COMPUTATIONAL
ANDSTRUCTURAL
BIOTECHNOLOGY
J O U R N A L

Deep learning for COVID-19 chest CT (computed tomography) image analysis: A lesson from lung cancer

Hao Jiang [a,b], Shiming Tang [c], Weihuang Liu [a,d], Yang Zhang [a,*]

[a] College of Science, Harbin Institute of Technology (Shenzhen), Shenzhen, Guangdong 518055, China
[b] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[c] School of Computing and Engineering, University of Missouri-Kansas City, MO, United States
[d] Department of Computer and Information Science, University of Macau, Macau, China

A B S T R A C T

As a recent global health emergency, the quick and reliable diagnosis of COVID-19 is urgently needed. Thus, many artificial intelligence (AI)-base methods are proposed for COVID-19 chest CT (computed tomography) image analysis. However, there are very limited COVID-19 chest CT images publicly available to evaluate those deep neural networks. On the other hand, a huge amount of CT images from lung cancer are publicly available. To build a reliable deep learning model trained and tested with a larger scale dataset, the proposed model builds a public COVID-19 CT dataset, containing 1186 CT images synthesized from lung cancer CT images using CycleGAN. Additionally, various deep learning models are tested with synthesized or real chest CT images for COVID-19 and Non-COVID-19 classification. In comparison, all models achieve excellent results in accuracy, precision, recall and F1 score for both synthesized and real COVID-19 CT images, demonstrating the reliable of the synthesized dataset. The public dataset and deep learning models can facilitate the development of accurate and efficient diagnostic testing for COVID-19.

1. Introduction

Since December of 2019, an outbreak of coronavirus disease (COVID-19) has spread rapidly throughout the world, with an ongoing risk of a pandemic [1,2]. In absence of specific therapeutic drugs for COVID-19, it is essential to diagnose this disease effectively and immediately. With currently reported cases and published papers, lung CT (computed tomography) imaging has been recommended as one of the effective screening tools for COVID-19 pneumonia [3,4].

Although CT images can be examined to identify the COVID-19 pneumonia regions with specific pattern by naked human eyes, CT screen is easy to miss those small and lightly infective regions especially in the early stage. Therefore, the sufficient training is necessary for radiologists to achieve an early-accurate diagnosis, which is indispensable not only for the prompt implementation of treatment but also for the population screening and response. However, the time-consuming and difficulty in professional train-

ing leads to the lack of qualified radiologists, which makes accurate diagnosis particularly challenging with the dramatically increasing cases nowadays [5].

To improve the reliability and speed of CT-based COVID-19 diagnosis, a more automatic and higher efficiency method is urgently demanded. Many researchers have noticed the Artificial Intelligence (AI), which already show great performance in other disease diagnosing cases [6,7], should also have the same feasibility in this novel pneumonia detection [8,9]. Lots of substantial evidences supporting the potential for deep learning in chest CT image analysis [10,11], particular in lung cancer analysis [12,13]. Thus, various deep learning-aided COVID-19 chest image analysis models were proposed with high accuracy and efficacy in disease diagnosis. Moreover, there are already hundreds of deep learning papers proposed relating to COVID-19 chest CT or X-ray images, and some of these results have been shown to be quite promising in terms of accuracy. Top cited deep learning studies using chest CT images of COVID-19 are summarized and listed in Table 1.

As shown in Table 1, Gozes et al proposed a deep learning-based automated CT image analysis tools for detection, quantification, and tracking of COVID-19 and demonstrated that they can differentiate patients from healthy ones [14]. In the segmentation step,

* Corresponding author.
E-mail addresses: zhangyang07@hit.edu.cn, yang.zhang2020@hotmail.com (Y. Zhang).

Table 1

Summary of top cited studies of deep learning-based COVID-19 analysis.

Ref	Backbone	Task	Dataset	Results
[14]	U-Net ResNet50	Segmentation Classification	6,150 CT images with lung abnormalities and their lung masks. 50 abnormal thoracic CT scans of patients that were diagnosed by a radiologist as suspicious for COVID-19. Cases were annotated for each image as normal (n = 1036) vs abnormal (n = 829).	AUC of 0.996, sensitivity of 98.2%, and specificity of 92.2%.
[15]	U-Net ResNet50	Classification	4352 3D chest CT images from 3,322 patients, consisting of 1,292 COVID-19, 1,735 community-acquired pneumonia, 1,325 Non-pneumonia.	AUC of 0.96, sensitivity of 90%, and specificity of 96%.
[16]	ResNet	Classification	618 CT images (219 COVID-19, 224 Influenza-A-viral-pneumonia, and 175 healthy case). 11,871 images (2,634 COVID-19, 2,661 Influenza-A viral pneumonia, and 6,576 healthy case).	Accuracy of 86.7%.
[17]	DRE-Net based on ResNet50	Classification	88 COVID-19 patients with 777 CT images, 100 bacterial pneumonia patients with 505 images, and 86 healthy people with 708 images.	AUC of 0.99 and recall of 0.93.
[18]	VB-Net	Segmentation	249 CT images of 249 COVID-19 patients were collected from other centers for training. 300 CT images from 300 COVID-19 patients were collected for validation.	Dice similarity coefficients of 91.6%±10.0%.
[19]	DenseNet-169	Classification	349 CT images of COVID-19 were extracted from 760 preprints about COVID-19 from medRxiv and bioRxiv. 463 Non-COVID-19 CT images were collected, consisting of 36 from LUNA, 195 from MedPix, 202 from PMC, and 30 from Radiopaedia.	F1 of 0.85, AUC of 0.95, and accuracy of 0.83.
[20]	Conditional GAN	COVID-19 image generating	CT images of patients were positive or suspected of COVID-19 or other viral and bacterial pneumonias (MERS, SARS, and ARDS).	Enhanced the identification and detection capacities of the classification models.
[21]	GAN	COVID-19 image generating	2143 chest CT images, containing 327 COVID-19 cases, were acquired from 12 sites across 7 countries.	Improve lung segmentation (+6.02% lesion inclusion rate) and abnormality segmentation (+2.78% dice coefficient).
[22]	Conditional GAN	COVID-19 image generating	829 lung CT slices from 9 COVID-19 patients.	Peak signal-to-noise ratio (PSNR) of 26.89, and structural similarity index (SSIM) of 0.8936.

they train U-net to extract the lung region of interest (ROI) using 6,150 CT images of cases with lung abnormalities (Non-COVID-19) taken from a U.S based hospital. Then, the ResNet-50 deep convolutional neural network pretrained on ImageNet is trained with suspected COVID-19 cases from several Chinese hospitals, which were annotated per slice as normal (n = 1036) vs abnormal (n = 829). They achieved an AUC of 0.996 (95%CI: 0.989–1.00) for classify COVID-19 confirmed cases from 56 patients vs normal thoracic CT scans from 51 patients without any abnormal lung findings in the radiologist's report. This study used a manually labeled dataset which limited the size of training samples. Although they used over thousand patches for training, these patches are from small amount patients' cases.

