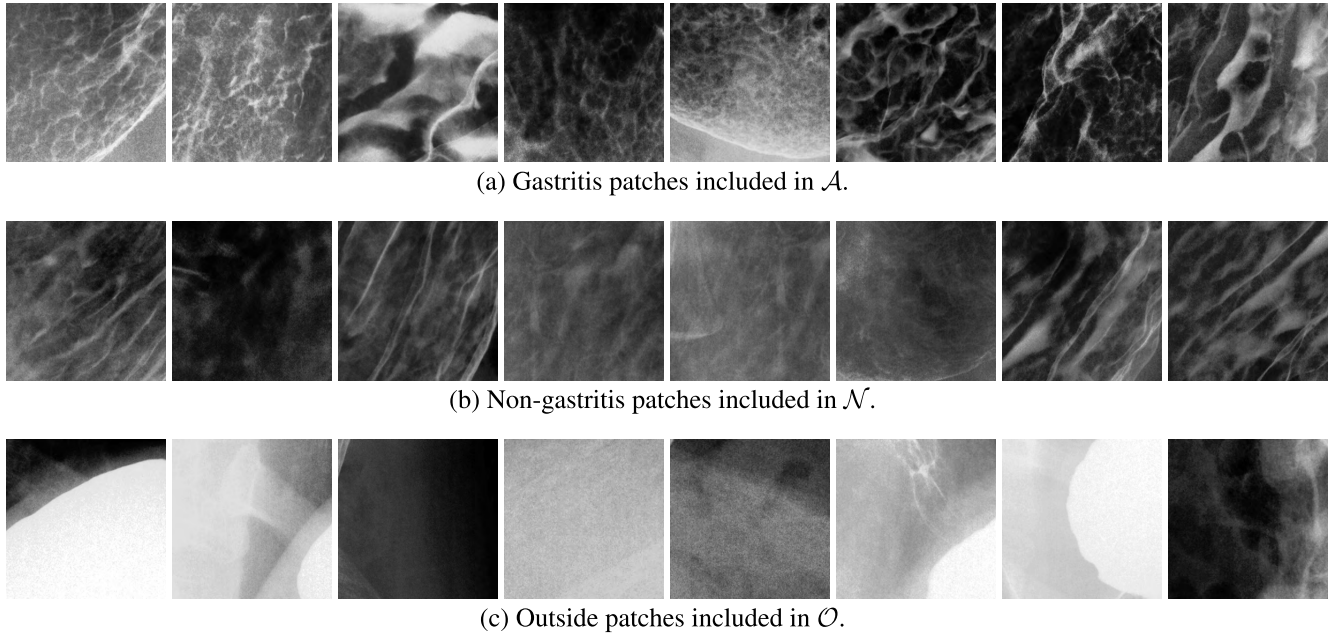


FIGURE 2. Examples of extracted patches (299×299 pixels).

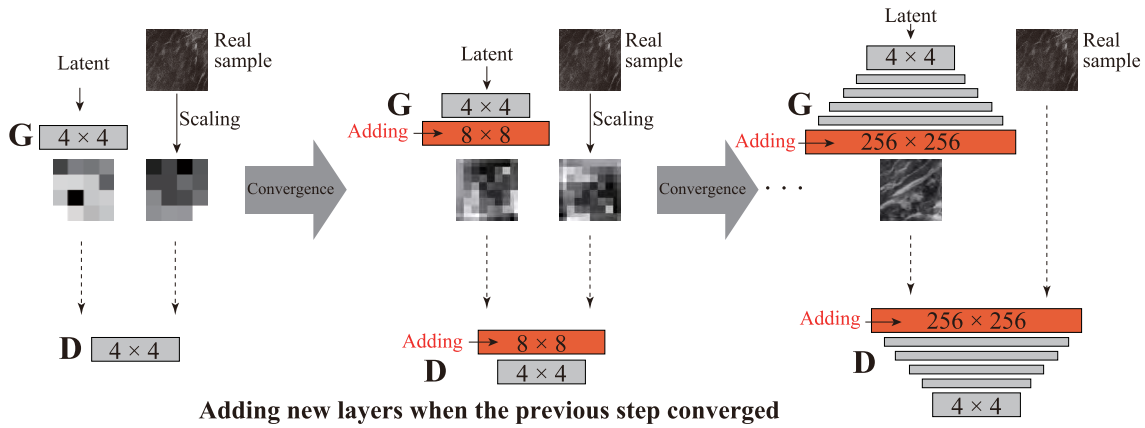


FIGURE 3. Details of the architecture of the progressive growing network. According to the training progression, new layers are added.

