2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

Enhancing Automated COVID-19 Chest X-ray Diagnosis by Image-to-Image GAN Translation

Zhaohui Liang
Dept.of Electrical Engineering & Computer Science
York University
Toronto, Canada
stan79@yorku.ca

Jimmy Xiangji Huang
School of Information Technology
York University
Toronto, Canada
jhuang@yorku.ca

Jun Li
Guangdong Provincal Hospital of Chinese Medicine
Guangzhou Univ Chinese Med
Guangzhou, China
lijun1401@163.com

Stephen Chan
Dapasoft INC.
Toronto, Canada
schan@dapasoft.com

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-7281-6215-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/BIBM49941.2020.9313466

Abstract—**The severe pneumonia induced by the infection of the SARS-CoV-2 virus causes massive death in the ongoing COVID-19 pandemic. The early detection of the SARS-CoV-2 induced pneumonia relies on the unique patterns of the chest X-Ray images. Deep learning is a data-greedy algorithm to achieve high performance when adequately trained. A common challenge for machine learning in the medical domain is the accessibility to properly annotated data. In this study, we apply a conditional adversarial network (cGAN) to perform image to image (Pix2Pix) translation from the non-COVID-19 chest X-Ray domain to the COVID-19 chest X-Ray domain. The objective is to learn a mapping from the normal chest X-Ray visual patterns to the COVID-19 pneumonia chest X-ray patterns. The original dataset has a typical imbalanced issue because it contains only 219 COVID-19 positive images but has 1,341 images for normal chest X-Ray and 1,345 images for viral pneumonia. A U-Net based architecture is applied for the image-to-image translation to generate synthesized COVID-19 X-Ray chest images from the normal chest X-ray images. A 50-convolutional-layer residual net (ResNet) architecture is applied for the final classification task. After training the GAN model for 100 epochs, we use the GAN generator to translate 1,100 COVID-19 images from the normal X-Ray to form a balanced training dataset (3,762 images) for the classification task. The ResNet based classifier trained by the enhanced dataset achieves the classification accuracy of 97.8% compared to 96.1% in the transfer learning mode. When trained with the original imbalanced dataset, the model achieves an accuracy of 96.1% compared to 95.6% in the training from train-by-scratch model. In addition, the classifier trained by the enhanced dataset has more stable measures in precision, recall, and F1 scores across different image classes. We conclude that the GAN-based data enhancement strategy is applicable to most medical image pattern recognition tasks, and it provides an effective way to solve the common expertise dependence issue in the medical domain.**

Keywords— COVID-19, generative adversarial network, GAN, image classification, deep learning

I. Introduction

Coronavirus disease 2019 (COVID□19) is an infectious disease caused by a novel coronavirus SARS-CoV-2. The most common clinical manifestation of COVID-19 infection is a special type of pneumonia which rapidly leads to severe acute respiratory infection symptoms and rapidly develops into acute respiratory distress syndrome (ARDS) [1]. The diagnostic methods of COVID-19 include new medical technologies from various domains. Though the gold standard for confirmation of COVID-19 is the real-time reverse-transcriptase polymerase chain reaction (RT-PCR), the test sensitivity is unsatisfactory at about 96.0%, and its performance is also related to the disease prevalence in the given population [2]. As a result, the diagnosis of COVID-19 needs to combine with other clinically accessible methods such as contact history, physical examination, and radiographic imaging.