Similarly, another high-cited paper also used a ResNet_50 based framework for the detection of COVID-19, referred to COVNet [15]. It can extract both 2D local and 3D global representative features. They also used the U-net for segmentation to extract the lung region as the region of interest (ROI). The preprocessed image is then passed to COVNet for the predictions. The newly developed method achieved high sensitivity and high specificity in detecting COVID-19, with the ability to differentiate COVID-19 and community-acquired pneumonia (CAP) from chest CT images. This study used a larger chest CT dataset, which contains 4352 chest CT images (i.e., 1292 COVID-19, 1735 community-acquired pneumonia (CAP), and 1325 Non-pneumonia) from 3322 patients. Nevertheless, the dataset is not open for public access. Thus, it is impossible to determine the imaging features used by this model to distinguish between COVID-19 and CAP.

Besides, Xu et al. established a deep learning model to distinguish COVID-19 pneumonia from Influenza-A viral pneumonia and healthy cases with pulmonary CT images using a location-attention classification model [16]. The study collected 618 CT samples, including 219 from 110 patients with COVID-19, 224 from 224 patients with Influenza-A viral pneumonia, and 175 healthy cases. In this proposed system, useful pulmonary regions were extracted firstly; then, a 3D CNN model was used to segment multiple image cube candidates. Next, a location-attention classi-

cation model based on ResNet is followed to categorize image patches to COVID-19, Influenza-A viral pneumonia, or irrelevant-to-infection. Each image patch from the same cube was voted to represent the entire candidate region. Finally, the comprehensive analysis report for one CT sample was calculated using the Noisy-or Bayesian function.

Ying et al. designed a Details Relation Extraction Neural Network (DRE-Net) based on pretrained ResNet_50, on which the Feature Pyramid Network (FPN) was added to extract the top-K details in the CT images. An attention module is coupled to learn the importance of each detail. By using the FPN and attention modules, DRE-Net achieved better performance in pneumonia classification and diagnosis compared to ResNet, DenseNet and VGG16 [17].

In another way, Shan et al. introduced a VB-Net based deep learning framework with a human-in-the-loop (HITL) strategy for segmentation of COVID-19 infection regions from chest CT scans [18]. The HITL strategy is adopted to assist radiologists to refine automatic annotation of each case. The system is trained using 249 COVID-19 patients and validated using 300 new COVID-19 patients, yielding high accuracy for automatic infection region delineation. Moreover, compared with the cases of fully manual delineation that often takes hours, the proposed human-in-the-loop strategy can dramatically reduce the delineation time to several minutes after three iterations of model updating.

In general, most of the listed highly cited works are based on ResNet to realize a patches classification; insufficient CT images are used to train deep learning models, which affects the generalization ability of the model. Notably, barely founded open-source codes and datasets of proposed deep learning methods limited the more profound understanding and improvement for the research community to help more patients in this particular pandemic period.

Besides, building a reliable deep learning model always has the challenging in requirement of vast amounts of data for training. Since the sudden happening and short research of COVID-19, the limited amount of lung CT images of COVID-19 are available and open source. The lack of raw data badly hindered the development

and evaluation of the deep learning model for COVID-19 detection. Yang et al. build an open-sourced dataset to relieve this predicament, which contains 349 COVID-19 CT images from 216 patients and 463 Non-COVID-19 CTs [19]. Those CT images of COVID-19 and Non-COVID-19 are extracted and collected from publications and online databases. Using this dataset, DenseNet-169 model was used for binary classification of COVID-19 or Non-COVID-19, achieving an F1 of 0.85, an AUC of 0.95, and an accuracy of 0.83. However, the scale of this dataset is not enough to train a reliable deep learning model. On the other hand, a huge amount of public lung CT images are accumulated because of the well-established studies in lung cancer, such as LUNA16, LIDC-IDRI [23,24]. Those chest CT images are publicly available with sufficient quantity and can be developed to aid the deep learning model training, especially in the lesion containing regions.

To solve the dataset limit challenge, a few proposed papers focused on synthesizing COVID-19 from other large-scale lung CT datasets based on GAN models. Both Jiang's team and Li's team used conditional GAN to synthesize the high-quality COVID-19 CT scans to provide more data for corresponding machine learning models. The generated CT scans can enhance the detecting and classifying capability of models [20,22]. Liu et al. also proposed a GAN model to generate COVID-19 related tomographic patterns on chest CTs from negative cases. Synthetic data are used to improve both lung segmentation and segmentation of COVID-19 patterns [21].

Inspired by the GAN models mentioned above, this paper tries to leverage the rich label information and large scale in lung cancer datasets by incorporating them into deep learning model for COVID-19 detection. Moreover, CT examination of patients with COVID-19 pneumonia showed extensive consolidation and ground-glass opacity (GGO) pattern [25]. The COVID-19 lesions with GGO pattern can be seen in Fig. 1A. With the guidance of this prior knowledge from the radiologist on GGO pattern, our deep learning study is the first research that uses this CT feature of COVID-19. With particular focus on the GGO pattern, a deep learning-based method is proposed to generate synthesized pneumonia CT images of COVID-19 from lung cancer images. This study might provide an alternative solution for the reliable COVID-19 automatic diagnosis.