- $\mathcal{N}$: data including non-gastritis patches,
- $\mathcal{O}$: patches from outside the stomach.

It should be noted that the region annotation of the stomach in this study was defined manually by a radiological technologist since the accuracy of automated stomach region estimation methods for gastric X-ray images is still insufficient for clinical use [34]. Image level labels are assigned for each gastric X-ray image, and divided patches $X_i^{(h,w)}$ have the same label of F_i . Examples of extracted patches (299×299 pixels) are shown in Fig. 2. The datasets $\mathcal{A}$, $\mathcal{N}$ and $\mathcal{O}$ are used for the following adversarial training procedure.

B. PROGRESSIVE GROWING NETWORK ARCHITECTURE

It is necessary to detect the fine differences between abnormal and normal characteristics when trying to generate

synthetic gastritis/non-gastritis images for a classification task. However, abnormal (gastritis) characteristics in a gastric X-ray image are often only subtly different from normal (non-gastritis) characteristics in gastric X-ray images and are difficult to understand. In order to detect the subtle differences between abnormal and normal images, we employ a progressive growing network architecture motivated by Progressive Growing GAN (PGGAN) [35]. PGGAN is a representative generative model in high-quality image generation tasks. PGGAN's training starts with low resolution images, and then progressively increases the resolution by adding new layers to the generator and the discriminator. The architecture of PGGAN is shown in Fig. 3. In order to stabilize the training processes, some chips (e.g., mini-batch standard deviation, equalized learning rate, and pixel-wise feature vector normalization) are introduced into PGGAN. Employing PGGAN's

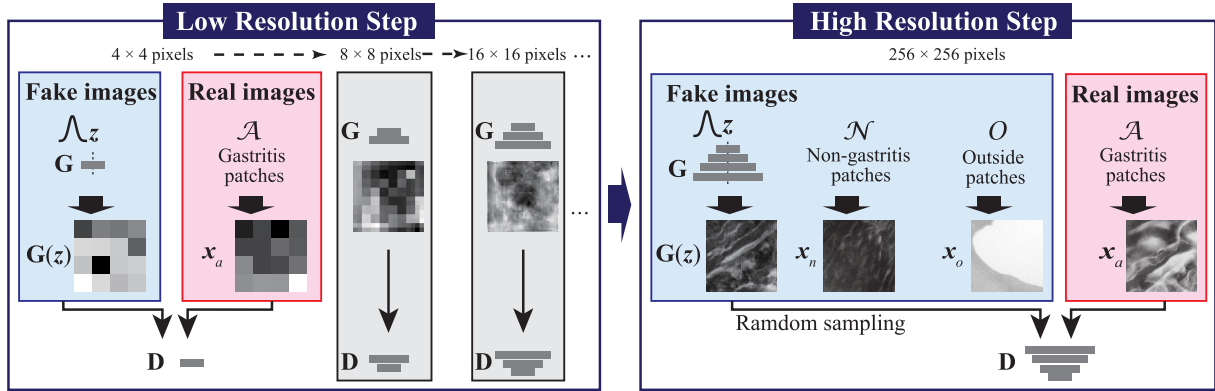


FIGURE 4. Overview of the proposed method. In the low resolution step, a PGGAN-based learning strategy is employed, and in the high resolution step, the conditional adversarial learning is employed.

learning process enables our networks to detect the characteristics of symptoms from coarse to fine.

In LC-PGGAN, we train our networks with a low spatial resolution of 4×4 pixels. The details of our network architecture are shown in Fig. 4. In the low resolution step, LC-PGGAN learns the broad outlines of training images. As the training advances, layers are incrementally added to the generator and the discriminator to reach the high resolution images. In the high resolution step, our LC-PGGAN learns the detailed regions of training images. By adopting these progressive training procedures, the generator can learn the characteristics of gastritis/non-gastritis shown in the training images. Conditional information is also added to generated synthetic images in the high resolution step, and this is another novel point of the proposed method. The conditional adversarial learning is explained in detail in Subsec. III-C.

C. LOSS FUNCTION-BASED CONDITIONAL ADVERSARIAL LEARNING

In training for a typical GAN, a generator learns only a single training distribution. There always exist two pairs of gastritis and non-gastritis images in gastric X-ray inspection. By utilizing this property, generation of synthetic gastritis images can be realized, thus contributing to the improvement of gastritis classification performance.