Deep neural network or deep learning (DL) is one of the greatest innovations of artificial intelligence (AI) for medical applications [4]. Since the outbreak of the COVID-19 pandemic, many initial studies on applying DL respective for CT [5-6], CXR [7], and LUS [8] has been published. However, DL is considered as a data-greedy and expensive algorithm. The DL performance largely relies on the adequate computing resources (e.g. GPU and high RAM) and large datasets to sufficiently learn the complex mapping from the input data to the output result. COVID-19 is a new disease appeared in December 2019 [9]. Therefore, the number of images in the accessible data collections (CT and CXR) are usually inadequate for develop a high-performance DL model if it is trained from scratch, or there will be a bottle net preventing from the model to achieve high performance. An alternative solution is to applied a transfer learning strategy. Transfer learning is a pragmatic method for image processing tasks when acquiring enough training samples is difficult. This strategy is effective for medical image classification for rare or new diseases. The deep neural network model relies on its deep architecture to capture the complex patterns to confirm the diagnosis, but the deeper the network is, the more parameters to be optimized thus the more labeled data samples it requires. This character constrains the applications of DL in medicine because the highly expertise dependency in the medical domain. In contrast, the transfer learning model contains most of the pattern capturing filters that have been already optimized, therefore, the network can be optimized with a relatively low cost. The previous studies indicate that the main restriction on effectively implementing DL for COVID-19 is the data accessibility for enough amount of labeled COVID-19 image examples. A DL model trained by an imbalanced dataset will reach a performance threshold much earlier than its theoretical capacity determined by the architecture. In addition to adding new annotated image to increase the quantity of the data samples, another approach is to add simulated images that is good enough to mimic the real COVID-19 CXR images. This idea becomes the intuition of our study.

II. RATIONALE AND METHOD

A. Deep Convolutional Generative Adversarial Networks

Deep convolutional generative adversarial Networks, or DCGAN, are deep learning architecture for generating images. A generator model can generate new artificial samples which plausibly come from an existing distribution of samples. In general, a GAN is comprised of generator (G) networks and discriminator (D) networks (can contain more than one G or D given different tasks). The generator is responsible for generating new samples from the domain, and the discriminator is responsible for classifying whether samples are real or fake (generated). The performance of the discriminator model is used to update both the model weights of the discriminator alone and the generator model inside the GAN. The main constraint of the original GAN is that it cannot generate images on a particular class because the D in the GAN architecture translates the image from a random latent space without control. Therefore, the conditional generative adversarial network, or cGAN is introduced. It is an extension of the basic GAN that involves the conditional generation of images by a generator model. Image generation can be conditional on a specific, which allows the cGAN model to encode particular patterns in the training process, so that the D network can generate output images with desired [11]. The cGAN architecture has been successfully applied to medical imaging tasks [12-13].

The cGAN architecture provides a method to generate images of a specific class via the GAN, which extends use the GAN to improve image classification for small or imbalanced datasets. It is useful for medical imaging because the acquisition of annotated medical images is usually expensive. The image to image translation or pix2pix is analogic to automatic language translation, where the input pixcel values is believed to follow a certain distribution. When the cGAN model is sufficiently optimize, the G network can learn the representation of the image of a given distribution, and it will generate plausible images which is considered as real images to be added to the training dataset for classification tasks [14]. The detailed of the cGAN for image to image translation will be discussed in the next section.

B. cGAN Architecture and Optimization

The original GAN model learns n a mapping from a latent space with random noise vector z to output image y, i.e. $G: z \rightarrow y$ [10]. In comparison, a cGAN learns a mapping from both a input image belonging to a desired class and the random noise vector z to y, i.e. $G: \{x, z\} \rightarrow y$. The G network is optimized to in the adversarial manner with the D network to produce plausible image that can confuse the D network. In an ideal condition, the D network cannot effectively discriminate the 'fake' images of a given class from the real image from the training set. Therefore, a well-trained cGAN can render qualified training samples to balance the training set to enhance the generalization of the trained classifier when we use the GAN architecture for classification tasks, which is a potential solution to expedite the GAN application to medical imaging.