Our previous work shows that a Cycle Generative Adversarial Network (CycleGAN) based deep learning method can transfer expert knowledge for microscopic image recognition [26]. CycleGAN is an unsupervised learning model for image-to-image translation, which learns the image style through competitive strategies. CycleGAN is based on GAN [27], which is widely used in medical image processing, including reconstruction [28], classification [29], detection [30] and segmentation [31]. The mutual generation of unpaired images is an important feature for CycleGAN, which is different from other GAN derivatives. Therefore, the CycleCAN is utilized in this paper to generate the chest CT images of COVID-19 pneumonia by image-to-image translation, which learns the style and pattern of GGO and applies this GGO pattern to CT lung cancer images. In detail, the model learns the GGO knowledge from COVID-19 images allowing the lung nodule image with labelled information to adapt this feature. To evaluate the efficacy of generated COVID-19 dataset on automatically AI diagnosis, we test various deep learning models for classification and demonstrate its excellent performance in COVID-19 CT image analysis.

2. Materials and methods

2.1. Dataset

Publicly available CT images of COVID-19 pneumonia and lung cancer are used in this study (Fig. 1A). This COVID-19 dataset

includes a total of 349 COVID-19 CT images and 397 Non-COVID-19 CT images [19]. These CT images are extracted from 760 preprints reporting COVID-19 on medRxiv and bioRxiv. This dataset is used as a priori knowledge domain to learn the ground-glass opacity (GGO) pattern in the CT image. As a result of the large-scale and rich label, LUNA16 dataset (lung cancer dataset) is used in the experiment to synthesize COVID-19 images for detection. This LUNA16 contains 888 lung cancer CT scans from 888 patients with pulmonary nodules annotated. The size of each original 3D image is 512 × 512 × 3 × S (S is the number of 2D slices). From these, a total of 1186 2D images with lung nodules are used for training the COVID-19 synthesizer, and the rest slides are used as positive samples, which constitutes the Non-COVID-19 dataset to train the COVID-19 classifiers [23]. For the further research, the codes and data sets that support findings of this study are available on https://github.com/jiangdat/COVID-19.

2.2. Generation of COVID-19 CT images from lung cancer based on style transfer

This work proposes a dataset-driven deep learning strategy based on style transfer for generating COVID-19 CT images (Fig. 1B). Using the large-scale lung cancer CT dataset with rich label information, the model learns the GGO style of COVID-19 to synthesize a COVID-19 dataset using a CycleGAN model. The synthetic COVID-19 dataset with the location label of the lesion is used to train deep learning models for classification. In comparison, the real COVID-19 is also tested to verify and validate the generated COVID-19 images. An unpaired mapping-based approach is used for generating COVID-19 CT images. Usually, a chest CT slice image contains the entire lung structure, parts of which are lesions, and the rest are normal regions. Their sizes, positions, and shapes vary significantly; thus, this study applied the unpaired strategy.

The architecture of COVID-19 generation in Fig. 1C briefly illustrates that CycleGAN are used to generate COVID-19 CT images for training the COVID-19 detection and classification models. The CycleGAN framework can randomly learn a mapping between COVID-19 CT images and unpaired lung nodule CT images. In other words, the GGO styles is combined in the COVID-19 pneumonia domain (C). And the deep learning is applied to take the advantages of annotations and data richness in lung nodule domain (N). Also, the adversarial loss and cycle consistency are used in a dual-GAN architecture to transfer between domain (C) and domain (N). This setup learns a reverse mapping from COVID-19 images to lung nodule images. A detailed explanation is followed in the below paragraphs.

As shown in Fig. 1C, our dual-GAN architecture consists of two domains (COVID-19 domain (C) and lung nodule domain (N)), and four networks including two generations (G_C and G_N) and two discriminators (D_N and D_C). G_C is from random COVID-19 to lung nodule CT image and G_N is from lung nodule to COVID-19 CT image, D_N and D_C are corresponding to G_N and G_C separately.

As mentioned above, the loss of training is joint of adversarial loss L_{GAN} and cycle consistency loss L_{cyc}, where the adversarial loss is applied in both generators. For the mapping function $G_C: C \rightarrow N$ and its discriminator D_N, the adversarial loss is expressed as:

$$L_{GAN}(G_C, D_N) = E_{n \sim p_{data}(n)}[\log D_N(n)] + E_{c \sim p_{data}(c)}[\log(1 - D_N(G_C(c)))] \quad (1)$$

