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

Hao Jiang
Harbin Institute of Technology (Shenzhen)
Shiming Tang
University of Missouri-Kansas City
Weihuang Liu
Harbin Institute of Technology (Shenzhen)
Yang Zhang (zhangyang07@hit.edu.cn)
Harbin Institute of Technology (Shenzhen) https://orcid.org/0000-0002-3503-5161

Research Article

Keywords: COVID-19; lung cancer; Chest CT image; CycleGAN; Image synthesis; Style transfer; Classification

Posted Date: August 14th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-58467/v1

Version of Record: A version of this preprint was published at Computational and Structural Biotechnology Journal on January 1st, 2021. See the published version at https://doi.org/10.1016/j.csbj.2021.02.016.

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

Hao Jiang[1,2], Shiming Tang[3], Weihuang Liu[1], Yang Zhang[1]*

[1]College of science, Harbin Institute of Technology (Shenzhen), Shenzhen, Guangdong, 518055, China.

[2] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610054, China.

[3] School of Computing and Engineering, University of Missouri-Kansas City, Missouri, US.

* To whom correspondence should be addressed. Email: zhangyang07@hit.edu.cn or yang.zhang2020@hotmail.com

Abstract

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, we build 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 CT images for COVID-19 and non-COVID-19 classification. In comparison, all models achieve excellent results (over than 90%) 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.

Keywords: COVID-19; lung cancer; Chest CT image; CycleGAN; Image synthesis; Style transfer; Classification

Key points:

- Recent advances in deep learning analysis of COVID-19 chest CT image are summarized and compared, particularly in classification performances.
- The major problem of current deep learning in COVID-19 CT image analysis is the lack of public datasets with well-defined labels to allow comparisons of different models.
- A style transferring strategy based on CycleGAN is employed to synthesize a publicly available COVID-19 dataset from lung cancer, containing 1186 chest CT images.
- Various deep learning models are trained and tested on the synthesized COVID-19 data set in order to systematically compare different deep learning models in COVID-19 CT image analysis.

Introduction

December 2019, an outbreak of pneumonia caused by a novel coronavirus occurred and has spread rapidly throughout world, with an ongoing risk of a pandemic. As of July 2020, coronavirus disease 2019 (COVID-19) has been confirmed around 11 million people worldwide and caused over 520 thousand deaths as reported. In absence of specific therapeutic drugs or vaccines for COVID-19, it is essential to diagnose this disease effectively, and immediately isolate the infected person from the healthy population. With currently reported cases and published papers, lung CT (computed tomography) imaging has been recommended as one of effective screening tool for COVID-19 pneumonia.[1,2,3,4]

CT images can be examined to identify the COVID-19 pneumonia regions with specific pattern by naked human eyes, which is easy to miss those small and lightly infective regions especially in the early stage. Therefore, the sufficiently training is necessary for radiologists to achieve an early-accurate diagnosis, which is essential not only for the prompt implementation of treatment but also the population screening and response. However, the time-consuming and difficulty in professional training leads to the lack of qualified radiologists, which makes the 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 automatically 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, should also have same feasibility in this novel pneumonia detection.[6,7] Lots of substantial evidences supporting the potential for deep learning in chest CT image analysis, particular in lung cancer analysis.[8,9,10,11] Thus, various deep learning-aided COVID-19 chest image analysis models were proposed showing high accuracy and efficacy in disease diagnosis. There are already hundreds of deep learning papers proposed for COVID-19 chest CT or X-ray images, and some of 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[12]. In the segmentation step, they train U-net to extract the lung region of interest (ROI) using 6,150 CT slices of cases with lung abnormalities (non-COVID-19) taken from a U.S based hospital. Then, a 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.[13] It is able to 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 4356 chest CT images (i.e., 1296 COVID-19, 1735 community-acquired pneumonia (CAP), and 1325 non-pneumonia) from 3322 patients. But the dataset is not open for public access. One disadvantage of this study mentioned by the authors is impossible to determine what 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.[14] The study collected a total of 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, effective pulmonary regions were extracted firstly, then a 3D CNN model was used to segment multiple image cube candidates. Next a location-attention classification 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 voted to represent the entire candidate region. Finally, the overall 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.[15]

In another way, Shan et al. introduced a VB-Net based deep learning framework with human-in-the-loop (HITL) strategy for segmentation of COVID-19 infection regions from chest CT scans.[16] 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 3 iterations of model updating.

In generally, most of 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 deeper understanding and improvement for the research community to help more patients in this special pandemic period.