Let a generator G and a discriminator D have parameters θ_G and θ_D , respectively. LC-PGGAN utilizes three data distributions, $\mathcal{A}$, $\mathcal{N}$ and $\mathcal{O}$, to consider the above-mentioned situations. Then let x_a , x_n , and x_o denote mini-batches of $\mathcal{A}$, $\mathcal{N}$, and $\mathcal{O}$, respectively. Note that the method for the following image generation is a method that tries to generate “gastritis” images. Generation of “non-gastritis” images can be realized by simply replacing the feeding data $\mathcal{A}$ with $\mathcal{N}$.

At the beginning of the training in the low resolution step, the objective function $J_{D,G}$ of LC-PGGAN is

$$J_{D,G}(\theta_D, \theta_G) = \mathcal{L}_A + \mathcal{L}_G, \quad (2)$$

where the loss function $\mathcal{L}_A$ is defined as:

$$\mathcal{L}_A = -\mathbb{E}_{x_a \sim \alpha_{x_a}} [\log D(x_a)], \quad (3)$$

where α_{x_a} represents the prior distribution of $\mathcal{A}$ and x_a represents the mini-batch of $\mathcal{A}$. If the discriminator D can correctly classify x_a as real images, $D(x_a)$ becomes larger, namely, $\mathcal{L}_A$ becomes smaller. Next, the loss function $\mathcal{L}_{G(z)}$ is defined as:

$$\mathcal{L}_G = -\mathbb{E}_{z \sim \alpha_z} [\log(1 - D(G(z)))], \quad (4)$$

where a synthetic image $G(z)$ is generated by feeding a noise vector z that is sampled from a latent distribution α_z to the generator G . The generator G tries to fool the discriminator D , and $\mathcal{L}_{G(z)}$ becomes larger when $G(z)$ passes through the discriminator D . We minimize Eq. (2) by optimizing the two loss functions $\mathcal{L}_A$ and $\mathcal{L}_G$ simultaneously.

In this way, we feed the data $\mathcal{A}$ as training images to the discriminator D , and the generator G learns the outline of gastritis images in the low resolution step. However, some generated synthetic images are still not effective as training data for the classification problem since the generator G learns only a one-class data distribution. To generate more high-quality synthetic gastritis images for enhancing the gastritis classification performance, we feed other data distributions to our adversarial learning, and update our objective function in the high resolution step. In the high resolution step (the 256×256 pixels resolution step), we update our objective function as follows:

$$\hat{J}_{D,G}(\theta_D, \theta_G) = \mathcal{L}_A + \mathcal{L}_G + \lambda(\mathcal{L}_N + \mathcal{L}_O), \quad (5)$$

where λ represents a weight coefficient, and λ is a constant used during the high resolution training step. Note that the loss function $\mathcal{L}_N$ is

$$\mathcal{L}_N = -\mathbb{E}_{x_n \sim \alpha_{x_n}} [\log(1 - D(x_n))], \quad (6)$$

where α_{x_n} represents the prior distribution of the data $\mathcal{N}$. In the same manner as $\mathcal{L}_N$, the loss function $\mathcal{L}_O$ is

$$\mathcal{L}_O = -\mathbb{E}_{x_o \sim \alpha_{x_o}} [\log(1 - D(x_o))], \quad (7)$$

where α_{x_o} represents the prior distribution of the data $\mathcal{O}$. The loss functions $\mathcal{L}_N$ and $\mathcal{L}_O$ provide conditional information to the generator G and the discriminator D . Specifically, the discriminator D should judge a generated “gastritis” image $G(z)$

as a fake image, but “non-gastritis” and “outside” images are judged as fake images by the updated loss function $\hat{J}_{D,G}$. The introduction of constraints on our objective function affects the training of the generator G . In the low resolution step, the generator G focuses on detecting only the outline of the “gastritis” distribution. On the other hand, in the high resolution step, the generator G has to generate more realistic “gastritis” samples since samples similar to “non-gastritis” and “outside” samples are rejected by the trained discriminator D . Consequently, the generation of “abnormal” images for gastritis classification can be realized.

By the training procedures of the progressive growing network architecture and loss function-based conditional adversarial learning of our LC-PGGAN, our generator can produce images that have conditional information. Although images generated by LC-PGGAN have conditional information for gastritis classification, a variety of training data is also an essential element in the classification task. Therefore, we mix the images generated by PGGAN that have rich diversity and the images generated by LC-PGGAN that have conditional information for training of the gastritis classification model.