in which G_C aims to generate lung nodule images $G_C(C)$ that are similar to real lung nodule images, while D_N aims to distinguish the generated images $G_C(C)$ from real images. And $L_{GAN}(G_C, D_N)$ is a binary cross entropy (BCE) loss of D_N in classifying real or fake. D_N and G_C play a min–max game to maximize and minimize this
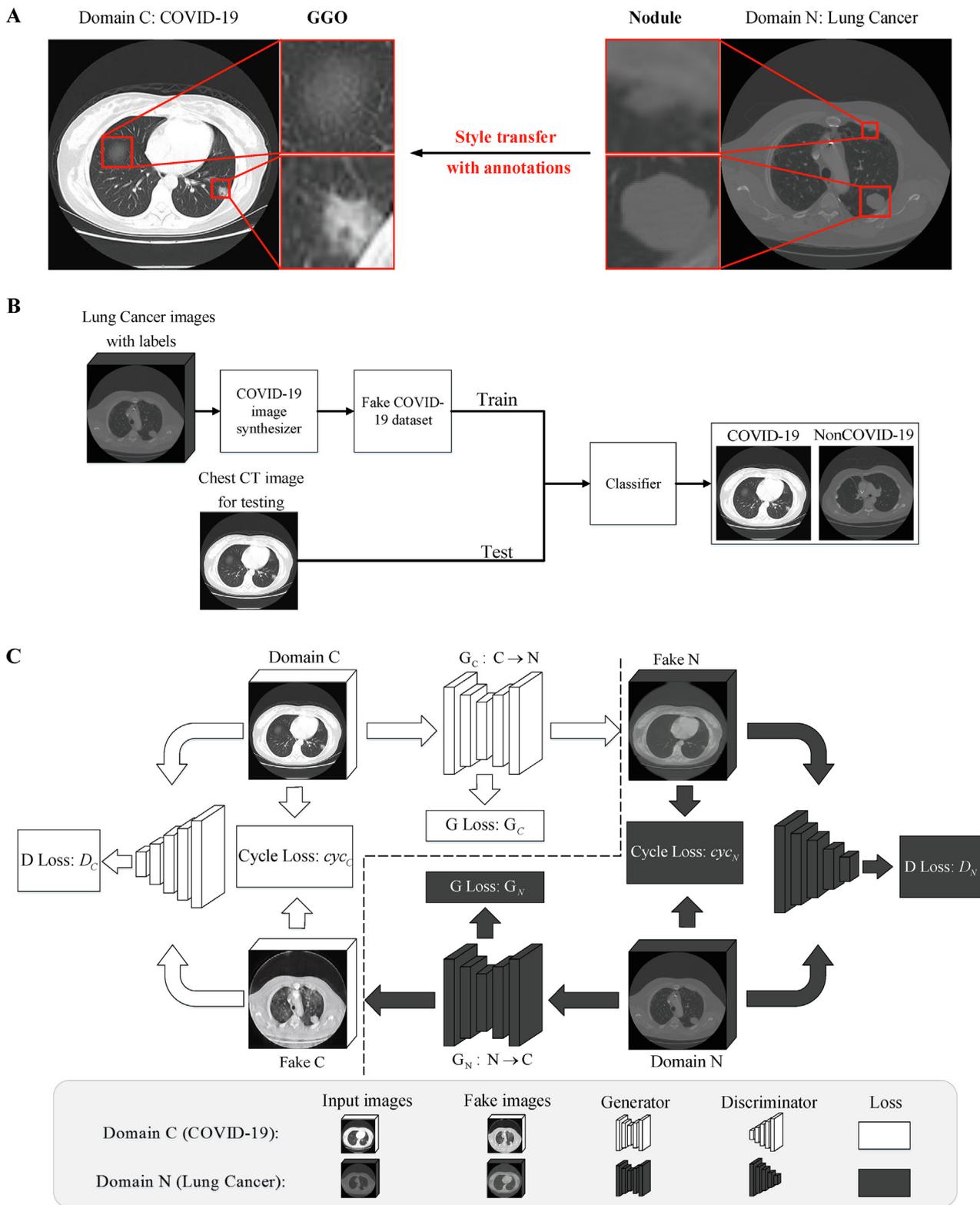
Fig. 1. Overview of CycleGAN-base deep learning for COVID-19. (A) Representative chest CT images. Publicly available COVID-19 pneumonia images have infected areas with GGO pattern, and images of lung cancer with distinct nodules source from LUNA16. (B) COVID-19 analysis model based on style transfer. The COVID-19 dataset synthesized from lung cancer images is used to train classifiers, and synthesized or real COVID-19 chest CT images are used for testing. (C) A graphical illustration of CycleGAN based deep learning for COVID-19 CT image construction. This structure is divided into two symmetrical parts, for domain C, Generator C tries to transform the GGO style of COVID-19 into the nodule style of lung cancer. The Discriminator C is used to compare the real COVID-19 with fake COVID-19 learned from domain N. Cycle loss is used for supervising the continuity of the input and the image circulated after two generations.

loss term respectively. And for the reverse $N \rightarrow C$ generation with a similar objective, its adversarial loss can be denoted as:

$$L_{GAN}(G_N, D_C) = E_{c \sim p_{data}(c)}[\log D_C(C)]$$
$$+ E_{n \sim p_{data}(n)}[\log(1 - D_C(G_N(N)))] \quad (2)$$

The cycle consistency loss L_{cyc} term ensures that the forward and back translations between the COVID-19 images and lung nodule images are lossless and cycle consistent, i.e., $G_N(G_C(C)) \approx C$ (forward) and $G_C(G_N(N)) \approx N$ (backwards). L_{cyc} is defined as below:

$$L_{cyc}(G_C, G_N) = \lambda_C E_{c \sim p_{data}(c)}[\| G_N(G_C(C)) - C \|_1]$$
$$+ \lambda_N E_{n \sim p_{data}(n)}[\| G_C(G_N(N)) - N \|_1] \quad (3)$$

where λ_C and λ_N control the relative importance of the two objectives, and the full objective for synthetic data generation can thus be written as:

$$\operatorname*{argmin}_{G_C, G_N} (\operatorname*{argmax}_{D_C, D_N}(L_{GAN}(G_C, D_N) + L_{GAN}(G_N, D_N) + L_{cyc}(G_C, G_N))) \quad (4)$$

During the training, the deep network iterates and optimizes the parameters based on the above objective to obtain a reliable COVID-19 synthesizer.

2.3. Deep learning-guided classifiers for COVID-19 CT images

In this study, the data and knowledge-driven high-precision classifiers to distinguish between COVID-19 and Non-COVID-19 are established and compared, by taking the advantages of a large scale of synthetic COVID-19 CT images. To evaluate the generated COVID-19 dataset, some of best performing CNNs, including VGG [32], ResNet [33], Inception-v3 [34], Inception_ResNet_v2 [35] and DenseNet [36], are deployed in the classification experiments.

2.4. Training and evaluation metric

The implementation of all networks is deployed on the TensorFlow framework [37] in Ubuntu 16.04. And the training hardware configure is one Tesla K40C GPU and 128-GB memory. Two datasets are used in our experiments, containing 1186 lung cancer and 349 COVID-19 CT images. The general data pre-processing (Augmentation + Resize) is applied: (i) the data augmentation procedure includes scaling, flipping, cropping, and rotating. After the augmentation, the two original datasets are expanded to a 3000-images dataset. (ii) the resize operation scales the COVID-19 dataset (various size) to the same size as lung cancer images ($512 \times 512 \times 3$). In the training of COVID-19 synthesizer, the general settings are followed the original CycleGAN [38], for example, the parameters λ_C and λ_N in formula (3) are both set as 10. And the regular training parameters batch size and learning rate are set as 4 and 2×10^{-4} individually. Besides, the network is optimized by Adam optimizer [39].

When training the COVID-19 classifier, the dataset includes 1000 images of synthetic COVID-19 and 1000 images of normal lung CT. And two test datasets are used to evaluate the effectiveness of the synthetic images and the classifier, including the real test dataset with 300 images of real COVID-19 and 300 images of Non-COVID-19, and the synthetic test dataset with 300 images of synthetic COVID-19 and 300 images of Non-COVID-19.

For measuring the performance of the classification task, accuracy, recall, precision and F1 score are calculated, and their formulas can be represented as:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$F_1 = 2\frac{Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \quad (8)$$

where TP, FP, FN, TN are True Positive, False Positive, False Negative and True Negative. As one of most important metrics, F1 is used to measure the accuracy of a test, by considering both the precision and the recall of the test to compute the score.