Besides, building a reliable deep learning model always have the challenging in requirement of vast amounts of data for training. Since the sudden happening and short research of COVID-19, only limited amount of lung CT images of COVID-19 are available and open source. This badly hindered the development and evaluation of deep learning model for COVID-19 detection. To relieve this predicament, Zhao et al. build an open-sourced dataset, which contains 349 COVID-19 CT images from 216 patients and 463 non-COVID-19 CTs. Those CT images of COVID-19 and non-COVID-19 are extracted and collected from publications and on line 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.[17] 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 very well-study in lung cancer in past few years. These chest CT images are public and can be developed to aid the deep learning model training, especially paying more attention to the lesion-containing regions.

Inspired by the work of deep learning in lung cancer, we try to leverage the rich label information and the large number of images from lung cancer datasets by incorporating them into developing deep learning model for COVID-19 detection. With the guidance this prior knowledge from radiologist that chest imaging showing extensive consolidation and ground-glass opacity (GGO) in COVID-19 patients,[18] the GGO pattern can be seen in Figure 1A. Our algorithm also utilizes labelled information from the lung cancer images. We propose a deep-learning-based method to generate synthesized pneumonia CT images of COVID-19 from lung cancer, which 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 knowledge of expert for microscopic images recognition.[19] CycleGAN is an unsupervised learning model for image-to-image translation, which learns the image style through the competitive strategies. CycleGAN is based on GAN,[20] which is widely used in medical image processing, including reconstruction, classification, detection and segmentation.[21,22,23,24] The mutual generation of unpaired images

is an important feature for CycleGAN that 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 ground class image and applies this 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.

Materials and Methods

Dataset

Publicly available CT images of COVID-19 pneumonia and lung cancer are used in this study (Figure 1A). This COVID-19 dataset includes a total of 349 COVID-19 CT images and 397 Non-COVID-19 CT images. These CT images are extracted from 760 preprints reporting COVID-19 on medRxiv and bioRxiv. This dataset is used as priori knowledge domain to learn the ground-glass opacity (GGO) pattern in the CT image. And 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. And 1186 patches, containing lung nodules, are segmented according to the annotations.[25] For the further research, the codes and data sets that support the findings of this study are available on https://github.com/jiangdat/COVID-19.

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

We propose a dataset-driven deep learning strategy based on style transfer for generating COVID-19 CT images (Figure 1B). By 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 infected area 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, part of which is lesion or infected,

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 Figure 1C briefly illustrates that we use CycleGAN 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 another words, the GGO styles is combined in 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. The detailed explanation is followed in below paragraphs.

As shown in Figure 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 $G_C: C \to N$ and its discriminator D_N, the adversarial loss is expressed as:

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

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

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

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., $G_N(G_{\text{C}}(C)) \approx C$ (forward) and $G_C(G_{\text{N}}(N)) \approx N$ (backwards). L_{cyc} is defined as below:

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

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:

$$\arg\min_{G_C, G_N} \arg\max_{D_C, D_N} [L_{GAN}(G_C, D_N) + L_{GAN}(G_N, D_N) + L_{cyc}(G_C, G_N)$$

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

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 NonCOVID-19 are established and compared which takes the advantages of larger-size synthetic COVID-19 CT images dataset. To evaluate the generated COVID-19 dataset, some of best performing CNNs, including VGG, ResNet, Inception and DenseNet, are deployed in the classification experiments.

VGG

VGG is a deep convolutional neural network proposed by Oxford University and Google DeepMind.[26] Compared with the earlier network, it has a smaller size convolution kernel (3×3) and pooling layer (2×2). In addition, it has more layers and a wider feature map which facilitates feature extraction. In the test phase, three full connections were replaced with three convolutions.

ResNet

Compared with the previous deep networks like VGG, ResNet solved the gradient dispersion or gradient explosion problem in the process of deepening the network.[27] It proposed a model of combination of multiple residual blocks to deep the network for a better performance. The strategy of residual block achieves the fusion of multi-level features, which greatly improves the classification accuracy. In residual learning, the output of the network is:

$$y = g(z^{l+2} + a^l)$$

Where z^{l+2} represents the output of the layer $l+2$ linear activation function, it can be denoted as follows:

$$z^{l+2} = W^{l+2}a^{l+1} + b^{l+2}$$

Inception Network

Inception Network was proposed to solve the overfitting and gradient disappearance problem.[28] It utilized the multiple different-size convolution kernels to capture receptive fields of different sizes. And a bottleneck layer (1x1 convolution) is used to reduce the amount of calculation. The improved version Inception_v3 has made considerable progress in classification speed and accuracy. In addition, another subsequent improved version, Inception_ResNet_v2, was also proposed as a high-performance classification model, its core idea is to merge the jump connection in ResNet into the Inception Network to further improve the performance of the network.[29] This model is also used in classification experiments.

