Multi-Scale Conditional Generative Adversarial Network for Small-Sized Lung Nodules Using Class Activation Region Influence Maximization

KYEONGJIN ANN[1,2], **(Member, IEEE), YEONGGUL JANG**[1,3], **HACKJOON SHIM**[2], **AND HYUK-JAE CHANG**[2,4]

[1]Brain Korea 21 PLUS Project for Medical Science, Yonsei University, Seoul 03722, South Korea
[2]CONNECT-AI Research Center, Yonsei University College of Medicine, Seoul 03721, South Korea
[3]ONtactHeAlth, Seoul 03764, South Korea
[4]Division of Cardiology, Severance Cardiovascular Hospital, Yonsei University College of Medicine, Yonsei University Health System, Seoul 03721, South Korea

Corresponding authors: Hyuk-Jae Chang (hjchang@yuhs.ac) and Yeonggul Jang (ygjang1722@yonsei.ac.kr)

ABSTRACT Automatic detection and classification of thoracic diseases using deep learning algorithms have many applications supporting radiologists' diagnosis and prognosis. However, in the medical field, the class-imbalanced problem is extremely common due to the differences in prevalence among diseases, making it difficult to develop these applications. Many GAN-based methods have been proposed to solve the class-imbalance problem on chest X-ray (CXR) data. However, these models have not been trained well for small-sized diseases because it is challenging to extract sufficient information with only a few pixels. In this paper, we propose a novel deep generative model called a class activation region influence maximization conditional generative adversarial network (CARIM-cGAN). The proposed network can control the target disease's presence, location, and size with a controllable conditional mask. We newly introduced class activation region influence maximization (CARIM) loss to maximize the probability of disease occurrence in the bounded region represented by a conditional mask. To demonstrate an enhanced generative performance, we conducted numerous qualitative and quantitative evaluations with the samples generated using a CARIM-cGAN. The results showed that our method has a better performance than other methods. In conclusion, because the CARIM-cGAN can generate high-quality samples based on information on the location and size of the disease, we can contribute to solving problems such as disease classification, -detection, and -localization, requiring a higher annotation cost.

INDEX TERMS Conditional generative model, medical image augmentation, class activation map, high resolution image, nodule classification and detection, chest X-ray.

I. INTRODUCTION

Building of a large balanced dataset is critical for generalization and convergence in supervised learning. Direct use of a small, highly imbalanced dataset as the training data without additional processes can lead to undesirable outcomes such as overfitting [1], [2]. In the medical field, a class imbalance is pervasive owing to the differences in prevalence among

diseases. For example, on a large-scale chest X-ray (CXR) dataset provided by the National Institutes of Health [3], the number of instances of infiltration, nodule, cardiomegaly, and hernia is 19,894, 6,331, 2,776, and 227, respectively, with a large imbalance ratio of 1:87 between the most and least classes.

Data augmentation techniques such as image transformations have been used to address this challenge of insufficient numbers and imbalances of medical imaging data and help prevent degrading the performance [4], [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Mei Zhang.

The representative, basic data augmentation techniques for transforming images include flipping, scaling, rotation, shearing, and changing the contrast. However, by simply augmenting data in such a way, it is difficult to improve the performance at a large level because the augmented data are also generated from a small number of real datasets and have the same diversity as the real data. Therefore, increasing the diversity of a real dataset is a fundamental solution that allows deep neural networks to learn richer features from the dataset.

The generative adversarial network (GAN), first introduced by Goodfellow *et al.* [6], is the most powerful framework for training a generative model. Many generative models based on GANs have been proposed and have shown a tremendous success in applications of various fields, including computer vision [7]–[9]. These models can generate new plausible synthetic samples, which alleviate the problems of data imbalance and lack of diversity in the real dataset, thereby improving the performance degradation caused by skewed class proportions [10]–[14].

In CXR data augmentation, GAN-based models can produce high-quality synthetic images with high resolutions for large-sized diseases (e.g., effusion, cardiomegaly, and pneumothorax) [15], [16]. However, it is challenging for small-sized diseases (e.g., nodule and mass) to generate a synthetic image that reflects the characteristics of the disease well. Similar to the detection of small objects in an object detection problem, it has a limitation in learning the realistic features of small-sized diseases because they are represented by only a few pixels in the image, which has a minimal contribution to learning [17], [18].

In this paper, we propose a novel deep generative model called a class activation region influence maximization conditional generative adversarial network (CARIM-cGAN) to solve the limitations mentioned above. The proposed network can generate CXR samples with small-sized diseases in the desired form using a controllable conditional mask consisting of binary bits as input. In addition, to preserve characteristics of target disease, we introduce class activation region influence maximization (CARIM) loss newly, which is used in combination with the generator loss during the training phase. The contributions of this study are as follows:

- The proposed network can control the presence, location, and size for the target disease using a controllable conditional mask. The proposed network generates samples of various types (i.e., normal, abnormal samples with different locations/sizes) and consequently contributes to drastically increasing the diversity of the data.
- Our model guarantees generation of CXR samples with a small-sized target disease using a novel CARIM loss that assigns additional restrictions on the bounded area by a given conditional mask. This loss function makes the network concentrate on learning realistic descriptive features for a small-sized disease.
- To demonstrate that the proposed model is superior to other generative models, we conducted experiments

from various perspectives, such as the t-SNE [19], image differencing, histogram, FID [20], PSNR [21], and classification.

II. RELATED WORKS

The deep convolutional GAN (DCGAN) proposed by Radford *et al.* [22] is one of the successful networks designed for image generation, which is composed of convolution layers without max pooling or fully connected layers and uses a convolutional stride and transposed convolution for the down- and up-sampling. Although this model improves learning stability compared to the previous model, it has difficulty in generating high-resolution (256 × 256) images.

The conditional GANs (cGANs) proposed by Mirza and Osindero *et al.* [23] uses conditional variables as input to the generator and discriminator. The representative cGAN-based studies include Pix2Pix [24], CycleGAN [25], DiscoGAN [26], and StarGAN [27], which can generate desirable image styles using a target condition.

The progressive growing of GANs (PGGANs) proposed by Karras *et al.* [28] is an extension to the GAN that enables stable training and high-quality image generation. It's training starts from low-resolution images to high-resolution images by adding new layers to the generator and discriminator progressively.

In [29] and [30], CPGGAN, a study related to cGAN and PGGAN, is described. First, the model proposed by [29] combines a conditional variable (class type) into PGGANs for the task of unsupervised CXR synthesis, which allows for the control of various diseases. It generated high-resolution images well, but this model has a disadvantage in that it cannot finely control the location or size of the disease. Second, the model proposed by [30] incorporates conditional variables (bounding boxes) into PGGANs to control brain metastases at the designated positions/sizes on the magnetic resonance (MR) images. In the case of general cGANs, the discriminator's loss function is implemented as the sum of the unconditional loss that distinguishes real from fake and the conditional loss that encourages the use of conditional information. However, this model does not have a conditional loss in the discriminator, and thus it is possible to generate only tumor samples, not normal cases. Neglecting normal sample generation is inefficient in that all information obtained by the mask is not utilized. Therefore, there is extensibility to cover both cases with and without a disease.