Additionally, the ROC (Receiver Operating Characteristic) curve is also plotted to show the performance of each model. The vertical axis of the ROC curve is True Positive Rate (TPR), and the horizontal axis is False Positive Rate (FPR), and AUC (Area Under Curve) is defined as the area under the ROC curve and the coordinate axis, the ROC curve is connected by $\{(x_1, y_2), (x_2, y_2), \dots, (x_m, y_m)\}$, AUC is expressed as:

$$AUC = \frac{1}{2}\sum_{i=1}^{m-1}(x_{i+1} - x_i) \cdot (y_i + y_{i+1}) \quad (9)$$

ROC curve and AUC are used to evaluate the performance of the model.

3. Results

3.1. Synthetic datasets description

As the result of the experiment and the main contribution, we have established a COVID-19 dataset, which are 2D images synthesized from CT images of lung cancer. This dataset contains 1186 CT images of COVID-19 pneumonia with the size of $512 \times 512 \times 3$. Each CT image contains apparent COVID-19 lesion features with GGO pattern. According to our experiments, this dataset can be used for further analysis of COVID-19, such as classification experiments. More details and images can be obtained at https://data.mendeley.com/datasets/kdn5v76wb3/draft?preview=1.

3.2. Generation of synthetic COVID-19 CT images from lung cancer

To construct COVID-19 synthetic images, the style strategy is applied to transfer the GGO features of COVID-19 into the lung cancer CT image. Therefore, a GAN-based deep learning model is trained to achieve an optimal result. The training dataset contains 349 images of COVID-19 and 1186 images of lung cancer. Each image patch has a size of 512×512 pixels, and the raw input lung cancer CT images to the network are collected from LUAN16. The results of the network are compared against the ground truth (real COVID-19 images). A random example of the network input image is shown in Fig. 2A, where the generated CT images of COVID-19 have GGO pattern not only similar to ground truth, but also with annotations of nodule regions. Details of GGO pattern are shown in the zoomed-in regions of interest (ROIs) in Fig. 2B, C. A pre-trained CycleGAN based deep neural network is applied to these input images (lung cancer and COVID-19), and the output is the GGO pattern enabled COVID-19 CT images with annotations, where GGO features of COVID-19 are clearly resolved. The model provides an excellent agreement with the ground truth images (COVID-19) shown in Fig. 2C.

Moreover, from the result in the Fig. 2, the generated image shows more lesions with GGO pattern compared to the COVID-19 images, and these generated GGO pattern located around the original lung nodules. This factor does not affect the further network evaluation for COVID-19 analysis as more lesions with GGO pattern might be the benefits of training. Note that all the network output images shown in this article are blindly generated by the deep network, that is, the input images are not previously seen by the network.
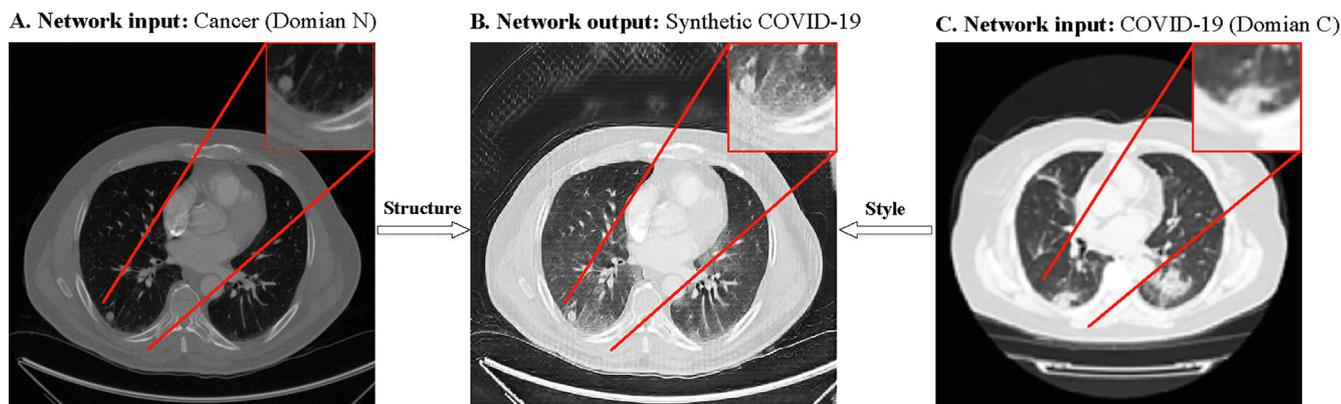
Fig. 2. Deep-learning enabled CT image transformation from lung cancer to COVID-19. (A) Input lung cancer CT image. (B) Reconstructed image obtained using the CycleGAN based deep learning method. (C) Input COVID-19 CT image. Zoomed-in regions of lesion in COVID-19 and lung cancer, highlights the success of generation of GGO pattern. Experiments are repeated through the whole lung cancer dataset, achieving similar results.

3.3. Effectiveness of our method

To compare the similarity of the synthetic and real COVID-19 images, statistical distribution and quantitative measurement are tested on our dataset [40–42]. For statistical analysis, the pixels of chest CT images are first counted, including CT images of lung cancer (source domain), synthetic COVID-19 (generation domain) and real COVID-19 (target domain). Then the histograms of three randomly sampled CT images are plotted in Fig. 3A. The image histogram is a gray-scale value distribution showing the frequency of occurrence of each gray-level value. The histogram analysis showed that the gray-scale values of lung cancer (purple curve) and both COVID-19 images (red curve: real image; green curve: synthetic image) are distinguishable (Fig. 3A). This results in curves shifted appearing on the histogram of lung cancer to two COVID-19 images. In addition, both gray-scale values of synthetic and real COVID-19 image are very close, indicating a high level of agreement. With respect to the lesions, the local distribution of lung nodule (purple curve) and GGO pattern of COVID-19 (red curve: real one; green curve: synthetic one) are plot with similar results obtained (Fig. 3B).

For quantitative measurement, KL divergence is computed to compare the agreement of both synthetic COVID-19 image and lung cancer image (Fig. 3C). KL divergence quantifies the level of agreement relative to real COVID-19 CT image. The KL divergence of synthetic image to real COVID-19 is 0.043, while the divergence

value of cancer image to real COVID-19 is 0.097, showing a strong correlation between real COVID-19 image and synthetic one.