DenseNet

The basic idea of DenseNet is the same as ResNet, however, it established a dense connection between all the front and back layer. This network reuses features by connecting the features on channels and each layer will accept all previous layers as its additional input. The process can be defined as:

$$x^l = H([x_0, x_1, ..., x_{l-1}])$$

Where l is the number of layers, and $H(\cdot)$ is a non-linear transformer, it is a combined operation, which includes a series of BN (Batch Normalization), ReLU, Pooling and Conv. In this study, DenseNet is considered as an improved CNN model (compared with ResNet) applied in the experiments.[30]

Training and performance evaluation

The implementation of all networks is deployed on the TensorFlow framework in Ubuntu 16.04.[31] And the training hardware configure is one Tesla K40C GPU and 128-GB memory. During training COVID-19 synthesizer, we generally follow the settings in original CycleGAN: the general data pre-processing (Augmentation + Resize) is applied.[19] In order to get a more generalized model, data augmentation is used in our data processing including

scaling, flipping, cropping, and rotating, and the two input datasets are expanded to 3000 images. And then, all images are resized to 256×256 as input of the network. The network is optimized by Adam with batch size of 4, and the learning rate is set as 2×10^{-4} .[32]

To measure the performance of classification, accuracy and F1 score are presented as the metrics. F1 is a measure of a test's accuracy. It considers both the Precision and the Recall of the test to compute the score. It is defined as:

$$F_1 = 2\frac{P\times R}{P+R} = \frac{2TP}{2TP+FP+FN}$$

Where P, R are Precision and Recall, TP, FP, FN are True Positive, False Positive and False Negative.

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 $\{(x_1, y_2),(x_2, y_2),\ldots,(x_m, y_m)\}$, AUC is expressed as:

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

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

Results

Generation of synthetic COVID-19 CT images from lung cancer

To construct COVID-19 synthetic images, style transfer strategy is applied to transfer the GGO features of COVID-19 into the lung cancer CT image. And a GAN-based deep learning model is trained to achieve an optimal result. The training dataset contains 349 COVID-19 and 1186 lung cancer CT images. 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). An example of the network input image is shown in Figure 2A, where the generated CT images of COVID-19 have GGO pattern similar to ground truth, but also with annotations of nodule regions. Figure 2B, C show some zoomed-in regions of interest (ROIs) revealing further details of GGO pattern. A pretrained 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 a very good agreement with the ground truth images (COVID-19) shown in Figure 2C. Moreover, the generated image shows more GGO patterns compared to the ground truth image. This result is in line with the fact that the generated whole COVID-19 images containing GOO patterns, therefore generates more GOO pattern regions besides only lung nodule region. This factor does not affect the further networks evaluation for COVID-19 analysis as much, because the real COVID-19 CT image to the network is similar to this with a much boarder infected GOO region. 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.

Classification performance on synthetic COVID-19 CT images

In this study, the 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 synthesized COVID-19 like CT images and 1000 non-COVID-19 CT images. 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 metrices 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 Figure 3 to help the performance evaluation.

According to the results (Table 2), all models achieve the excellent results (all the average metrices 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 general 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 metrices 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 Figure 3, the ROC curves generally show the coincident result with numeric metrics. The DenseNet has the largest AUC area in both Synthetic and Real Tests. Besides, the AUC
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.

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.

By using this generated COVID-19 CT dataset, five widely used deep learning models, including VGG16, ResNet-50, Inception_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 metrices 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.

Authors' contributions

H.J. and Y.Z. collected data. H.J., S.T. and Y.Z. designed the experiments and analyzed the results. H.J. and W.L. wrote the code and trained the algorithms. Y.Z. conceptualized and supervised the project.

Funding

The work was supported by the Shenzhen Science and Technology Innovation Commission (Shenzhen Basic Research Project No. JCYJ20180306172131515). The funders of the study had no role in data collection, analysis, interpretation, or writing of the paper. The authors had not been paid to write this article by a pharmaceutical company or other agency. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflict of interests

The authors declare no conflict of interests.