GAN-based models for augmenting CXR datasets have been studied to address the imbalanced classes, a chronic problem in the utilization of medical data. Related studies [15] and [16] trained a DCGAN as a generative model independently of each disease to generate synthetic CXR images for all target diseases. A classification network, based on AlexNet [31], was then trained using both real and synthesized data. The authors showed that increasing the diversity of a dataset with CXR samples generated using a DCGAN helps improve the classification performance and avoid overfitting.
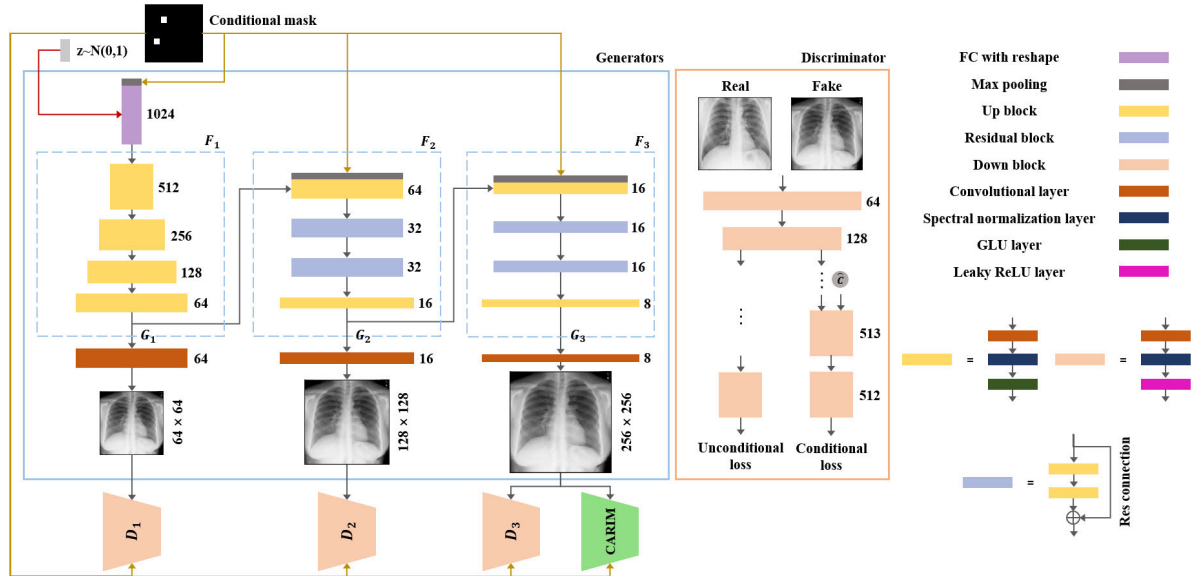
FIGURE 1. The architecture of the proposed CARIM-cGAN for synthetic CXR image generation. Generators *G* gradually produce samples at multiple scales, and discriminators *D* take synthesized samples and real samples at the corresponding scale and evaluate them for authenticity.

As another example, [32] used a CycleGAN to balance the data through CXR data augmentation in the preprocessing step of the binary classification task. CycleGAN is a framework for training image-to-image translation models without paired examples and is usually used to deal with the problem of data unpairing in the image conversion process. CycleGAN in [32] was utilized to generate a paired dataset that refers to a relationship such as fibrosis versus no fibrosis.

However, the above studies show that the classification accuracy was improved for large-sized diseases (e.g., pneumonia, pleural-thickening), but not for small-sized diseases (e.g., fibrosis) after the data augmentation.

Therefore, to address the poor classification performance depending on the size of the disease, it is necessary to build a robust generative model capable of generating CXR datasets that realistically reflect the disease property regardless of its size.

III. METHODOLOGY

This section details the proposed CARIM-cGAN. Our goal is to generate high-resolution synthetic CXR samples with small-sized diseases. In this study, a lung nodule was selected as the target disease among the major thoracic diseases. This nodule is a tiny globular lesion in the lung approximately 3cm in length. Section 3.A introduces the proposed multi-scale cGAN using a controllable conditional mask. And CARIM, a novel method for improving the generative performance for small-sized diseases, is described in section 3.B.

A. MULTI-SCALE CONDITIONAL GENERATIVE ADVERSARIAL NETWORK USING A CONTROLLABLE CONDITIONAL MASK

For synthetic CXR images to be used as training data, they must be generated at high-resolution that can fully reflect the

characteristics of the thoracic structure and organs. We used StackGAN++ [33] as a backbone structure, which is one of the best solutions for generating high-resolution samples. StackGAN++ has multiple generators and discriminators in a tree-shaped structure, gradually learning the distribution of samples at multiple scales. At each branch, the generator produces samples at the corresponding scale, and the discriminator estimates the probabilities of whether the given samples came from training data of that scale. Therefore, the network progressively generates the samples from low to high-resolution through multiple branches during the generation phase.

In this study, we have made four additional modifications to the previous study to achieve our goal. First, the proposed network uses a conditional mask to control the presence, location, and size of the disease. Second, we used the least-square GAN loss [34] to penalize samples according to how far they were from the decision boundary. Third, to eliminate the occurrence of checkerboard artifacts, the nearest-neighbor method [35] was used for all branches as an up-sampling layer. Finally, spectral normalization (SN) [36] was used for generators and discriminators to stabilize the training.

Fig. 1 shows the architecture of the proposed CARIM-cGAN, which takes a 100-dimensional latent vector such that $z \sim N(0, 1)$ and a conditional mask C as inputs and generates samples at multiple scales. The conditional mask is a binary image, where the presence, location, and size of the target disease are controlled by a bounded region represented by 0s and 1s. It is resized to the corresponding scale for each branch and concatenated with the hidden representation of the previous branch to be used as input for the next branch. We compute the hidden representation h_i of the ith branch using the following equation:

$$h_i = F_i (h_{i-1}, C_i), \quad i = 1, \ldots, M, \tag{1}$$

where C_i is the conditional mask for the ith branch, F_i is the convolutional module of the ith branch, and M is the total number of branches. In addition, h_i is computed by F_i using C_i and the hidden representation h_{i-1} from the previous branch. The noise vector z is projected to a spatially extended initial hidden representation h_0 with 1,024 feature maps. Based on these hidden representations, the generators produce synthetic samples at various scales, ranging from low to high, i.e.,

$$s_i = G_i(h_i), \quad i = 1, \ldots, M, \qquad (2)$$

where G_i is the generator of the ith branch, and s_i indicates the sample generated by G_i.