The objective of a cGAN has two folds. On the one hand, the generator network or G network tries to minimize the loss against an adversarial discriminator or D network. On the other hand, the D network tries to maximize the loss. Thus, the optimization objective is written as:

$$G = arg\ min_G max_D L_{cGAN}(G, D) \quad (1)$$

where the loss of the cGAN is expressed as:

$$L_{GAN}(G, D) = \mathbb{E}_y[\log D(y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x, z)))] \quad (2)$$

In Equation (2), the task of the D network is to discriminate the real images from the 'fake' image by the G network, but the task of the G network is not only to confuse the D network but also to come close to the ground truth output by measuring the L1 loss [14]. The final optimization objective of the image to image translation cGAN is written as:

$$G = arg\ min_G max_D L_{cGAN}(G, D) + \lambda L_1(G) \quad (3)$$

In Equation (3), it is easy to find that the network can learn a mapping from x to y even without the random z from the latent space, A concern on this setting is that the G will eventually learn the delta function only without any distribution from the training data. This problem can be solved by using real images as the true background other than a random latent space, because our task is to generate COVID-19 CXR images from normal CXR images. The discriminable patterns for COVID-19 in CXR is defined as: normal CXR lung pattern + mottling and ground-glass opacity on the lung area [1], which implies the normal lung CXR images can serve as the best inputs for the G network. The overall cGAN architecture is illustrated in Fig. 1.

As shown in Fig.1, the cGAN model uses the normal CXR images to train the G network and the D network in the adversarial manner. They are also used as the 'template' to generate synthesized COVID-19 CXR images after the model is optimized. On the other hand, the real COVID-19 CXR images from the dataset is used to compute the L1 loss with the output of the G network with a arbitrarily lambda (set to 100 in this study) as the total loss of the generator. The discriminator loss is defined as the summation of two loss values: the real image loss and the generated image loss. The real image loss is a sigmoid cross entropy loss of a real image and an array of ones (assuming all real images); the generated image loss is a sigmoid cross entropy loss of a generated image and an array of zeros (assuming all fake images). By this network configuration, we use the real COVID-19 CXR images as the constraint to optimize the cGAN to generate COVID-19 images instead of randomly producing both COVID-19 and normal CXR images.

Since the cGAN architecture is computationally intensive, we need to simplify the network architecture. In our study, we use the U-net architecture for both the G network and the D network. U-Net is a convolutional neural network for biomedical image segmentation with a relative light layer configuration. The input and output size of the GAN is set to 224 x 224, 3 channels for the convenience for transfer learning in the later classification tasks.

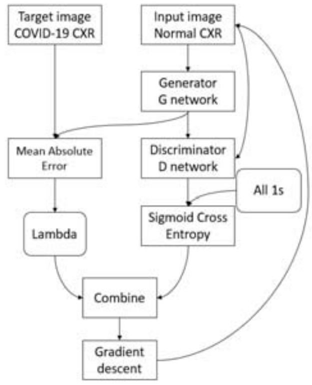

Fig. 1. Architecture of image to image cGAN for COVID-19 CXR

III. EXPERIEMENT AND RESULTS

A. Data Source and Preprocessing

The COVID-19 image set in this study contains 219 COVID-19 positive images, 1,341 normal images and 1,345 viral pneumonia images. It is donated by research team from Qatar University, Doha, Qatar and the University of Dhaka, Bangladesh, and with the collaborators from Pakistan and Malaysia. The raw dataset is accessible on the Kaggle website at: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database. The original study used a classic sequential CNN architecture and reported the classification accuracy ranges from 93% to 98%, but when the overfitting issue is excluded, the average accuracy is about 93%.

The experiments are implemented with the Tensorflow API on Google Colab Pro Cloud platform. For a robust result, we separate 60 COVID-19 CXR image into the test set for all the comparisons. All images are rescaled to 224 x 224 with 3 channels for both the cGAN and the classification experiments. To improve the diversity of the training samples, a random horizontal flip and a random 5-degree rotation is applied for data augmentation.

B. Image to Image cGAN Optimizatioin

The cGAN model for image to image translation is composed of a generator network (G network) and a discriminator network (D network). The G network has the architecture similar to U-Net with the encoder block having the Convolution + Batch normalization + Leaky ReLU activation structure, and with the decoder block having the Transposed Convolution + Batch normalization + Dropout (applied to the first 3 blocks) + ReLU activation structure. In order to preserve the gradient across the network, skip connections are added between the encoder and decoder blocks. The D network uses the network architecture connected by the Convolution + Batch normalization + ReLU activation blocks. The whole model is trained by 100 epochs. To form an enhanced dataset for COVID-19 image classification, we use the trained cGAN to generate 1,100 COVID-19 images from the normal CXR images and randomly select 30 of them for the separated test dataset, and the rest of them are added to the training set. The final cGAN-enhanced dataset contains 3,762 images (1,254 images for each class) for training and 180 images (60 images for each class) for test.