References

1. Favre G, Pomar L, Musso D, et al. 2019-nCoV epidemic: what about pregnancies?. *The Lancet* 2020; 395.
2. Kupferschmidt K, Jon C. ‘This beast is moving very fast.’ Will the new coronavirus be contained—or go pandemic?. *Science* 2020.
3. Wu J T, Leung K, Leung G M, et al. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 2020, 395: 689-697.
4. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus‐Infected Pneumonia. *The New England Journal of Medicine* 2020.
5. Kanne J P. Chest CT Findings in 2019 Novel Coronavirus (2019-nCoV) Infections from Wuhan, China: Key Points for the Radiologist. *Radiology* 2020, 295: 16-17.
6. Mei X, Lee H, Diao K, et al. Artificial intelligence‐enabled rapid diagnosis of patients with COVID-19. *Nature Medicine* 2020.
7. Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine* 2020.
8. Mcbee M P, Awan O, Colucci A, et al. Deep Learning in Radiology. *Academic Radiology* 2018; 25: 1472-1480.
9. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine* 2019; 25: 954-961.

10. Ding J, Li A, Hu Z, et al. Accurate Pulmonary Nodule Detection in Computed Tomography Images Using Deep Convolutional Neural Networks. *medical image computing and computer assisted intervention* 2017;559-567 .
11. Zhu W, Liu C, Fan W, et al. DeepLung: Deep 3D Dual Path Nets for Automated Pulmonary Nodule Detection and Classification. *workshop on applications of computer vision* 2018; 673-681.
12. Gozes O, Fridadar M, Greenspan H, 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: Computer Vision and Pattern Recognition* 2020.
13. Li L, Qin L, Xu Z, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. *Radiology* 2020; 200905-200905.
14. Xu X, Jiang X, Ma C, et al. Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia. *Medical Physics* 2020.
15. Song Y, Zheng S, Li L, et al. Deep learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT images. *medRxiv* 2020.
16. Shan F, Gao Y, Wang J, et al. Lung Infection Quantification of COVID-19 in CT Images with Deep Learning. arXiv: *Computer Vision and Pattern Recognition* 2020.
17. Yang XY, He XH, et al. COVID-CT-Dataset: A CT Scan Dataset about COVID-19. *preprint arXiv* 2020.
18. Chung M, Bernheim A, Mei X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiology* 2020; 295: 202-207.
19. Zhu J, Park T, Isola P, et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *international conference on computer vision* 2020; 2242-2251.
20. Goodfellow I, Pougetabadie J, Mirza M, et al. Generative Adversarial Nets. *neural information processing systems* 2014; 2672-2680.
21. Yang G, Yu S, Dong H, et al. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging* 2018; 37: 1310-1321.
22. Fridadar M, Diamant I, Klang E, et al. GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification. *Neurocomputing* 2018; 321-331.
23. Schlegl T, Seebock P, Waldstein S M, et al. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. *international conference information processing* 2017; 146-157.
24. Kamnitsas K, Baumgartner CF, Ledig C, et al. Unsupervised Domain Adaptation in Brain Lesion Segmentation with Adversarial Networks. *international conference information processing* 2017; 597-609.
25. Setio AA, Traverso A, De Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical Image Analysis* 2017; 1-13.
26. Simonyan K, Zisserman A. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. *computer vision and pattern recognition* 2014.
27. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. *computer vision and pattern recognition* 2016; 770-778.
28. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. *computer vision and pattern recognition* 2015; 1-9.
29. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision. *computer vision and pattern recognition* 2016; 2818-2826.

30. Huang G, Liu Z, Der Maaten L V, et al. Densely Connected Convolutional Networks. computer vision and pattern recognition 2017; 2261-2269.
31. Abadi M, Barham P, Chen J, et al. TensorFlow: A system for large-scale machine learning. *preprint arXiv* 2016; 265-283.
32. Kingma D P, Ba J. Adam: A Method for Stochastic Optimization. *arXiv: Learning* 2014.

Tables and Figures

Table 1. Summary of top cited studies of deep learning-based COVID-19 analysis.