The most important aspect is that images generated by LC-PGGAN do not have one-to-one correspondence to real images. One of the bottlenecks of medical image analysis is that medical images should be treated with high confidentiality. This problem must be solved to accelerate research on medical image analysis, particularly for minor inspections. Images generated by LC-PGGAN can contribute to a solution of this challenging problem since it can be used as data that do not include individual information.

IV. EXPERIMENT

In this section, we quantitatively and qualitatively evaluate synthetic images generated by LC-PGGAN. Experimental settings are shown in Subsec. IV-A, and quantitative and qualitative evaluation results are shown in Subsec. IV-B.

A. EXPERIMENTAL SETTINGS

As clinical data, 815 (240 gastritis and 575 non-gastritis) patients’ gastric X-ray images were used. The ground truth of gastritis/non-gastritis was determined by endoscopic and X-ray image interpretation results with double-checking by clinicians. Gastric X-ray images were gray-scale and $2,048 \times 2,048$ pixels, and they were divided into multiple patches of 299×299 pixels with a sliding interval of 50 pixels. The sizes of the patches were experimentally determined. In the image generation procedure, these patches were resized for training. We randomly selected 100 gastritis and 100 non-gastritis gastric X-ray images from our original data, and we constructed our training data for image generation. In other words, 200 gastric X-ray images were allocated as training data. The numbers of divided training patches of the data $\mathcal{A}$, $\mathcal{N}$ and $\mathcal{O}$ were 45,127, 42,785 and 48,385, respectively. Synthetic image generation was implemented using data for $\mathcal{A}$, $\mathcal{N}$ and $\mathcal{O}$. The remaining 615 X-ray images

were allocated to test data. Test data were also divided into multiple patches in the same manner as that for training data, namely, 1,225 patches were extracted from each test gastric X-ray image. Estimated labels of gastritis/non-gastritis were determined for each patch, and the final image level estimation result was determined by the simplest majority voting method.

Many evaluation metrics such as Inception score [36], Fréchet Inception distance (FID) [37], and Sliced Wasserstein distance (SWD) [35] have been proposed for evaluation of the quality of generated images. However, these metrics are not suitable for the evaluation of images for classification problems. A classification-based metric, known as GAN-train, for the evaluation of generated images was proposed by Shmelkov *et al.* in 2018 [38]. GAN-train evaluates the classification performance of a classifier trained on generated synthetic images and tests the performance on a set of real images. If an optimal GAN model that perfectly captures the target distribution can generate a set of images, they are indistinguishable from the original training set. Assuming this set has the potential of the same classification performance in GAN-train. Since anonymized generated images were used for gastritis classification in our study, we used GAN-train as our evaluation index.

In the experiment, the support vector machine (SVM) [39] was used as an estimator for the gastritis classification task in GAN-train. In terms of gastritis classification accuracy, a deep learning-based estimator is the first choice. However, such an estimator has many parameters and the classification performance heavily relies on the settings of the parameters. Therefore, we employed the simplest SVM as our estimator to fairly evaluate the effectiveness of generated images. The types of features also affect the classification performance. Hand-crafted features are an old-fashioned approach, and we therefore extracted high-level semantic features from pre-trained deep models, namely, pre-trained VGG-16 [40], Inception-v3 [41] and, ResNet-50 [42] models, in the experiment. Specifically, 4,096-dimensional features were extracted from the fully connected layer (fc_7) of VGG-16, and 2,048-dimensional features were extracted from the pool_3 layer of Inception-3 and the flatten layer of ResNet-50, respectively. The features of generated and real images that were obtained were used for the SVM-based evaluation.

As comparative methods, synthetic images generated by an original PGGAN and a basic deep convolutional GAN (DCGAN) [30] were used. In this GAN-train experiment, we generated 10,000 patches from latent distributions through learned image generation models. From this set, we randomly sampled 5,000 generated positive/negative patches and constructed GAN-train data. Note that the SVM training data “LC-PGGAN + PGGAN” is constructed by randomly sampled from 20,000 (10,000 from LC-PGGAN and 10,000 from PGGAN) patches. Sensitivity (Sen), specificity (Spe), and harmonic mean of Sen and Spe (HM) were utilized for the evaluation. These criteria can be defined as