The discriminator D_i takes either the real sample x_i or synthesized sample s_i at the scale of the ith branch and the disease condition as inputs, and learns both conditional and unconditional perspectives. The conditional perspective determines whether the sample and condition match or not, and the unconditional perspective distinguishes whether the sample is real or synthesized.

$$\begin{aligned}
\mathcal{L}_{\mathcal{D}_i} = \frac{1}{2}(&\mathbb{E}_{x_i \sim \mathcal{P}_{data_i}}\left[(\mathcal{D}_i(x_i) - 1)^2\right] \\
&+ \mathbb{E}_{s_i \sim \mathcal{P}_{G_i}}\left[(\mathcal{D}_i(s_i))^2\right] \\
&+ \mathbb{E}_{x_i \sim \mathcal{P}_{data_i}}\left[(\mathcal{D}_i(x_i|\hat{C}) - 1)^2\right] \\
&+ \mathbb{E}_{s_i \sim \mathcal{P}_{G_i}}\left[(\mathcal{D}_i(s_i|\hat{C}))^2\right]),
\end{aligned} \qquad (3)$$

where $\hat{C}$ is the class information obtained from the conditional mask, which is expressed as 0 for a normal state and as 1 for a nodule.

The generator G_i jointly approximates unconditional and conditional distributions of samples at the scale of the ith branch, which is optimized by minimizing the loss $\mathcal{L}_{G_i}$. The generator G_i and discriminator D_i at the ith branch are alternatively trained against each other.

$$\begin{aligned}
\mathcal{L}_{G_i} = \frac{1}{2}\left(\mathbb{E}_{s_i \sim \mathcal{P}_{G_i}}\left[(\mathcal{D}_i(s_i) - 1)^2\right] \right. \\
\left. + \mathbb{E}_{s_i \sim \mathcal{P}_{G_i}}\left[(\mathcal{D}_i(s_i|C_i) - 1)^2\right]\right) \qquad (4)
\end{aligned}$$

B. CLASS ACTIVATION REGION INFLUENCE MAXIMIZATION

The class activation map (CAM) was first introduced by Zhou et al. [37] to visualize what the convolutional neural network (CNN) is looking for in classifying images, highlighting the importance of the image region to the prediction. Selvaraju et al. [38] proposed Grad-CAM, a generalization of the CAM, which compensates for shortcomings such as dependencies on the global average pooling (GAP) layer. This approach defines the class-discriminative localization map M^c for any class c at the feature maps A^k of the kth convolutional layer through the following

equation:

$$M^c = ReLU(\sum_k a_k^c A^k); \quad a_k^c = \frac{1}{wh}\sum_{i=1}^{w}\sum_{j=1}^{h}\frac{\partial y^c}{\partial A_{ij}^k}, \qquad (5)$$

where $\frac{\delta y^c}{\delta A_{ij}^k}$ is the gradient of the score y^c for a target class c with respect to the feature map $A^k \in R^{w \times h}$ of the width w and height h, indexed by i and j, respectively.

The neuron importance weight a_k^c of the feature map k for a target class c is defined as the global-average-pooled gradient over the width and height dimensions. The class-discriminative localization map M^c is computed by performing a linear combination of feature maps and the neuron importance weight and then applying the ReLU activation function to consider the positive influence on the target class. To obtain fine-grained pixel-scale representations, we use up-sampled activation maps.

In this paper, we propose a novel CARIM loss, an extended version of Grad-CAM, to ensure that the disease is generated in the bounded region defined by the conditional mask. The CARIM loss maximizes the values of the CAM in the bounded region through the following equation:

$$\mathcal{L}_{CARIM} = \frac{1}{|\Omega|}\sum_{x \in \Omega}(1 - M_{norm}^c(x))^2, \qquad (6)$$

where $M_{norm}^c(x)$ is the value of the CAM normalized by $\max(M^c)$ at location x; Ω is the bounded region defined by the conditional mask, i.e., the region where the nodule should appear ($x \in \Omega$); and $|\Omega|$ is the number of elements within Ω.

This loss function shifts the network attention in classifying the disease to the bounded region by the conditional mask, making the region significantly impact the classification. Thus, it increases the likelihood of a targeted disease lesion caused by a conditional mask and can help learn realistic descriptive features for small-sized diseases.

Fig. 2 shows the overall process of maximizing the values of the CAM in the bounded region by the conditional mask. Computing the CAM requires additional pre-trained networks for targeted disease classification. In this study, CheXNet [39], a 121-layer densely connected CNN trained on the largest publicly available CXR dataset [3], was selected as the pre-trained classification network. This network calculates the CAM referred to (5) using synthetic samples as the input. The values of the CAM within the bounded region by the given conditional mask are maximized by the loss function referred to (6). As shown in Fig. 1, only samples at the last scale are used for a CARIM loss computation because samples at low-scales do not adequately represent characteristics for small-sized diseases, resulting in unnecessary computational overhead. In conclusion, the final loss function for the generator is expressed by the weighted sum of the multi-scale generative loss and the CARIM loss,
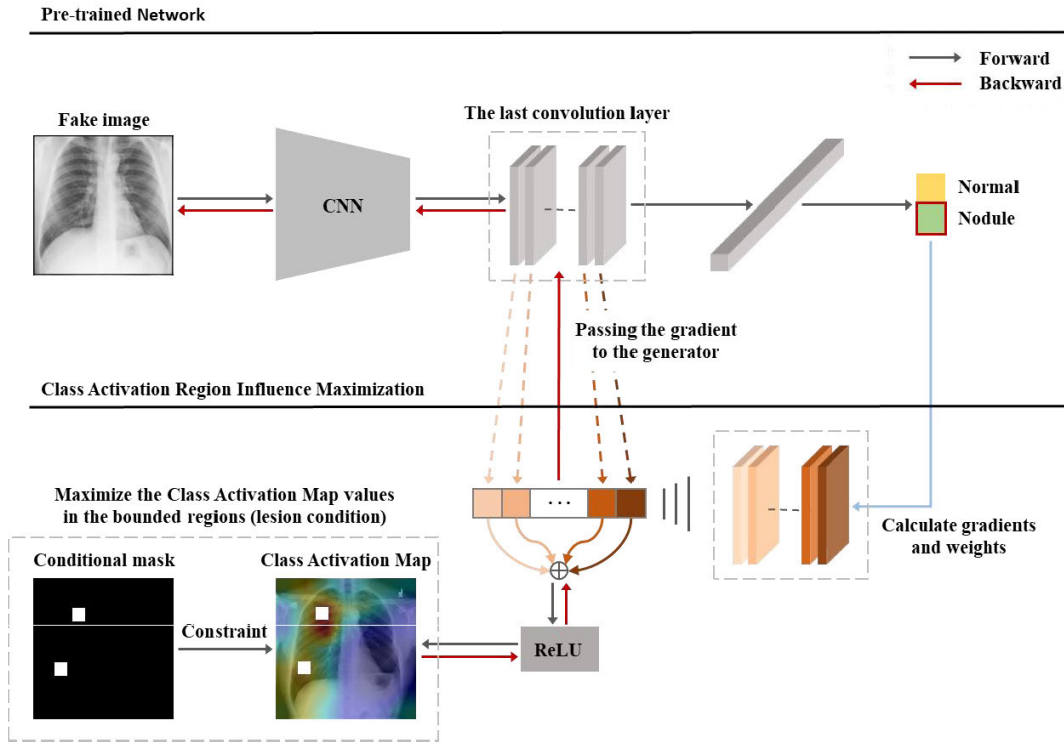
FIGURE 2. The overall process of maximizing the class activation map (CAM) values under a given conditional mask with the pre-trained classification network. In the forward propagation (gray arrows), the CAM of a synthesized image is obtained using pre-trained CheXNet. Then class activation region influence maximization (CARIM) loss is calculated under the constraint of the bounded regions by the conditional mask. In the backward propagation (red arrows), the gradient by CARIM loss is passed to the generator through convolutional layers of the pre-trained classification network.