Ref	Backbone	Task	Dataset	Results
[12]	U-Net	Segmentation	6,150 CT slices of cases with lung abnormalities and their corresponding lung masks	AUC of 0.996, sensitivity of 98.2%, and specificity of 92.2%
	ResNet50	Classification	50 abnormal thoracic CT scans (slice thickness, {5,7,8,9,10}mm) of patients that were diagnosed by a radiologist as suspicious for COVID-19. Cases were annotated for each slice as normal (n=1036) vs abnormal (n=829)	
[13]	U-Net ResNet50	Classification	4356 3D chest CT exams from 3,322 patients, consisting of 1,296 COVID-19, 1,735 community-acquired pneumonia (CAP), 1,325 non-pneumonia	AUC of 0.96, sensitivity of 90%, and specificity of 96%
[14]	Location-attention ResNet	Classification	618 CT samples (219 COVID-19, 224 Influenza-A-viral-pneumonia, and 175 healthy case). 11,871 image patches (2,634 COVID-19, 2,661 Influenza-A viral pneumonia, and 6,576 healthy case).	Accuracy of 86.7%
[15]	Details Relation Extraction neural network (DRE-Net) based on ResNet50	Classification	88 COVID-19 patients with 777 CT images, 100 bacterial pneumonia patients with 505 slices, and 86 healthy people with 708 slices.	AUC of 0.99 and recall (sensitivity) of 0.93
[16]	VB-Net with human-in-the-loop (HITL) strategy	Segmentation	300 CT images from 300 COVID-19 patients were collected for validation. 249 CT images of 249 COVID-19 patients were collected from other centers for training.	Accuracy (Dice similarity coefficients between automatic and manual segmentations): 91.6%±10.0%
[17]	DenseNet-169	Classification	The 349 CT images of COVID-19 are 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.