3.4. Classification performance on synthetic COVID-19 CT images

In this study, various deep learning-based COVID-19 classification methods are evaluated on this newly synthesized COVID-19 CT image dataset. Specifically, VGG16, ResNet-50, Inception_ResNet_v2, Inception_v3, and DenseNet-169 models are trained in the generated dataset, which is consisted of 1000 images of synthesized COVID-19 and 1000 images of Non-COVID-19. To compare, various classification models are trained on the synthetic COVID-19, and tested with synthetic COVID-19 dataset and real COVID-19 dataset and the two test sets are named as Synthetic Test and Real Test respectively. Here, the synthetic COVID-19 dataset contains 300 generative COVID-19 images and 300 Non-COVID-19 images. While, the real COVID-19 dataset includes 300 real COVID-19 images and 300 Non-COVID-19 images. The performance metrics of above mentioned 5 classification models, including accuracy, precision, recall and F1 score, are summarized in Table 2. And the ROC curves are also plotted in Fig. 4 to help the performance evaluation.

According to the results (Table 2), all models achieve the excellent results (all the average metrics are greater than 90%), which means that the reliable model for diagnosing COVID-19 can be trained by the synthesized dataset. In the Synthetic Test, the gen-
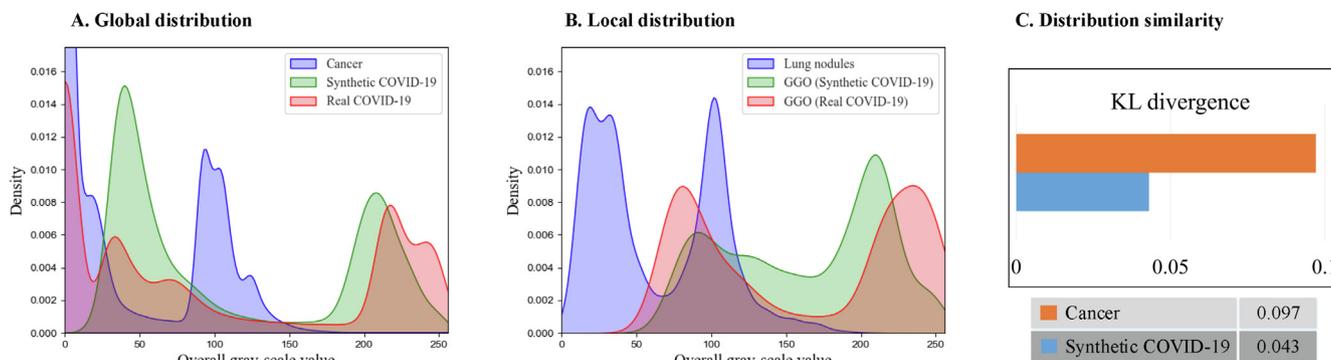

Fig. 3. Comparison the synthetic and real COVID-19 CT images. A and B show the grayscale image of global and local distribution using the histogram method, respectively. As shown in the histogram, the pixels in both synthetic image and real COVID-19 CT reside more on the larger scale, and hence the synthetic CT image is similar to the real COVID-19 one. C. The KL divergence is computed to compare the agreement of both synthetic COVID-19 image and lung cancer image by quantifying the level of agreement relative to real COVID-19 CT image. The increased value indicated a lower level of similarity.

Table 2
Test result of different classification models on real and synthetic COVID-19 CT images.

| Model	Synthetic Test				Real Test			
	Accuracy	Recall	Precision	F1	Accuracy	Recall	Precision	F1
VGG16	94.19	88.15	100.00	93.70	94.80	88.15	98.52	93.05
ResNet_50	94.83	89.47	100.00	94.44	94.10	89.47	95.32	92.30
Inception_v3	96.55	96.05	96.90	96.47	95.32	96.05	92.40	94.19
Inception_ResNet_v2	95.91	91.67	100.00	95.65	96.70	91.67	100.00	95.65
DenseNet_169	98.92	97.80	100.00	98.89	98.09	97.80	97.37	97.92
Average	96.08	92.63	99.38	95.83	95.80	92.63	96.72	94.62
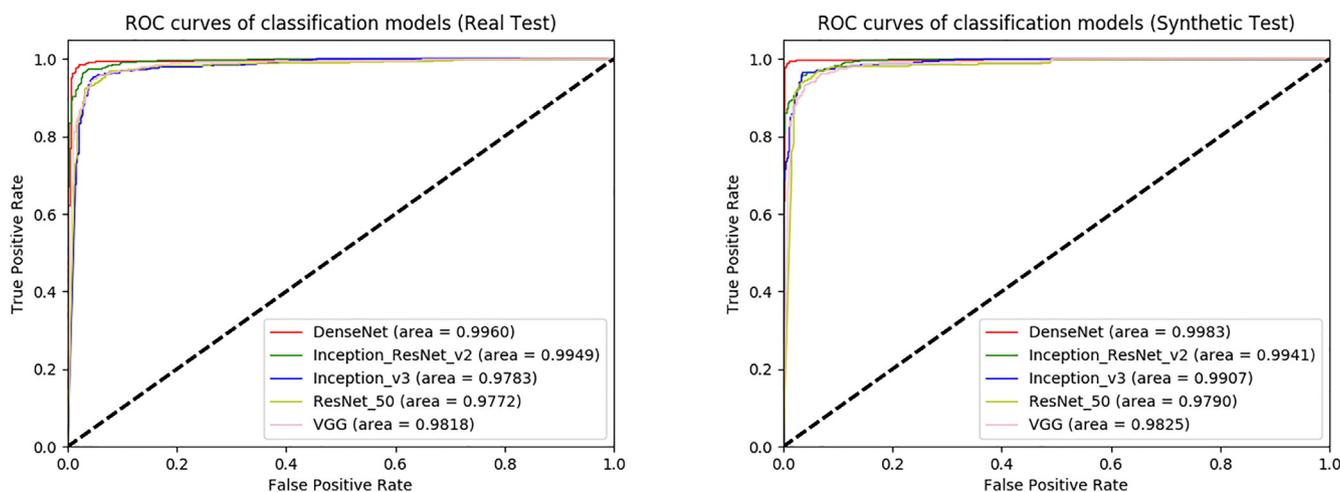
Fig. 4. ROC curves of five classification models on real or synthetic dataset. The lines colored by red, green, blue, yellow, pink are the ROC curves of DenseNet_169, Inception_ResNet_v2, Inception_v3, ResNet_50 and VGG16 respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

eral performance is slightly better than the Real Test: the average accuracy is around 96% and the average precision even over 99%. This is an acceptable phenomenon because the models are trained by the synthetic data and the synthetic test data has more similarity with the train data. For the Real Test, all average metrics are over the 90%, which reflects that the classification models still can recognize COVID-19 cases, even though the models are only trained by generated data without one real COVID-19 images. The Real Test has achieved a 95.80% average accuracy and 94.62% F1 score, which are very closed to Synthetic Test. This result is already better than the performances reported by other AI-based diagnosis methods.