C. Residual-Net based Classifier

The evaluation of the GAN networks is usually subjective, because the ideal situation is to let the generator and the discriminator model be optimized by maintaining an equilibrium. However, the purpose of our study is to improve the classification performance of a CNN for the COVID-19 X-Ray images by using the cGAN to enhance the training sample balance, we can directly use the classification metrics to evaluate the cGAN model. Therefore, we use the original raw dataset, and the enhanced balance dataset with the generated images to respectively train a 50-layer residual network (ResNet). The two datasets are respectively trained from scratch and by transfer learning with the model pre-trained by the ImageNet dataset. The measurement includes classification accuracy and F scores.

At first, we trained a ResNet model from scratch with 60 epochs. To compare the effect of transfer learning with training from scratch, we also trained the same ResNet model with the pre-trained feature embedding layers optimized by the ImageNet dataset. The pretrained layers are attached with a global average pooling layer to flatten the output and it is followed by a dense layer with three nodes for classification. The pre-trained model was first trained for 50 epochs, then the last block of the pre-trained model was unfrozen to be further tuned for another 10 epochs with the learning rate of 1×10^{-5}.

The next step is to optimize the 50-layer ResNet with the enhanced dataset by the trained cGAN image-to-image model. The training on the enhanced dataset converges more rapidly compared to the original dataset, which is more obvious in the training from scratch manner. When comparing with the transfer learning process, the balance feature of the enhanced dataset helps the fine-tuning process at the last 10 epochs after several feature embedding layers on the top of the architecture are unfrozen. The learning curve does not appear a sudden jump when we convert the top feature embedding layer to be trainable. This implies that the balanced dataset with the generated COVID-19 images by the cGAN model can sufficiently simulate the general COVID-19 pathological patterns on the chest X-Ray images.

When comparing the classification metrics, all the trained classifiers have good performance with an overall accuracy more than 95%. The details are shown in TABLE 1.

The highest accuracy is 97.8% by the transfer learning model with the cGAN enhanced dataset, and the lowest accuracy is 95.6% by the training from scratch model with the original imbalanced dataset. By observing the precision, recall and F1 scores of all classes, we find that scores for the COVID-19 class is always higher than the other two classes. However, this phenomenon is interpreted differently. When trained with the original imbalanced dataset, the high scores for the COVID-19 class is due to the limit training samples in this class, because we duplicated the image samples in the COVID-19 by 8 times in order to match the number of samples in the other two class.

This strategy might bring a side-effect to the classifier to infer the test image by random chance instead of by the concrete learned patterns. This assumption can be confirmed by the relatively low precision score in the normal lung CXR class (0.935, in the transfer learning experiment), and obviously low precision score in the viral pneumonia class (0.906, training from scratch).

TABLE I. CLASSIFICATION PERFORMANCE

Measure	**Class**		
	Normal	***COVID-19***	***Viral***
Original imbalance dataset (transfer learning)			
Accuracy	0.961		
Precision	0.935	1.000	0.950
Recall	0.967	0.967	0.950
F1 score	0.951	0.983	0.950
Original imbalance dataset (train from scratch)			
Accuracy	0.956		
Precision	0.967	1.000	0.906
Recall	0.967	0.933	0.967
F1 score	0.967	0.966	0.935
cGAN enhanced dataset (transfer learning)			
Accuracy	0.978		
Precision	0.952	1.000	0.983
Recall	0.983	1.000	0.950
F1 score	0.967	1.000	0.966
cGAN enhanced dataset (transfer from scratch)			
Accuracy	0.961		
Precision	0.935	0.984	0.965
Recall	0.967	1.000	0.917
F1 score	0.951	0.992	0.940