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

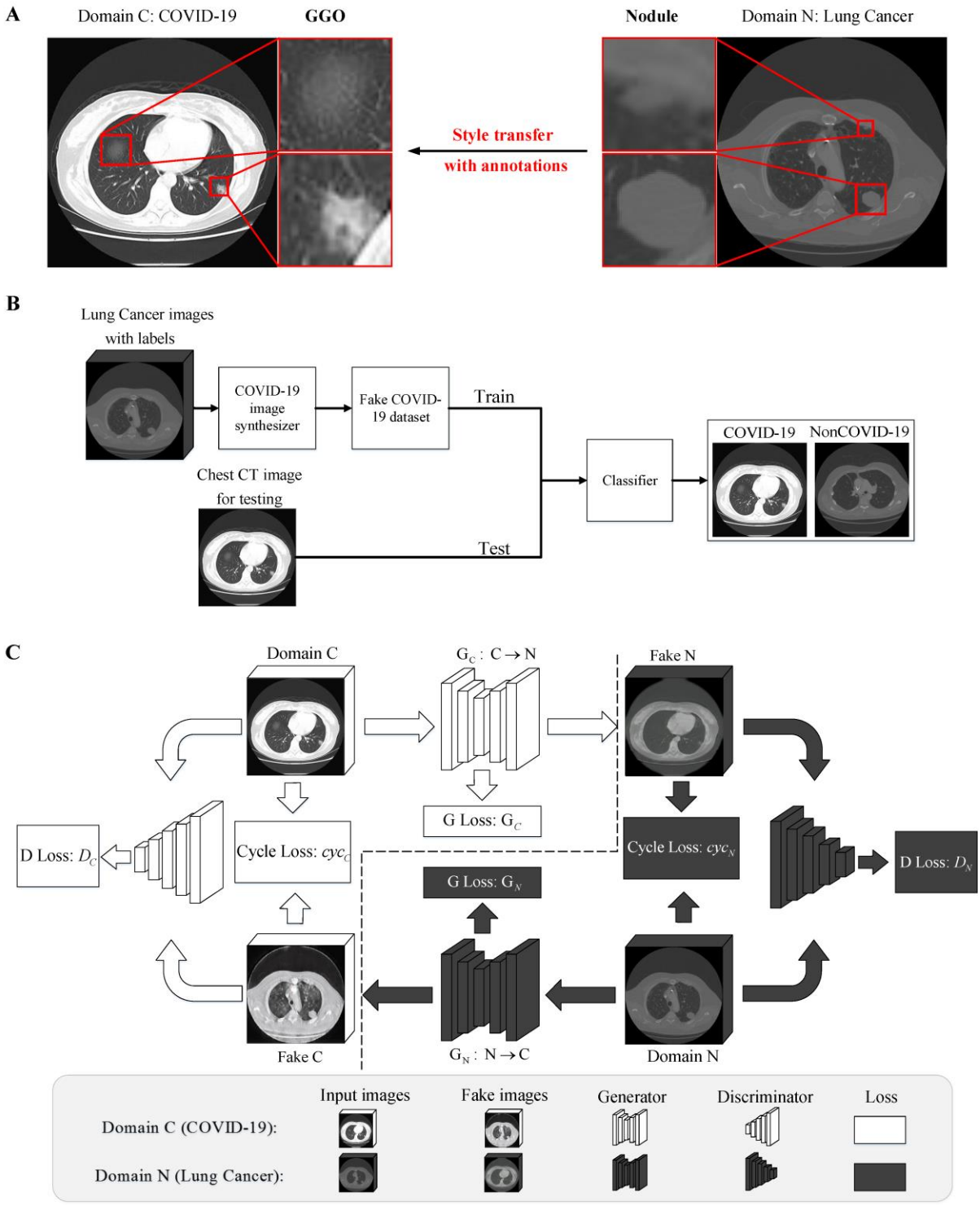

Figure 1. Overview of CycleGAN-base deep learning for COVID-19. **(A)** Representative chest CT images. Publicly available COVID-19 pneumonia images have infected areas with GGO pattern, and images of lung cancer with distinct nodules source from LUNA16. **(B)** COVID-19 analysis model based on style 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.

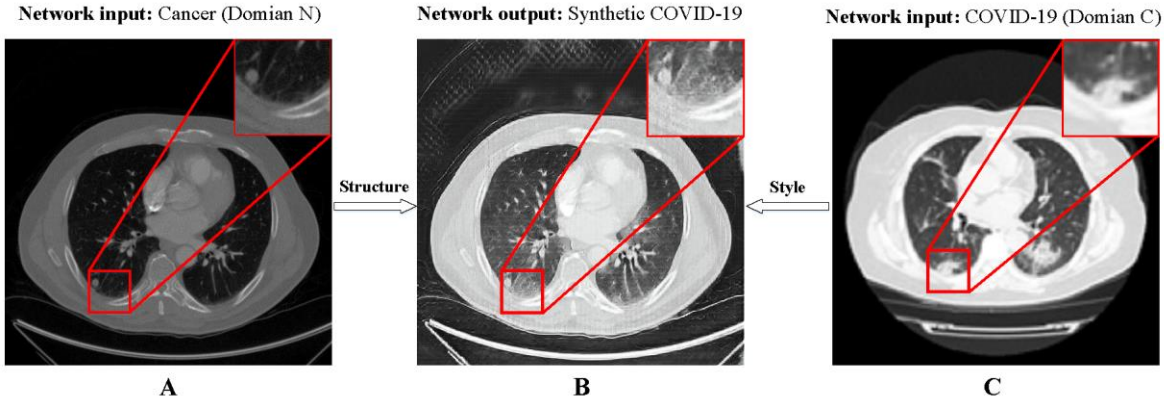

Figure 2. Deep-learning enabled CT image transformation from lung cancer to COVID-19. (A) Input lung cancer CT image. **(B)** Reconstructed image obtained using the CycleGAN based deep learning method. **(C)** Input COVID-19 CT image. Zoomed-in regions of lesion in COVID-19 and lung cancer, highlighting the success of generation of GOO pattern. Experiments are repeated through the whole lung cancer dataset, achieving similar results.

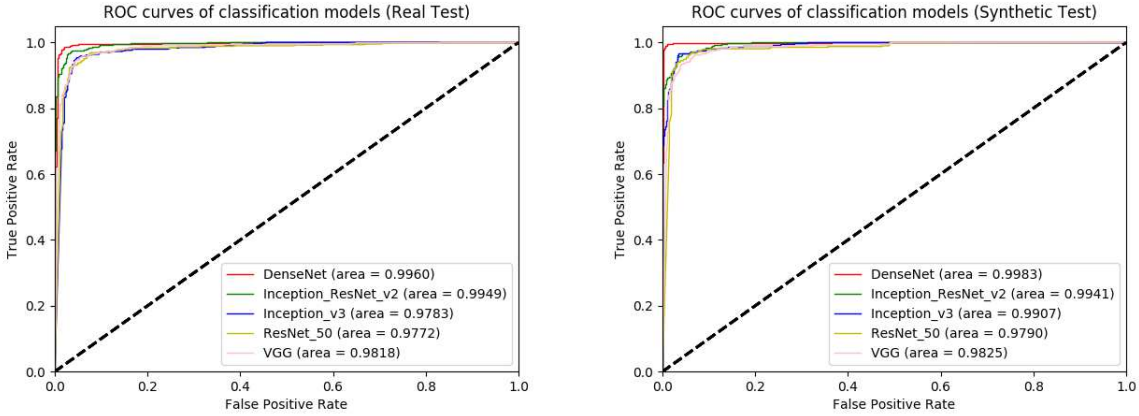

Figure 3. ROC curves of five classification models on real or synthetic dataset. The lines colored by red, green, blue, yellow, pink are the ROC curves of DenseNet_169, Inception_ResNet_v2, Inception_v3, ResNet_50 and VGG16 respectively.