as referred to (7), which makes it possible to learn a network in an end-to-end manner.

$$\mathcal{L}_G = \frac{1}{2} \sum_{i=1}^{m} \left(\mathbb{E}_{s_i \sim \mathcal{P}_{G_i}} \left[(\mathcal{D}_i(s_i) - 1)^2 \right] \right.$$
$$\left. + \mathbb{E}_{s_i \sim \mathcal{P}_{G_i}} \left[(\mathcal{D}_i(s_i|C) - 1)^2 \right] \right) + \lambda \mathcal{L}_{CARIM}, \quad (7)$$

where λ is the weighting parameter for the CARIM loss, and the optimal value for λ is determined empirically as 0.001.

As a result, the network generates high-resolution samples that depict the details of the thoracic structure by learning the data distributions at multiple scales. In addition, it ensures the generation of a disease with a rich representation in the bounded region by a given conditional mask.

IV. EXPERIMENTS

This part introduces evaluations to demonstrate the performance of the CARIM-cGAN. In section 4.C, for evaluating our method's capacity qualitatively, we conducted the following experiments. First, to prove that our model is better than previous state-of-the-art methods, we visualized the samples generated by each model. In addition, to verify that the synthetic samples were generated close to the real samples, we visualized the distribution of real and synthetic samples using t-distributed stochastic neighbor embedding (t-SNE). Finally, to show that the synthetic sample matches the conditional mask, we experimented with samples

generated by CARIM-cGAN utilizing the image differencing and histogram technique. In section 4.D, for interpreting the performance of the models quantitatively, we calculated the Fréchet Inception Distance (FID) and Peak Signal to Noise Ratio (PSNR), which are an indicator to assess the quality of the generated data, for all comparative models, and trained the classification model based on DenseNet [40].

A. DATASET

We used the ChestX-ray14 dataset, a large scale CXR dataset released by [3]. It comprises 112,120 frontal-view X-ray samples of 30,805 unique patients with labels for 14 diseases (where each sample can have multiple labels) from the associated radiological reports. In the database, only 51,759 samples contain one or more pathologies, and the remaining 60,361 samples are normal cases. There are 6,331 nodule samples among the samples containing pathologies. Specifically, the 79 nodule samples have ground-truth bounding boxes marked with the location of the disease.

We also used Grad-CAM to obtain nodule localization maps for real nodule samples from CheXNet, a classification model pretrained with many nodule samples. The pre-trained model can perform weakly supervised rough localization of a nodule and its classification, and extracted localization maps were used as the prior position distribution for a conditional mask. Therefore, it is possible to obtain a conditional mask
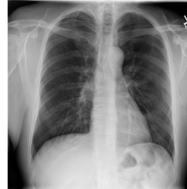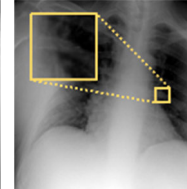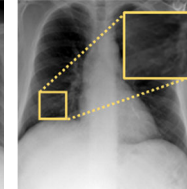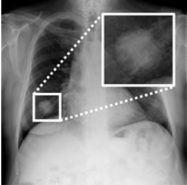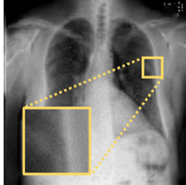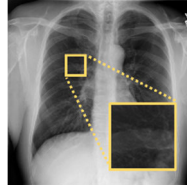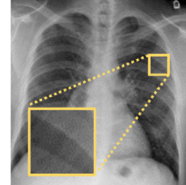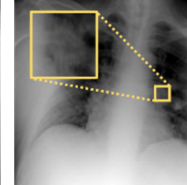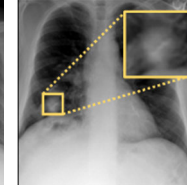
Model	-	DCGAN	PGGAN	Multi-scale conditional GAN		
Method	Real samples	[15], [16], [42]	[29]	[43]	CARIM-cGAN (w/o CARIM Loss)	**CARIM-cGAN (Proposed)**
Normal						
Nodule						

FIGURE 3. Comparison of synthetic samples between other state-of-the-art methods and the proposed method. The white and yellow boxes indicate the ground truth of nodule lesion for real samples and the given conditional mask, respectively. The synthetic normal and nodule sample from each model were generated with the same z noise but a different condition. The samples generated by [15], [16], [42], [29], and [43] are difficult to extract nodule regions because they do not use a conditional mask as input. Therefore, to provide reliable regions of interest, the nodule regions in these samples are marked by one expert radiologist with three years of experience. If there is no nodule region, we obtained the nodule regions by extracting the heatmaps from these samples through the CheXNet and specifying parts with high values on these heatmaps.

that realistically reflects the position of the nodule from the prior position distribution, which is more effective than generating the conditional mask randomly.

B. IMPLEMENTATION DETAILS

We adopted a two-step learning strategy to train the proposed model effectively. In the first step, the generation model is trained without the CARIM loss, and in the second step, the CARIM loss is included in the training. This learning strategy has the following advantages.

First, it can increase the stability of the model. This model learns the image's overall structure in the first step and detailed descriptive features in the second step. It helps the model clarify what needs to be done at each step and consequently contributes to convergence and stability of GAN training.

Second, it can reduce unnecessary computation overhead. It is ineffective to apply CARIM loss from the scratch because there is not enough information to learn early in the learning process. Therefore, applying CARIM loss after the model has sufficient feature representation is cost-effective.

In more detail, in the first step, the proposed model was trained for normal and nodule CXR samples using a mini-batch size of 24. The number of training epochs was set to 43. Also, to update the network weights, Adam optimization [41] with a learning rate of 2×10^{-5} was used. After training, we saved the model that achieved the best generative performance. In the second step, the trained model was then subsequently fine-tuned with CARIM loss. For this transfer learning, the mini-batch size was set to 16, and the number of training epochs was set to 13. The optimization method was the same as in the first step.