In detail, the simplest VGG16 model has relatively worse performance in synthetic set and ResNet_50 in the real set, but their F1 score are still over 90%. This may be in result of its shallower structure as the basic model. Furthermore, the DenseNet model has showed the great capacity in the COVID-19 recognition in both Real and Synthetic Tests. It has obtained approximate 98% accuracy in Real Test, which can be totally believed in real-world application.

From the Fig. 4, the ROC curves generally show the coincident result with numeric metrics. The DenseNet has the largest AUC area in both Synthetic and Real Tests. It can be summarized from the AUC results of the two test sets that we synthesized a large-scale and high-quality COVID-19 dataset and constructed a reliable COVID-19 automated diagnostic model.

This experiment demonstrates that CycleGAN-synthesized COVID-19 CT images trained deep learning-based COVID-19 classification models are reliable approaches. Therefore, the generated dataset can be used to enlarge the data for deep learning training in the COVID-19 classification. Instead of collecting large clinically

relevant data, a small amount real data is enough to build a large synthesize COVID-19 chest CT images dataset for the deep learning to realize the diagnosis. This shows the potentials to solve the biomedical data-lacking problem in many deep learning trainings. The further research can focus on the fine tuning the classification model by a mixture dataset, including small amount real data and large amount synthetic data, to improve the diagnosis performance further.

4. Discussion and conclusions

This work illustrates the feasibility of the synthetic data in solving the lack of data in deep learning-based COVID-19 analysis. Recent advances in deep learning-based COVID-19 analysis are systematically reviewed, including the CT-images based classification, detection, and segmentation. These earlier researches reveal that one of the most important factors, hindering the advancement of deep learning in COVID-19 analysis, is lacking a large number of clinical COVID-19 CT images. The more data used in training the networks, the more reliable deep learning model can be obtained. Thus, big dataset preparation is fatal in intelligent automatic diagnosis implementation in COVID-19. Using the strategy of CycleGAN style learning, this study establishes a public COVID-19 CT image dataset from lung cancer CT image, comprising 1186 synthetic COVID-19 CT images. The combination of those converted CT images with additional annotations and labels turns the dataset into a rich resource for the development and the evaluation of deep learning algorithms for COVID-19 CT image analysis.

Using this generated COVID-19 CT dataset, five widely used deep learning models, including VGG16, ResNet-50, Incep-

tion_ResNet_v2, Inception_v3, and DenseNet-169 are trained, and their performances are compared with the synthetic or real COVID-19 CT images. All models achieve the excellent results with over than 90% classification accuracy in both synthetic and real dataset. Although the classification models have never been learned on the real COVID-19 CT images, synthetic and real test set have similar results, which demonstrate they have similar features and the real COVID-19 can be replaced by our generated dataset. Among those models, DenseNet_169 model achieve best performance with all the metrics beyond 97% in both test datasets, which is much better than previously reported deep learning-based classification models. This also proves that the GGO characteristics can be learned very well in CycleGAN model.

Besides, CycleGAN-based image convert model learns the characteristics of the entire lung and generates a large infection area with GGO pattern. Although this is unexpectable, the phenomenon has not affected the results in evaluation of deep learning models for classification. Nevertheless, the further research is worth to conduct in order to identify the clinically relevant regions while ignoring the unrelated ones. One possible improvement is to establish attention and localization-based conversion models for lesion regions.

In summary, recent progresses in deep learning for COVID-19 chest CT image analysis are reviewed and compared. To address the major obstacles of current deep learning in COVID-19 CT image analysis, CycleGAN is utilized to generate a public COVID-19 CT image dataset from lung cancer. The proposed model can synthesize large amount of COVID-19 data from a small amount of publicly available data, which can assist the development of deep learning-based COVID-19 diagnostic method. The lesson learned from the work of deep learning in lung cancer will reduce data curation and model development time dramatically for COVID-19, by relieving the shortness of high-quality labelled data and well-established model.

5. Importance

As of Feb. 2021, COVID-19 has been confirmed around 107 million people worldwide and caused over 2.3 million deaths as reported. Lung CT imaging is one of effective screening tool for COVID-19. The time-consuming and labour intensity make the accurate diagnosis particularly challenging. Deep learning shows great potential in CT image analysis. However, the major problem is the lack of public datasets with well-defined labels to allow comparisons of different deep learning models. A style transferring strategy based on CycleGAN is employed to synthesize a publicly available COVID-19 dataset from lung cancer. Various deep learning models are trained and tested on the synthesized COVID-19 dataset in order to systematically compare different models. The lesson learned from the work of deep learning in lung cancer will reduce data curation and model development time dramatically for COVID-19, by relieving the shortness of high-quality labelled data and well-established model.

Funding

CRediT authorship contribution statement

Hao Jiang: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Shiming Tang:** Investigation, Writing - original draft, Formal analysis. **Weihuang Liu:** Data curation, Formal analysis. **Yang Zhang:** Conceptualization, Investigation, Methodology, Supervision, Resources, Writing - original draft, Writing - review & editing, Validation, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Favre G, Pomar L, Musso D, Baud D. 2019-nCoV epidemic: what about pregnancies?. Lancet 2020;395:. https://doi.org/10.1016/S0140-6736(20)30311-1e40.

[2] Kupferschmidt K. 'This beast is moving very fast'. Will the new coronavirus be contained—or go pandemic?. Science 2020. https://doi.org/10.1126/science.abb1701.

[3] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020;395:689–97. https://doi.org/10.1016/S0140-6736(20)30260-9.

[4] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020;382:1199–207. https://doi.org/10.1056/NEJMoa2001316.

[5] Kanne JP. Chest CT findings in 2019 novel coronavirus (2019-NCoV) infections from Wuhan, China: Key points for the radiologist. Radiology 2020;295:16–7. https://doi.org/10.1148/radiol.2020200241.