IV. CONCLUSION AND DISCUSSIONS

The study demonstrates a novel data enhance strategy to improve medical image classification performance. It is helpful to enhance the DL performance towards small medical datasets with imbalanced classes. COVID-19 a new disease appearing in December 2019 [1]. Our assumption is that the patterns of COVID-19 on CXR images follows the rule: the visual pattern of a healthy lung + the unique patterns of COVID-19 pneumonia, which is the start point of the experiment setting. This assumption is supported by the medical expertise on COVID-19 by the current descriptive studies [1, 15]. Therefore, the image-to-image translation strategy with a well designed cGAN model and trained by a relatively small number of epochs becomes feasible for this task.

In conclusion, this study demonstrates a method to enhance the automated COVID-19 X-Ray image classification by an image-to-image cGAN. The results confirm that it is a solution to improve the balance of the training dataset and further the final performance of the trained image classifiers. We believe that this GAN enhancing strategy will become a low-cost and feasible method to improve the AI performance in the medical imaging domain.

REFERENCES

[1] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, et al, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," Lancet. vol. 395, pp. 507-513, February 2020.

[2] JJ. Deeks, J. Dinnes, Y. Takwoingi, C. Davenport, R. Spijker, S. Taylor-Phillips, et al, "Antibody tests for identification of current and past infection with SARS-CoV-2," Cochrane Database Syst Rev, vol. 6, pp. CD013652, June, 2020.

[3] L. Guo, L. Ren, S. Yang, M. Xiao, D. Chang, F. Yang, et al, "Profiling early humoral response to diagnose novel coronavirus disease (COVID-19)," Clin Infect Dis, vol. 71, pp.778-785, July, 2020.

[4] A. Sufian, A. Ghosh, AS. Sadiq, F. Smarandache, "A Survey on Deep Transfer Learning and Edge Computing for Mitigating the COVID-19 Pandemic," Journal of Systems Architecture. vol. 108, no. 101830, September, 2020.

[5] J. Wang, Y. Bao, Y. Wen, H. Lu, H. Luo, Y. Xiang, et al, "Prior-Attention Residual Learning for More Discriminative COVID-19 Screening in CT Images," in IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2572-2583, August. 2020.

[6] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, et al, "Accurate Screening of COVID-19 using Attention Based Deep 3D Multiple Instance Learning," IEEE Transactions on Medical Imaging. vol. 39, no. 8, pp. 2584-2594, August. 2020.

[7] Y. Oh, S. Park, JC. Ye, "Deep learning covid-19 features on cxr using limited training data sets," IEEE Transactions on Medical Imaging. vol. 39, no. 8, pp. 2688-2700, August, 2020.

[8] S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori C, "Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound," IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2676-2687, August, 2020.

[9] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, et al, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," The Lancet, vol. 395, no. 10223, pp. 507-13, Feburary, 2020.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al, "Generative adversarial nets,", In Proceedings of the 27th International Conference on Neural Information Processing Systems, pp. 2672-2680, December, 2014.

[11] J. Gauthier, "Conditional generative adversarial nets for convolutional face generation," In: Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester 2014; 2014: 2.

[12] P.M. Johnson, M. Drangova, "Conditional generative adversarial network for 3D rigid-body motion correction in MRI," Magn Reson Med., pp. 901-910, vol. 82, no. 3, September, 2019.

[13] E.L. Denton, S. Chintala, R. Fergus, "Deep generative image models using a Laplacian pyramid of adversarial networks," In Proceedings of the 28th International Conference on Neural Information Processing Systems, vol. 1, pp. 1486-1494, 2015.

[14] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, "Image-to-image translation with conditional adversarial networks," In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, USA, pp. 1125-1134, July, 2017.

[15] Z. Liang, A. Powell, I. Ersoy, M. Poostchi, K. Silamut, K. Palaniappan, et al. "CNN-based image analysis for malaria diagnosis," In Proceedings of 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 493-496, Dec 15, 2016.