C. QUALITATIVE EVALUATIONS

1) COMPARISON OF CARIM-CGAN WITH STATE-OF-THE-ART METHODS

Fig. 3 shows the normal and nodule samples obtained from [15], [16], [42], [29], [43], a CARIM-cGAN without a CARIM loss, and a CARIM-cGAN (from left to right). All models in Fig. 3 were trained using the same ChestX-ray14 dataset. Here, nodules appear as round, white shadows with a diameter of 3cm or less on a CXR.

As indicated in Fig. 3, in [15], [16], and [42], it is difficult to interpret the state of the major organs due to blurred regions of the sample and an unclear shape of the ribs or lung bronchi. In addition, the study in [29] and [43] generally have well-modeled chest characteristics, but the nodule is invisible in the synthetic sample. Thus, we can interpret that [15], [16], [42], [29], and [43] have a significantly lower quality than the real data and the samples generated by the proposed generative model. By contrast, the samples generated from a CARIM-cGAN without a CARIM loss and a CARIM-cGAN have a substantial quality similar to real data. However, in the nodule case, the target disease in the sample generated by a CARIM-cGAN looks more like a real nodule than that of CARIM-cGAN without a CARIM loss. In conclusion, these results demonstrate that the proposed model can properly produce data including structural features as well as detailed elements.

2) COMPARISON OF SYNTHETIC SAMPLES BY A CARIM-CGAN WITH REAL SAMPLES: VISUALIZATION USING T-SNE

The t-SNE is a nonlinear method for a dimensionality reduction, which models each high-dimensional object by a point in
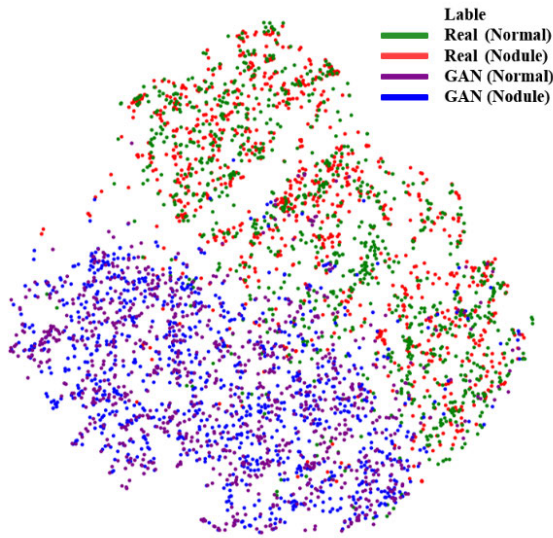
FIGURE 4. t-SNE results with 1,000 256 × 256 samples per category: (a) real normal samples, (b) real nodule samples, (c) synthetic normal samples by CARIM-cGAN, and (d) synthetic nodule samples by CARIM-cGAN.

FIGURE 5. t-SNE results with 79 or 1,000 50 × 50 samples per category: (a) 1,000 real normal samples resized randomly in the lung, (b) 79 real nodule samples resized based on the ground-truth, (c) 1,000 synthetic normal samples resized based on the conditional mask by CARIM-cGAN, and (d) 1,000 synthetic nodule samples resized based on the conditional mask by CARIM-cGAN.

a low-dimensional map by reflecting the degree of similarity between objects. We visually analyze the distributions of real and synthetic datasets with the following samples: (*i*) 1,000 real normal CXR samples, (*ii*) 79 real nodule CXR samples, and (*iii-iv*) 1,000 synthetic normal and nodule CXR samples by a CARIM-cGAN.

As shown in Fig. 4, real and synthetic CXR samples are distributed differently owing to the strong anatomical consistency of real CXR samples even though they belong to the same class group. In other words, samples from two different classes have similar distributions for real and synthetic datasets. Since the nodule is represented by only a few pixels in the CXR image and its influence is relatively low, it is not easy to obtain a distinct distribution between samples of two class groups using t-SNE. However, the distribution is widespread with real and synthetic samples separated, indicating the increased data diversity.

To address the above limitation, in Fig. 5, we visualize the local feature distributions using cropped samples based on the disease region. The distributions for the synthetic normal and nodule samples are generated close to those of the real normal and nodule samples, respectively, and the real and synthetic normal samples have a distinct distribution from the real and synthetic nodule samples. These results indicate that the proposed model generates synthetic CXR images that reflect the intrinsic characteristics of the two different class groups and increases data diversity.

3) DIFFERENCE MAP AND CLASS ACTIVATION MAP HISTOGRAM
To demonstrate that the disease is generated only in the bounded region by a given conditional mask, we visualized a difference map between synthetic normal and nodule samples
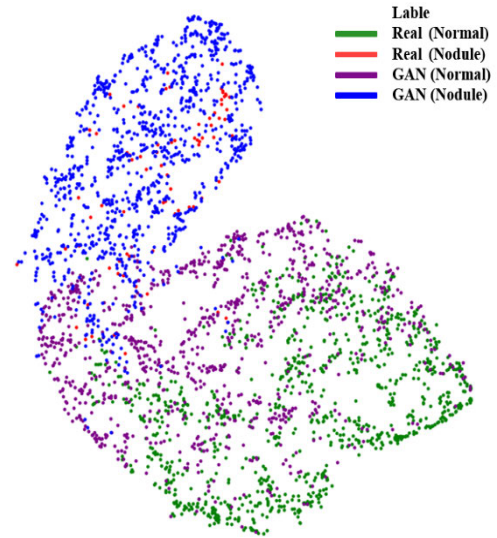
according to the conditional mask setting, and class activation map distributions between the bounded regions and others.

For calculating the difference map, normal and nodule samples are generated with the conditional mask filled with 0s and the conditional mask including bounded regions represented by 1s, respectively, using the same z noise. Fig. 6 shows that the brightest part of the difference map precisely matches the bounded regions by the given conditional mask. It means that the proposed model generates CXR samples by reflecting the presence, size, and location information of the target disease using conditional masks.

Also, we obtained class activation maps for 2,000 synthetic nodule samples from CheXNet using Grad-CAM. We then plotted the histograms of map values inside and outside the bounded regions by the conditional masks. Fig. 7 shows that the histogram of class activation map values inside the bounded regions is closer to the maximum value (min-max normalization) than that of outer regions.

A paired t-test is performed to analyze the difference between pixels inside and outside the bounding boxes obtained in the same heatmap. There is a statistically significant difference between the two groups (p-value < 0.001), which means that the target disease is generated within the bounded region by the conditional mask.