[6] Lee S, Jung J, Park I, Park K, Kim DS. A deep learning and similarity-based hierarchical clustering approach for pathological stage prediction of papillary renal cell carcinoma. Comput Struct Biotechnol J 2020;18:2639–46. https://doi.org/10.1016/j.csbj.2020.09.029.

[7] Komura D, Ishikawa S. Machine Learning Methods for Histopathological Image Analysis. Comput Struct Biotechnol J 2018;16:34–42. https://doi.org/10.1016/j.csbj.2018.01.001.

[8] Mei X, Lee HC, Diao K yue, Huang M, Lin B, Liu C, et al. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. Nat Med 2020;26:1224–8. doi: 10.1038/s41591-020-0931-3.

[9] Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020;26:1037–40. https://doi.org/10.1038/s41591-020-0916-2.

[10] McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep learning in radiology. Acad Radiol 2018;25:1472–80. https://doi.org/10.1016/j.acra.2018.02.018.

[11] Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25:954–61. https://doi.org/10.1038/s41591-019-0447-x.

[12] Ding J, Li A, Hu Z, Wang L. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. Med Image Comput Comput Interv MICCAI 2017:559–67. https://doi.org/10.1007/978-3-319-66179-7_64.

[13] Zhu W, Liu C, Fan W, DeepLung XX. Deep 3D dual path nets for automated pulmonary nodule detection and classification. IEEE Winter Conf Appl Comput Vis WACV 2018:673–81. https://doi.org/10.1109/WACV.2018.00079.

[14] Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, et al. Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv preprint arXiv:200305037; 2020.

[15] Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology 2020;296:. https://doi.org/10.1148/radiol.2020200905200905.

[16] Xu X, Jiang X, Ma C, Du P, Li X, Lv S, et al. A deep learning system to screen novel Coronavirus disease 2019 pneumonia. Engineering 2020;6:1122–9. https://doi.org/10.1016/j.eng.2020.04.010.

[17] Ying S, Zheng S, Li L, Zhang X, Zhang X, Huang Z, et al. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. MedRxiv 2020. https://doi.org/10.1101/2020.02.23.20026930.

[18] Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Lung infection quantification of COVID-19 in CT images with deep learning. arXiv preprint arXiv:200304655; 2020.

[19] Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-Dataset : A CT Image Dataset about COVID-19. arXiv preprint arXiv:200313865; 2020.

[20] Li H, Hu Y, Li S, Lin W, Liu P, Higashita R, et al. CT scan synthesis for promoting computer-aided diagnosis capacity of COVID-19. Int. Conf. Intell. Comput. ICIC 2020:413–22. https://doi.org/10.1007/978-3-030-60802-6_36.

[21] Liu S, Georgescu B, Xu Z, Yoo Y, Chabin G, Chaganti S, et al. 3D Tomographic Pattern Synthesis for Enhancing the Quantification of COVID-19. arXiv preprint arXiv:200501903; 2020.

[22] Jiang Y, Chen H, Loew MH, Ko H. COVID-19 CT Image Synthesis with a Conditional Generative Adversarial Network. IEEE J Biomed Heal Informatics 2020:1–1. https://doi.org/10.1109/JBHI.2020.3042523.

[23] Setio AAA, Traverso A, de Bel T, Berens MSN, van den Bogaard C, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Med Image Anal 2017;42:1–13. https://doi.org/10.1016/j.media.2017.06.015.

[24] Armato SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys 2011;38:915–31. https://doi.org/10.1118/1.3528204.

[25] Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT imaging features of 2019 novel coronavirus (2019-NCoV). Radiology 2020;295:202–7. https://doi.org/10.1148/radiol.2020200230.

[26] Li S, Yang Q, Jiang H, Cortés-Vecino JA, Zhang Y. Parasitologist-level classification of apicomplexan parasites and host cell with deep cycle transfer learning (DCTL). Bioinformatics 2020;36:4498–505. https://doi.org/10.1093/bioinformatics/btaa513.

[27] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Adv Neural Inf Process Syst NIPS 2014:2672–80. https://doi.org/10.3156/jsoft.29.5_177_2.

[28] Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans Med Imaging 2018;37:1310–21. https://doi.org/10.1109/TMI.2017.2785879.

[29] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 2018;321:321–31. https://doi.org/10.1016/j.neucom.2018.09.013.

[30] Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. 2017 Int Conf Inf Process ICIP 2017:146–57. https://doi.org/10.1007/978-3-319-59050-9_12.

[31] Kamnitsas K, Baumgartner C, Ledig C, Newcombe V, Simpson J, Kane A, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. Int Conf Inf Process ICIP 2017:597–609. https://doi.org/10.1007/978-3-319-59050-9_47.

[32] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556; 2015.

[33] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conf Comput Vis Pattern Recognit CVPR 2016:770–8. https://doi.org/10.1109/CVPR.2016.90.

[34] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. IEEE Conf Comput Vis Pattern Recognit CVPR 2015:1–9. https://doi.org/10.1109/CVPR.2015.7298594.

[35] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. IEEE Conf Comput Vis Pattern Recognit CVPR 2016:2818–26. https://doi.org/10.1109/CVPR.2016.308.

[36] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. IEEE Conf Comput Vis Pattern Recognit CVPR 2017:2261–9. https://doi.org/10.1109/CVPR.2017.243.

[37] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A system for large-scale machine learning. arXiv preprint arXiv:160508695; 2016.

[38] Zhu JY, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. IEEE Int Conf Comput Vis ICCV 2017:2242–51. https://doi.org/10.1109/ICCV.2017.244.

[39] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:14126980; 2015.

[40] Dirvanauskas D, Maskeliūnas R, Raudonis V, Damaševičius R, Scherer R. HEMIGEN: Human embryo image generator based on generative adversarial networks. Sensors 2019;19:3578. https://doi.org/10.3390/s19163578.

[41] Ayala-Rivera V, McDonagh P, Cerqueus T, Murphy L. Synthetic Data Generation using Benerator Tool. arXiv preprint arXiv:13113312; 2013.

[42] Liu Q, Dou Q, Yu L, Heng PA. MS-Net: multi-site network for improving prostate segmentation with heterogeneous MRI data. IEEE Trans Med Imaging 2020;39:2713–24. https://doi.org/10.1109/TMI.2020.2974574.