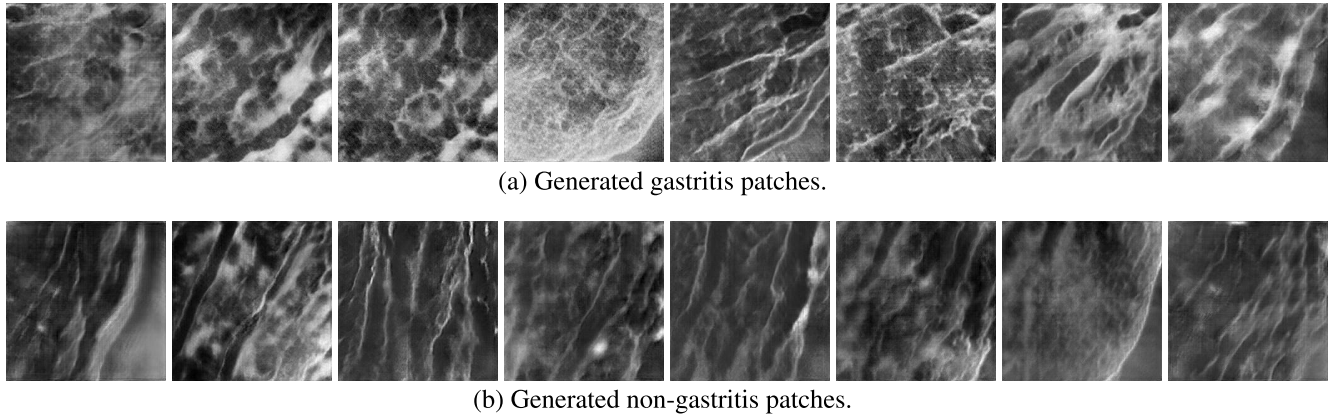


FIGURE 5. Examples of patches generated by LC-PGGAN.

TABLE 1. Classification performance of LC-PGGAN and the comparative methods with VGG-16 features.

Training Data	Sen	Spe	HM
LC-PGGAN + PGGAN	0.779	0.747	0.763
LC-PGGAN	0.850	0.716	0.777
PGGAN	0.807	0.705	0.753
DCGAN	0.571	0.619	0.594
Real data (Ideal)	0.879	0.880	0.879

TABLE 2. Classification performance of LC-PGGAN and the comparative methods with Inception-v3 features.

Training Data	Sen	Spe	HM
LC-PGGAN + PGGAN	0.771	0.912	0.836
LC-PGGAN	0.807	0.720	0.761
PGGAN	0.843	0.806	0.824
DCGAN	0.650	0.651	0.650
Real data (Ideal)	0.943	0.855	0.897

follows:

$$Sen = \frac{TP}{TP + FN}, \quad (8)$$

$$Spe = \frac{TN}{TN + FP}, \quad (9)$$

$$HM = \frac{2 \times Sen \times Spe}{Sen + Spe}, \quad (10)$$

where TP, TN, FP and FN are the numbers of true positive samples, true negative samples, false positive samples and false negative samples, respectively.

In the experiment, it took one week to train our networks with two NVIDIA GeForce GTX 1080Ti GPUs. The mini-batch size of LC-PGGAN and PGGAN was changed depending on the available memory budget (4×4 : 32, 8×8 : 32, 16×16 : 16, 64×64 : 16, 128×128 : 16, and 256×256 : 8). In the comparative methods, PGGAN was trained with the above two GPUs, and DCGAN was trained with a single NVIDIA GeForce GTX 1080Ti GPU.

B. RESULTS AND DISCUSSION

The goal of our method is the generation of realistic synthetic images for sharing and updating clinical data easily. It is expected that anonymized synthetic data will be as effective as real data for classification problems. We compared the gastritis classification performance using synthetic data as quantitative evaluation. Results of the classification performance in GAN-train are shown in Tables 1-3. From the results,

TABLE 3. Classification performance of LC-PGGAN and the comparative methods with ResNet-50 features.

Training Data	Sen	Spe	HM
LC-PGGAN + PGGAN	0.743	0.851	0.793
LC-PGGAN	0.807	0.749	0.777
PGGAN	0.636	0.796	0.707
DCGAN	0.643	0.749	0.692
Real data (Ideal)	0.821	0.916	0.866

we can see that LC-PGGAN outperformed the comparative methods in gastritis classification performance despite the fact that the performance does not outperform when real images were used as training data. Although PGGAN showed a certain level of classification performance, its performance was not as high as the performance of LC-PGGAN. Moreover, collaborative use of LC-PGGAN and PGGAN showed the best performance in Inception-v3 and ResNet-50 features. On the other hand, we can see that the model trained on generated images by DCGAN cannot correctly classify real data. Overall, we confirmed that the progressive growing network architecture is effective for detecting the real data distribution since DCGAN does not have such a network architecture.