D. QUANTITATIVE EVALUATIONS
1) FRÉCHET INCEPTION DISTANCE (FID)
The FID is a measure used to calculate the similarity, i.e., the Fréchet distance between two multivariate gaussians with mean μ and covariance Σ, between real and synthetic samples defined as the following equation:

$$FID = \|\mu_x - \mu_s\|_2^2 + Tr\left(\Sigma_x + \Sigma_s - 2\left(\Sigma_x \Sigma_s\right)^{\frac{1}{2}}\right), \quad (8)$$
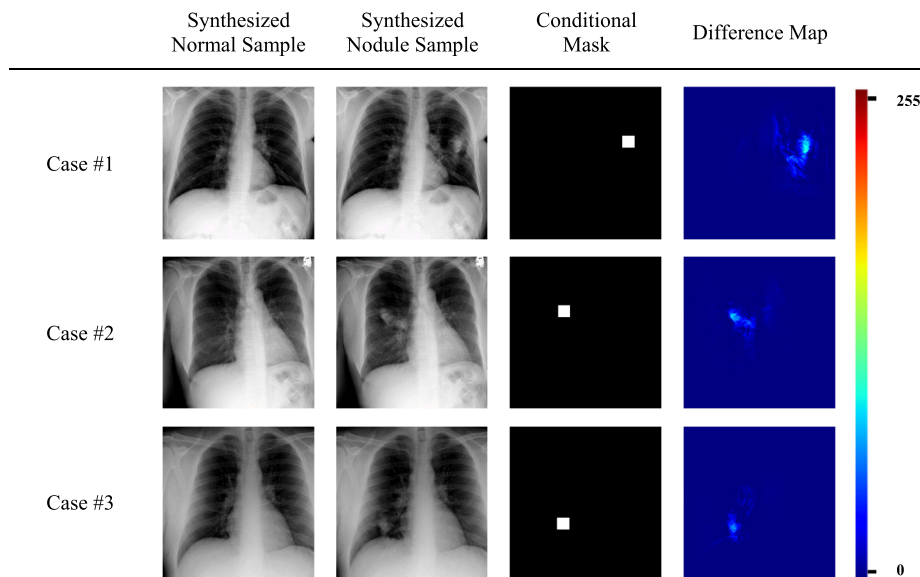
FIGURE 6. Example difference maps between the synthetic normal samples and the nodule samples which are generated by CARIM-cGAN with the same z noise but a different condition. In all cases, the bounded region by the conditional mask and the region with the largest difference in the difference map match each other.
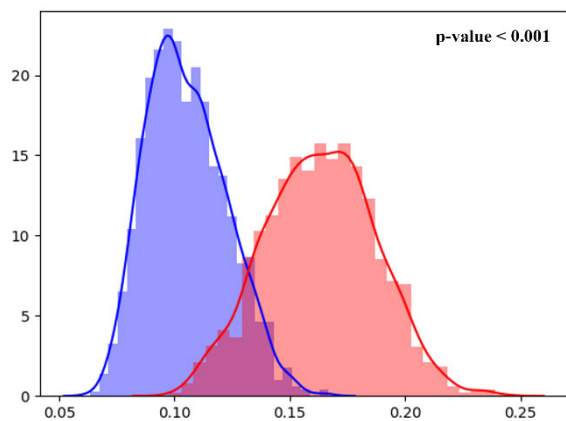

FIGURE 7. The histograms of class activation map values inside (red) and outside (blue) the bounded regions by conditional masks. The x-axis of the histogram refers to the pixel values ranging from 0.00 to 0.255, and the y-axis refers to the density.

where x is the real dataset, and s is the synthetic samples. The lower FID score means that the synthetic samples are high quality and resemble real samples.

To compare the performance among the generative models quantitatively, we calculated the FID score as a performance indicator using the following samples: (*i-ii*) 1,000 real normal and nodule CXR samples and (*iii-iv*) 1,000 synthetic normal and nodule CXR samples. Table 1 shows the FID scores along with the properties for the comparative models. In this experiment, the FID score of the proposed model was about 25% lower than that of [15], [16], [42] and was about 46% lower than that of [43]. In addition, the FID score of the proposed model was about 7% lower than that of [29] for nodules and higher than that of [29] for normal. As an ablation study, the application of the CARIM loss resulted in an additional 14% performance improvement over the case where it was not.

In conclusion, these results suggest that the proposed model has better generative performance, which is comparable to state-of-the-art results for normal and better for small-sized diseases.

2) PEAK SIGNAL TO NOISE RATIO (PSNR)
We calculated an image quality metric, the peak signal-to-noise ratio (PSNR), for quantifying the power of a signal and corrupting noise.

To calculate the PSNR, we prepared the following samples: (*i-ii*) 1,000 uncropped real and synthetic normal samples, (*iii*) 79 real nodule samples cropped based on ground truth location, (*iv*) 1,000 synthetic nodule samples cropped to include nodule region.

Unlike the proposed method, because synthetic nodule samples by other techniques [15], [16], [42], [29], and [43] that do not use a conditional mask as input have no information on where the nodule should occur, it is not easy to extract nodule regions. Therefore, one expert radiologist with three years of experience cropped the nodule regions in these samples. If there are no nodule regions, nodule regions are extracted at an arbitrary location inside the lung to make cropped samples.

As presented in Table 1, we demonstrated that our method achieves better performance in the PSNR than other comparative models. As a result, it means that the synthetic samples by the proposed model look more realistic the real than other models in signal properties.

3) CLASSIFICATION PERFORMANCE
To verify the classification performance improvement by the proposed method, we trained DenseNet-121 as a

TABLE 1. Fréchet inception distance (FID) score and peak signal to noise ratio (PSNR) results for [15], [16], [42], [29], [43], a CARIM-cGAN without a CARIM loss, and a CARIM-cGAN.

Model	Method	Condition	FID		PSNR [db]	
			Normal	Nodule	Normal	Nodule
DCGAN	[15], [16], [42]	-	70.9023	73.8846	58.3316	54.1645
PGGAN	[29]	Class	**49.5071**	57.5197	64.5207	65.2349
Multi-scale conditional GAN	[43]	Class	103.1294	100.0190	61.2435	58.1423
	CARIM-cGAN (w/o CARIM loss)	Mask	64.9082	62.2617	62.9645	60.5120
	CARIM-cGAN (Proposed)	**Mask**	55.2935	**53.1294**	**67.5231**	**66.4537**

TABLE 2. Accuracy, sensitivity, specificity, precision, F1 score, and AUC scores for nodule classification on DS1, DS2, DS3, and DS4.