Figures

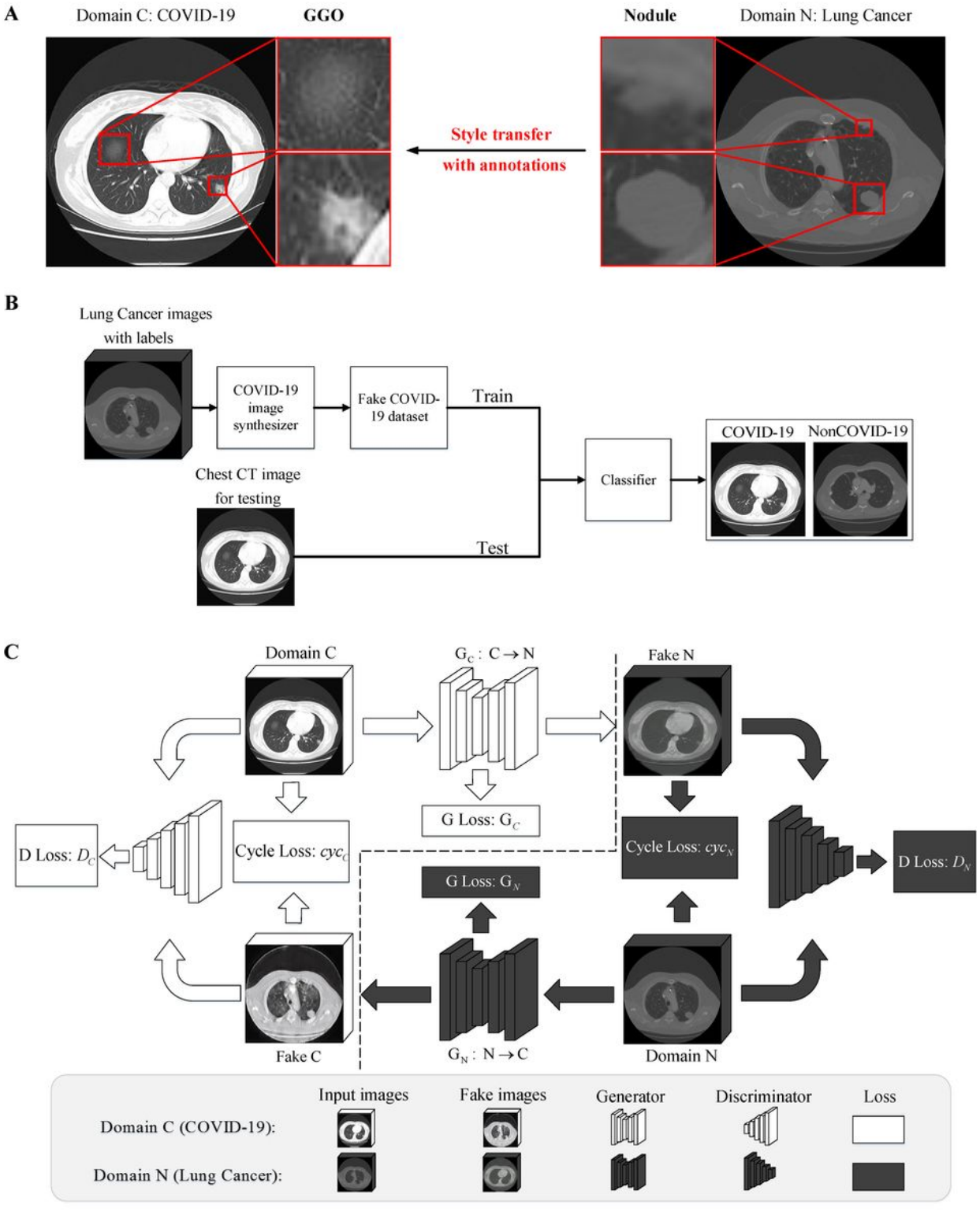

Figure 1

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.