Next, we discuss the visual quality of the generated images. Generated patch samples are shown in Figs. 5-7. Each figure includes generated gastritis and non-gastritis sample

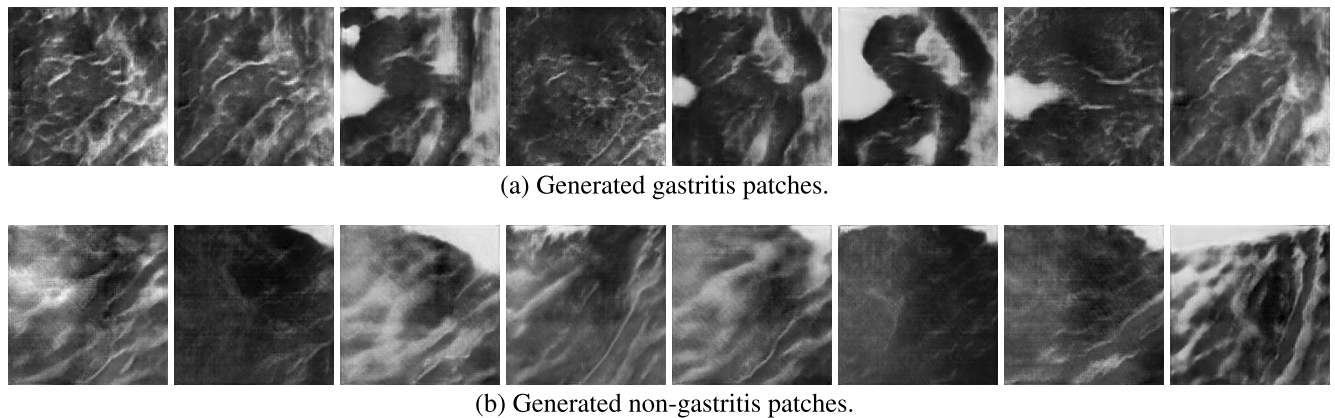


FIGURE 6. Examples of patches generated by PGGAN.

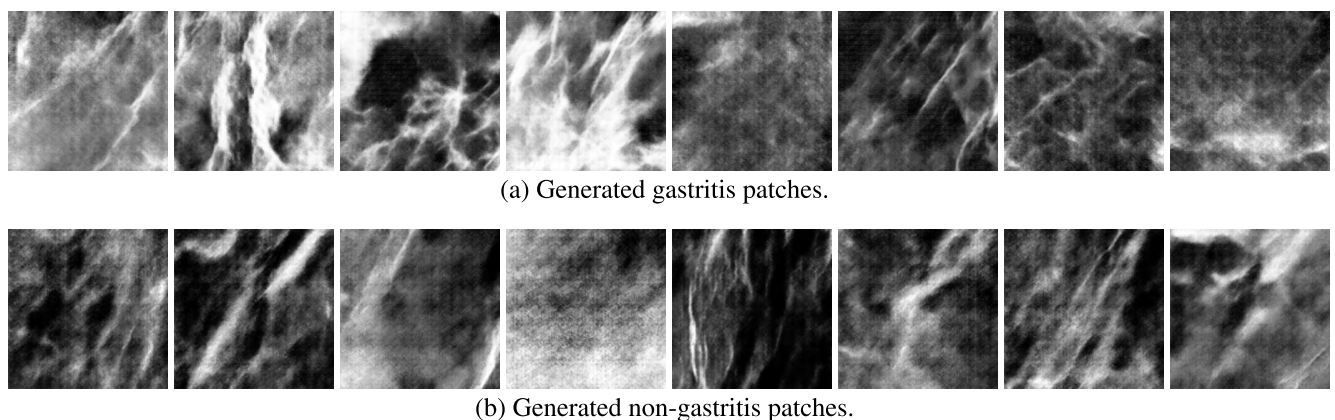


FIGURE 7. Examples of patches generated by DCGAN.

patches. As shown in Fig. 2, characteristics of gastritis are coarse mucosal surface patterns and non-straight folds, and those of non-gastritis are uniform mucosal surface patterns and straight folds. We can see that the images generated by LC-PGGAN can correctly detect these characteristics. In particular, we can see that the images generated by LC-PGGAN (in Fig. 5) have more specific characteristics of gastritis/non-gastritis. Also, the images generated by PGGAN (in Fig. 6) have the same characteristics as those of real data. In addition, we can see that generated images have wide varieties. This may contribute to the improvement in gastritis classification performance. On the other hand, the images generated by DCGAN (in Fig. 7) have some noise.