	Imbalanced Dataset		Balanced Dataset	
	DS1	DS2	DS3	DS4
Accuracy	0.9100	0.6034	0.6113	**0.6208**
Sensitivity	0.9107	0.5780	0.5807	**0.5898**
Specificity	0.0000	0.6533	0.6793	**0.6843**
Precision	0.9999	0.7661	**0.8009**	0.7930
F1 score	0.9532	0.6589	0.6733	**0.6765**
AUC	0.4553	0.6157	0.6300	**0.6371**

classification model using four combinations of real and synthetic CXR samples as follows:

- **Imbalanced real dataset (DS1):** This is an extremely imbalanced dataset. Of all of the available real CXR samples, the normal condition has 59,728 samples, while the nodule has only 5,698 samples.
- **Balanced real dataset (DS2):** A balanced version of the real dataset is set to the minimum number of available real CXR samples. In this case, because the nodule dataset is smaller than the normal dataset, 5,698 nodule samples and only 5,698 randomly extracted normal samples were used to build this dataset.
- **Balanced real dataset augmented with synthesized CXR using CARIM-cGAN without CARIM loss (DS3):** This dataset is composed by adding 2,000 samples generated from a CARIM-cGAN without a CARIM loss to DS2 per class.
- **Balanced real dataset augmented with synthesized CXR using CARIM-cGAN (DS4):** This dataset is composed by adding 2,000 samples generated from the proposed CARIM-cGAN to DS2 per class.

A confusion matrix is a performance measurement method for classification. From the confusion matrix, true

positive (T_p), true negative (T_n), false positive (F_p), and false negative (F_n) values are obtained, and six metrics (e.g., accuracy, sensitivity, specificity, precision, recall, and F1-score) are calculated for the performance evaluation using the following formulas: accuracy $= (T_n + T_p)/(T_n + T_p + F_n + F_p)$, sensitivity $= T_p/(T_p + F_n)$, specificity $= T_n/(T_n + F_p)$, precision defined as $P = T_p/(T_p + F_p)$, and recall defined as $R = T_p/(T_p + F_n)$. The weighted average of the precision and recall is called the F1-score, which is defined as $f_1 = (2 \cdot P \cdot R)/(P + R)$. In this section, we only use five metrics, excluding the recall, and additionally include the area under the curve (AUC), which is another measure for evaluating the ability of the classifier. We used 1,266 real samples (633 normal samples and 633 nodule samples) as a test dataset for evaluation and calculated five metrics and AUC scores on the test dataset for models trained with DS1, DS2, DS3, and DS4, respectively, as presented in Table 2.

The performances for imbalanced dataset DS1 are unreliable despite being generally higher than those of the other datasets, except for the specificity and AUC. In the case of binary classification with an imbalanced dataset, it is a common problem that a minority group containing significantly fewer samples is biased toward the majority group. The bias toward the majority class can be alleviated by

down-sampling the dataset containing more samples among the two classes. Thus, training with a balanced dataset DS2 can yield reliable results. DS3 and DS4 are the augmented datasets with DS2 and synthetic samples, and the performance results for DS3 and DS4 are on average 2% higher than those of the real balanced dataset DS2 for all measures. In addition, the classification model with DS4 achieves a better performance than DS3 for all metrics except the precision.

The results demonstrated that further use of synthesized samples achieves a better classification performance than with only real samples, and CARIM loss helps improve the generative performance effectively.

V. DISCUSSION

This paper proposes a high-resolution image generative model for small-sized diseases using a CARIM loss function specializing in maximizing the probability of target diseases. The proposed CARIM loss shifts the classification network attention to the bounded regions by a conditional mask when classifying diseases, allowing those regions to influence classification significantly. Therefore, it may increase the likelihood of target disease lesions by the conditional mask and help learn realistic descriptive characteristics for small-sized diseases.

We performed qualitative and quantitative evaluations to demonstrate our model's capacity. The difference map and histogram showed that the nodules on the synthetic image generated by the CARIM loss appear only in the bounded regions defined by the conditional mask. In t-SNE visualization, synthetic samples by the proposed model have similar distributions to real samples. FID, and PSNR experiments verified that our model has better performance than other competitors in nodule cases. On the other hand, the PGGAN-based model [29], trained with more data for normal and multiple diseases together, shows better performance in normal cases. For this reason, the proposed model can be improved with more samples. Based on these results, we can conclude that the proposed model can accurately and realistically generate the disease at the desired location using the conditional mask, which other models do not use, and has a similar or better generative quality than the recent studies.

In this paper, CARIM-cGAN is implemented with Grad-CAM [38], known as the general version of CAM, but other improved variants [44]–[50] of Grad-CAM have been recently proposed. Although these variations have not been covered in this paper, they can obtain weights for each feature channel, are differentiable. Therefore, they can apply to the CARIM loss function so that the proposed model learns more fine-grained features and provides optimal interpretation for lung nodules. In addition, the state-of-the-art generative baseline models such as StyleGAN [51] could separate high-level attributes (e.g., pose, disease type) and stochastic variations (e.g., lung, heart, ribs) in the generated images. As a result, the use of these techniques might make it possible to intuitively scale-specifically control, resulting in the generation of more diverse and high-quality samples with small-sized diseases. In future research, we plan to expand this study by applying better techniques such as improved variants of Grad-CAM or StyleGAN.

VI. CONCLUSION

Many GAN-based generative methods have been developed to address the need for a balanced large-scale medical dataset. However, it is still extremely challenging to accurately generate CXR data because the amount of extracted meaningful information for training depends on the size of the disease. In this paper, to solve the limitations on small-sized diseases, we proposed a CARIM-cGAN. The proposed model uses a controllable conditional mask as an input condition to generate synthetic CXR samples with the target disease in the desired location and size. In addition, to preserve characteristics of target disease, we introduce class activation region influence maximization (CARIM) loss newly, which makes the network concentrate on learning realistic descriptive features for a small-sized disease. As a result, our model can generate high-quality, realistic synthetic CXR samples and contributes to minimizing the annotation efforts of expert physicians through controllable conditions on a conditional mask, such as the presence, location, and size of the target disease.

REFERENCES

[1] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018.

[2] Z. Li, K. Kamnitsas, and B. Glocker, "Overfitting of neural nets under class imbalance: Analysis and improvements for segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Oct. 2019, pp. 402–410.

[3] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.

[4] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. Int. Interdiscipl. PhD Workshop (IIPhDW)*, May 2018, pp. 117–122.

[5] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 27, 2014, pp. 2672–2680.

[7] H. Alqahtani, M. Kavakli-Thorne, and G. Kumar, "Applications of generative adversarial networks (GANs): An updated review," *Arch. Comput. Methods Eng.*, vol. 28, no. 2, pp. 525–552, 2021.

[8] M. Uřičář, P. Křížek, D. Hurych, I. Sobh, S. Yogamani, and P. Denny, "Yes, we GAN: Applying adversarial techniques for autonomous driving," *Electron. Imag.*, vol. 2019, no. 15, pp. 48-1–48-17, Jan. 2019.

[9] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," 2020, *arXiv:2001.06937*. [Online]. Available: http://arxiv.org/abs/2001.06937

[10] S. A. Israel, J. H. Goldstein, J. S. Klein, J. Talamonti, F. Tanner, S. Zabel, P. A. Sallee, and L. McCoy, "Generative adversarial networks for classification," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2017, pp. 1–4.

[11] U. Hwang, D. Jung, and S. Yoon, "HexaGAN: Generative adversarial nets for real world classification," *Proc. Mach. Learn. Res.*, vol. 97, pp. 2921–2930, May 2019.