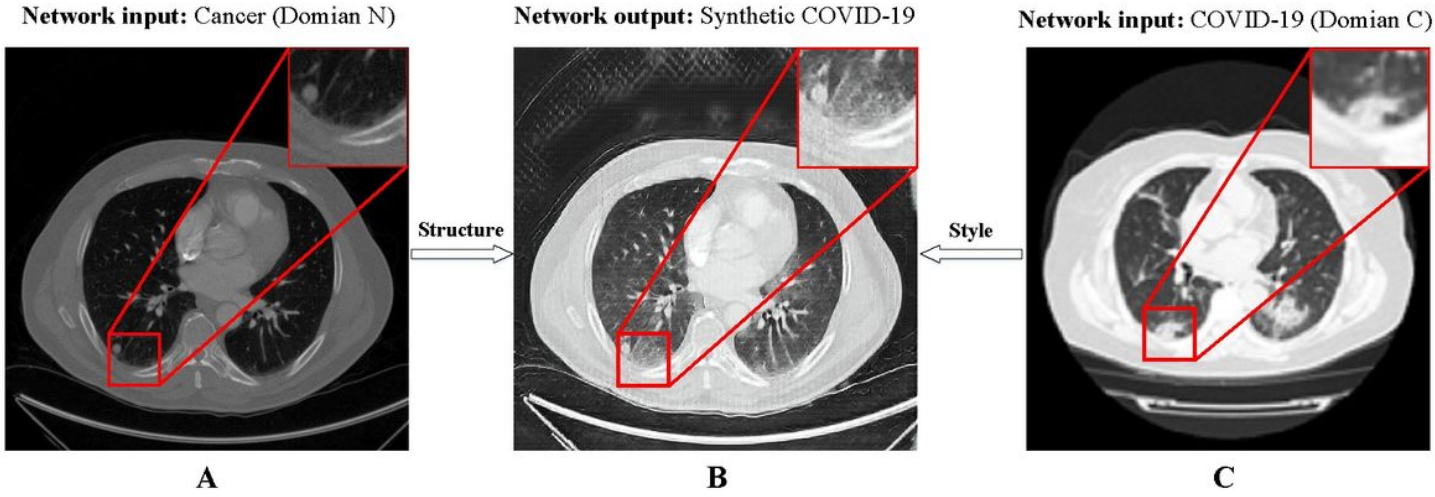

Figure 2

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, highlighting the success of generation of GOO pattern. Experiments are repeated through the whole lung cancer dataset, achieving similar results.

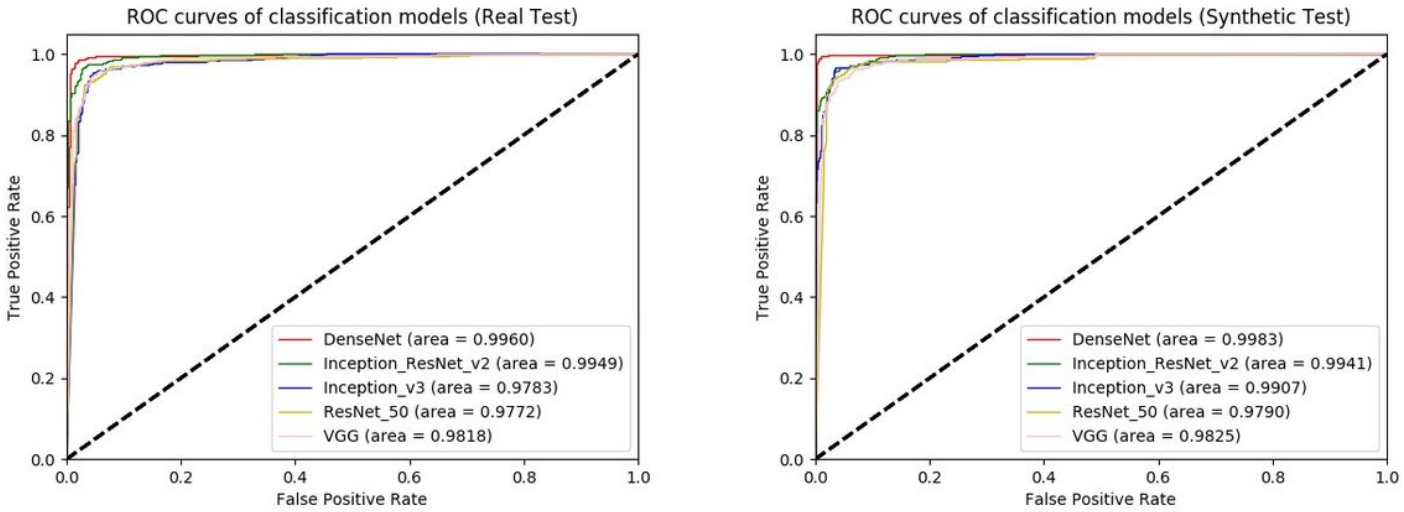

Figure 3

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.