Generally, GANs easily face the mode collapse problem caused by the meshing of the training progress of a generator and a discriminator. If mode collapse occurs, the learned generator produces only similar images. In the proposed method, we updated our loss function $J_{D,G}$ to $\hat{J}_{D,G}$ during the training. Despite this regularization, LC-PGGAN achieved successful generation of high-quality synthetic images.

This study has some limitations. The classification performance of real gastritis/non-gastritis images in this study is not of a sufficient level for clinical applications. In the

experiment, instead of using deep neural networks that require complicated parameter tuning processes, we used the simplest SVM models as our estimator since we focused on the evaluation of the quality of generated images. We have already achieved high classification performance using real images based on deep learning [33], and this attempt using synthetic images is the next step for data sharing.

V. CONCLUSION

We have presented a synthetic gastritis image generation method with progressive growing adversarial learning, which is a novel high-quality image generation method for realizing sharing and updating of clinical data for machine learning techniques more easily. Besides the fact that our anonymized generated images were useful for gastritis classification, we confirmed our these images had characteristics of gastritis/non-gastritis like real data.

ACKNOWLEDGMENT

The authors would like thank Katsuhiro Mabe of the Hakodate National Hospital and Nobutake Yamamichi of The University of Tokyo. Data obtained from The University of Tokyo Hospital in Japan were used.