[12] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using GAN for improved liver lesion classification," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 289–293.

[13] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[14] C. Han, Y. Kitamura, A. Kudo, A. Ichinose, L. Rundo, Y. Furukawa, K. Umemoto, Y. Li, and H. Nakayama, "Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2019, pp. 729–737.

[15] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "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. 2018, pp. 990–994.

[16] H. Salehinejad, E. Colak, T. Dowdell, J. Barfett, and S. Valaee, "Synthesizing chest X-ray pathology for training deep convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1197–1206, May 2019.

[17] C. Deng, M. Wang, L. Liu, and Y. Liu, "Extended feature pyramid network for small object detection," 2020, *arXiv:2003.07021*. [Online]. Available: http://arxiv.org/abs/2003.07021

[18] K. Tong, Y. Wu, and F. Zhou, "Recent advances in small object detection based on deep learning: A review," *Image Vis. Comput.*, vol. 97, May 2020, Art. no. 103910.

[19] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

[20] 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. Adv. Neural. Inf. Process. Syst. (NIPS)*, 2017, pp. 1–12.

[21] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2010, pp. 2366–2369.

[22] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[23] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784

[24] 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.

[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," *Proc. Mach. Learn. Res.*, vol. 70, pp. 1857–1865, Jul. 2017.

[27] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.

[28] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*. [Online]. Available: http://arxiv.org/abs/1710.10196

[29] B. Segal, D. M. Rubin, G. Rubin, and A. Pantanowitz, "Evaluating the clinical realism of synthetic chest X-rays generated using progressively growing GANs," *Social Netw. Comput. Sci.*, vol. 2, no. 4, pp. 1–17, Jul. 2021.

[30] C. Han, K. Murao, T. Noguchi, Y. Kawata, F. Uchiyama, L. Rundo, H. Nakayama, and S. Satoh, "Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 119–127.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[32] T. Malygina, E. Ericheva, and I. Drokin, "Data augmentation with GAN: Improving chest X-ray pathologies prediction on class-imbalanced cases," in *Proc. Int. Conf. Anal. Image Social Netw. Texts. (AIST)*, Jul. 2019, pp. 321–334.

[33] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "StackGAN++: Realistic image synthesis with stacked generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1947–1962, Aug. 2019.

[34] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2794–2802.

[35] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, p. e3, 2016.

[36] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," 2018, *arXiv:1802.05957*. [Online]. Available: http://arxiv.org/abs/1802.05957

[37] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.

[38] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[39] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, *arXiv:1711.05225*. [Online]. Available: http://arxiv.org/abs/1711.05225

[40] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[42] S. K. Venu and S. Ravula, "Evaluation of deep convolutional generative adversarial networks for data augmentation of chest X-ray images," *Future Internet*, vol. 13, no. 1, p. 8, Dec. 2020.

[43] K. Ann, Y. Jang, S. Ha, B. Jeon, Y. Hong, H. Shim, and H.-J. Chang, "Generation of high-resolution chest X-rays using multi-scale conditional generative adversarial network with attention," *J. Broadcast Eng.*, vol. 25, no. 1, pp. 1–12, 2020.

[44] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 839–847.

[45] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 24–25.

[46] S. Desai and H. G. Ramaswamy, "Ablation-CAM: Visual explanations for deep convolutional network via gradient-free localization," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 983–991.

[47] R. Fu, Q. Hu, X. Dong, Y. Guo, Y. Gao, and B. Li, "Axiom-based grad-CAM: Towards accurate visualization and explanation of CNNs," 2020, *arXiv:2008.02312*. [Online]. Available: http://arxiv.org/abs/2008.02312

[48] M. B. Muhammad and M. Yeasin, "Eigen-CAM: Class activation map using principal components," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7.

[49] P.-T. Jiang, C.-B. Zhang, Q. Hou, M.-M. Cheng, and Y. Wei, "LayerCAM: Exploring hierarchical class activation maps for localization," *IEEE Trans. Image Process.*, vol. 30, pp. 5875–5888, 2021.

[50] Y. Lei, Y. Tian, H. Shan, J. Zhang, G. Wang, and M. K. Kalra, "Shape and margin-aware lung nodule classification in low-dose CT images via soft activation mapping," *Med. Image Anal.*, vol. 60, Feb. 2020, Art. no. 101628.

[51] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.

KYEONGJIN ANN (Member, IEEE) received the B.S. degree in computer science engineering from Hankuk University of Foreign Studies, South Korea, in 2019. She is currently pursuing the M.S. degree in medical science with Yonsei University, Seoul, South Korea. Since 2018, she has been a Researcher with the Medical Image Computing and Analysis Laboratory, CONNECT-AI Research Center. Her research interests include the application of computer vision techniques from image processing and deep learning to segmentation and localization in medical images.

YEONGGUL JANG received the B.S. degree in computer science engineering from Hankuk University of Foreign Studies, South Korea, in 2014. He is currently pursuing the Ph.D. degree in medical science with Yonsei University, Seoul, South Korea. Since 2012, he has been a Researcher with the Medical Image Computing and Analysis Laboratory, CONNECT-AI Research Center. His research interests include the application of computer vision techniques, geometric analysis, image processing, and deep learning to segmentation and localization in medical images.

HACKJOON SHIM received the B.S. and Ph.D. degrees from Seoul National University, Seoul, South Korea, in 1993 and 2007, respectively. From 2008 to 2010, he was a Research Professor with the Department of Electrical and Computer Engineering, Seoul National University College of Engineering, Seoul. Since 2011, he has been a Research Professor with Severance Hospital, Yonsei University, Seoul. His research interests include computer vision and medical image analysis.
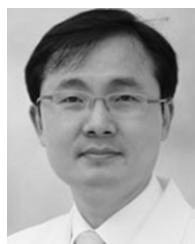
HYUK-JAE CHANG received the M.D. and M.M. degrees from the College of Medicine, Yonsei University, Seoul, South Korea, in 1994 and 1999, respectively, and the Ph.D. degree in medicine from the Graduate School of Medicine, Ajou University, Suwon, South Korea, in 2003. From 2002 to 2004, he was an Assistant Professor with the Department of Internal Medicine, College of Medicine, Ajou University. From 2004 to 2009, he was an Assistant Professor with the Department of Internal Medicine, Seoul National University College of Medicine, Seoul. From 2007 to 2008, he was a Postdoctoral Fellow with the Department of Internal Medicine, Johns Hopkins University, Baltimore, MD, USA. Since 2009, he has been a Professor with the Department of Internal Medicine, Yonsei University College of Medicine. Since 2016, he has been the CIO with the Division of Medical Information and Technology, Severance Hospital. Since 2017, he has been CEO with PHI Digital Healthcare, Inc., Seoul. His research interests include cardiovascular disease, pulmonary hypertension, and medical image processing.

● ● ●