REFERENCES

- [1] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," 2016, *arXiv:1605.07678*. [Online]. Available: <https://arxiv.org/abs/1605.07678>
- [2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [3] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.
- [4] X. Zhen, Z. Wang, A. Islam, M. Bhaduri, I. Chan, and S. Li, "Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation," *Med. Image Anal.*, vol. 30, pp. 120–129, May 2016.
- [5] D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, "Segmentation of brain tumor tissues with convolutional neural networks," in *Proc. MICCAI-BRATS*, Sep. 2014, pp. 36–39.
- [6] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.
- [7] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.
- [8] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
- [9] J. Luo, M. Wu, D. Gopukumar, and Y. Zhao, "Big data application in biomedical research and health care: A literature review," *Biomed. Inform. Insights*, vol. 8, Jan. 2016, Art. no. BII531559.
- [10] B. Fabian, T. Ermakova, and P. Junghanns, "Collaborative and secure sharing of healthcare data in multi-clouds," *Inf. Syst.*, vol. 48, pp. 132–150, Mar. 2015.
- [11] J. J. Berman, "Confidentiality issues for medical data miners," *Artif. Intell. Med.*, vol. 26, nos. 1–2, pp. 25–36, Sep./Oct. 2002.
- [12] J. Stadelmann, H. Schulz, U. van der Heide, and S. Renisch, "Pseudo-CT image generation from mDixon MRI images using fully convolutional neural networks," *Proc. SPIE*, vol. 10953, Mar. 2019, Art. no. 109530Z.
- [13] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with context-aware generative adversarial networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Sep. 2017, pp. 417–425.
- [14] P. Costa, A. Galdran, M. I. Meyer, M. D. Abràmoff, M. Niemeijer, A. M. Mendonça, and A. Campilho, "End-to-end adversarial retinal image synthesis," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 781–791, Mar. 2017.
- [15] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.
- [16] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012," *Int. J. Cancer*, vol. 136, no. 5, pp. E359–E386, Mar. 2015.
- [17] H. Ohata, S. Kitauchi, N. Yoshimura, K. Mugitani, M. Iwane, H. Nakamura, A. Yoshikawa, K. Yanaoka, K. Arii, H. Tamai, Y. Shimizu, T. Takeshita, O. Mohara, and M. Ichinose, "Progression of chronic atrophic gastritis associated with *helicobacter pylori* infection increases risk of gastric cancer," *Int. J. Cancer*, vol. 109, no. 1, pp. 138–143, Mar. 2004.
- [18] A. Nashimoto, K. Akazawa, Y. Isobe, I. Miyashiro, H. Katai, Y. Kadera, S. Tsujitani, Y. Seto, H. Furukawa, I. Oda, H. Ono, S. Tanabe, and M. Kaminishi, "Gastric cancer treated in 2002 in Japan: 2009 annual report of the JGCA nationwide registry," *Gastric Cancer*, vol. 16, no. 1, pp. 1–27, Jan. 2013.
- [19] K. Sugano, "Screening of gastric cancer in Asia," *Best Pract. Res. Clin. Gastroenterology*, vol. 29, no. 6, pp. 895–905, Dec. 2015.
- [20] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2014, pp. 2672–2680.
- [21] Y. Bengio, E. Laufer, G. Alain, and J. Yosinski, "Deep generative stochastic networks trainable by backprop," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jan. 2014, pp. 1–9.
- [22] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
- [23] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 1–9.
- [24] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2016, pp. 1–9.
- [25] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
- [26] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Aug. 2017, pp. 1857–1865.
- [27] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Aug. 2017, pp. 2642–2651.
- [28] N. Park, M. Mohammadi, K. Gorde, S. Jajodia, H. Park, and Y. Kim, "Data synthesis based on generative adversarial networks," *Proc. VLDB Endowment*, vol. 11, no. 10, pp. 1071–1083, Jun. 2018.
- [29] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barlett, "Generalization of deep neural networks for chest pathology classification in X-rays using generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2017, pp. 990–994.
- [30] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Jan. 2016, pp. 1–16.
- [31] R. Togo, K. Ishihara, T. Ogawa, and M. Haseyama, "Estimation of salient regions related to chronic gastritis using gastric X-ray images," *Comput. Biol. Med.*, vol. 77, pp. 9–15, Oct. 2016.
- [32] R. Togo, K. Ishihara, K. Mabe, H. Oizumi, T. Ogawa, M. Kato, N. Sakamoto, S. Nakajima, M. Asaka, and M. Haseyama, "Preliminary study of automatic gastric cancer risk classification from photofluorography," *World J. Gastrointestinal Oncol.*, vol. 15, no. 2, pp. 62–70, Feb. 2018.
- [33] R. Togo, N. Yamamichi, K. Mabe, Y. Takahashi, C. Takeuchi, M. Kato, N. Sakamoto, K. Ishihara, T. Ogawa, and M. Haseyama, "Detection of gastritis by a deep convolutional neural network from double-contrast upper gastrointestinal barium X-ray radiography," *J. Gastroenterology*, vol. 54, no. 4, pp. 321–329, Apr. 2019.
- [34] Y. Kita, "Elastic-model driven analysis of several views of a deformable cylindrical object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 12, pp. 1150–1162, Dec. 1996.
- [35] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Feb. 2017, pp. 1–26.
- [36] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2016, pp. 2234–2242.
- [37] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 6626–6637.
- [38] K. Shmelkov, C. Schmid, and K. Alahari, "How good is my GAN?" in *Proc. IEEE Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 213–229.
- [39] C. Cortes, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Apr. 2014, pp. 1–14.
- [41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (ICPR)*, Jun. 2016, pp. 770–778.



REN TOGO (S'16–M'19) received the B.S. degree in health sciences from Hokkaido University, Japan, in 2015, and the M.S. and Ph.D. degrees from the Graduate School of Information Science and Technology, Hokkaido University, Japan, in 2017 and 2019, respectively. He is also Radiological Technologist. His research interest includes machine learning and its applications.



TAKAHIRO OGAWA (S'03–M'08–SM'18) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2003, 2005, and 2007, respectively, where he is currently an Associate Professor with the Faculty of Information Science and Technology. His research interest includes multimedia signal processing and its applications. He has been an Associate Editor of the *ITE Transactions on Media Technology and Applications*.

He is a member of the EURASIP, IEICE, and Institute of Image Information and Television Engineers.



MIKI HASEYAMA (S'88–M'91–SM'06) received the B.S., M.S., and Ph.D. degrees in electronics from Hokkaido University, Japan, in 1986, 1988, and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University, as an Associate Professor, in 1994, where she is currently a Professor with the Faculty of Information Science and Technology. She was a Visiting Associate Professor with Washington University, USA, from 1995 to 1996. Her research interests include image and video processing and its development into semantic analysis. She is a member of the IEICE, ITE, and the Information Processing Society of Japan IPSJ. She has been the Vice-President of the Institute of Image Information and Television Engineers, Japan (ITE). She has been the Editor-in-Chief of the *ITE Transactions on Media Technology and Applications*. She has also been the Director of the International Coordination and Publicity of The Institute of Electronics, Information, and Communication Engineers (IEICE).

...