A Weakly-Supervised Framework for COVID-19 Classification and Lesion Localization From Chest CT

Xinggang Wang, *Member, IEEE*, Xianbo Deng, Qing Fu, Qiang Zhou, Jiapei Feng, Hui Ma, Wenyu Liu, *Senior Member, IEEE*, and Chuansheng Zheng

***Abstract*— Accurate and rapid diagnosis of COVID-19 suspected cases plays a crucial role in timely quarantine and medical treatment. Developing a deep learning-based model for automatic COVID-19 diagnosis on chest CT is helpful to counter the outbreak of SARS-CoV-2. A weakly-supervised deep learning framework was developed using 3D CT volumes for COVID-19 classification and lesion localization. For each patient, the lung region was segmented using a pre-trained UNet; then the segmented 3D lung region was fed into a 3D deep neural network to predict the probability of COVID-19 infectious; the COVID-19 lesions are localized by combining the activation regions in the classification network and the unsupervised connected components. 499 CT volumes were used for training and 131 CT volumes were used for testing. Our algorithm obtained 0.959 ROC AUC and 0.976 PR AUC. When using a probability threshold of 0.5 to classify COVID-positive and COVID-negative, the algorithm obtained an accuracy of 0.901, a positive predictive value of 0.840 and a very high negative predictive value of 0.982. The algorithm took only 1.93 seconds to process a single patient's CT volume using a dedicated GPU. Our weakly-supervised deep learning model can accurately predict the COVID-19 infectious probability and discover lesion regions in chest CT without the need for annotating the lesions for training. The easily-trained and high-performance deep learning algorithm provides a fast way to identify COVID-19 patients, which is beneficial to control the outbreak of SARS-CoV-2. The developed deep learning software is available at https://github.com/sydney0zq/covid-19-detection.**

***Index Terms*— COVID-19, CT, deep learning, weak label, SARS-CoV-2, DeCoVNet.**

I. Introduction

SINCE Dec 2019, a large and increasing outbreak of a novel coronavirus was reported in Wuhan, Hubei province of China [1], [2], which can cause acute respiratory illness and even fatal acute respiratory distress syndrome (ARDS) [3]. The new coronavirus was named as SARS-CoV-2 by International Committee on Taxonomy of Viruses (ICTV) [4] and the infectious diseases infected by this coronavirus was named as Coronavirus Disease 2019 (COVID-19) by World Health Organization (WHO) [5]. The new coronavirus has been confirmed of human-to-human transmission [6], [7], and due to the massive transportation and large population mobility before the Chinese Spring Festival, this new coronavirus has spread fast to other areas in China with considerable morbidity and mortality. According to the data from the National Health Commission of the People's Republic of China [8], update till 24 o'clock of Mar 29, 2020, China has reported 82447 identified cases with SARS-CoV-2, including 3,311 death cases; 82.2% (67,801/82,447) of the identified cases came from Hubei province and identified cases in Wuhan, the very center of epidemic area of Hubei province, accounted about 73.8% (50,006/67,801) of the data in Hubei province. Moreover, COVID-19 cases outside China have been reported in more than 200 countries, areas or territories. Until to 18:00 Central European Time of Mar 29, 2020, a total of 638,146 confirmed cases with 30,039 deaths cases globally was reported according to the COVID-19 situation dashboard in the World Health Organization (WHO) website [9]. Countries with the most numbers of confirmed cases included United States of America with 103,321 cases, Italy with 92,472 cases, China with 82,356 cases, Spain with 72,248 cases, Germany with 52,547 cases, Iran with 38,309 cases, France with 37,145 cases

Manuscript received May 5, 2020; accepted May 11, 2020. Date of publication May 20, 2020; date of current version July 30, 2020. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61876212, Grant 81873919, and Grant 61733007, in part by the Fundamental Research Funds for the Central Universities under Grant 2020kfyXGYJ021, in part by the CCF-Tencent Open Research Fund, and in part by the Zhejiang Laboratory under Grant 2019NB0AB02. *(Xinggang Wang, Xianbo Deng, Qing Fu, and Qiang Zhou contributed equally to this work.) (Corresponding author: Chuansheng Zheng.)*

Xinggang Wang is with the Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China, also with the Hubei Province Key Laboratory of Molecular Imaging, Wuhan 430022, China, and also with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: xgwang@hust.edu.cn).

Xianbo Deng, Qing Fu, Hui Ma, and Chuansheng Zheng are with the Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China, and also with the Hubei Province Key Laboratory of Molecular Imaging, Wuhan 430022, China (e-mail: dengxianbo@hotmail.com; fuqing1986@yeah.net; dmahui@hotmail.com; hqzcsxh@sina.com).

Qiang Zhou, Jiapei Feng, and Wenyu Liu are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: theodoruszq@hust.edu.cn; fjp@hust.edu.cn; liuwy@hust.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2020.2995965

and so on. With the tremendously fast spread of SARS-CoV-2, it has been declared to be a Public Health Emergency of International Concern (PHEIC) by WHO on 30 January 2020 [10], and the global level of assessment of the risk of spread and the risk of impact of COVID-19 has been increased to be Very High by WHO on Feb 28, 2020 [11] which poses a great threat to the international human health.

Even though real-time reverse transcriptase polymerase chain reaction (RT-PCR) has been considered as the gold standard for SARS-CoV-2 diagnosis, the very limited supply and strict requirements for laboratory environment would greatly delay accurate diagnosis of suspected patients, which has posed unprecedented challenges to prevent the spread of the infection, particularly at the center of the epidemic areas. In contrast with it, chest computed tomography (CT) is a faster and easier method for clinical diagnosis of COVID-19 by combining the patient's clinical symptoms and signs with their recent close contact, travel history, and laboratory findings, which can make it possible for quick diagnosis as early as possible in the clinical practice. It is also effectively helpful to isolate infected patients timely and control the epidemic, especially for the severely epidemic areas. In a word, chest CT is a key component of the diagnostic procedure for suspected patients and its CT manifestations have been emphasized in several recent reports [1], [12]–[15].

In a word, accurate and rapid diagnosis of COVID-19 suspected cases at the very early stage plays a crucial role in timely quarantine and medical treatment, which is also of great importance for patients' prognosis, the control of this epidemic, and the public health security. During Wuhan's COVID-19 outbreak period, a large number of patients, including suspected cases, identified cases and follow-up cases, were needed to undergo chest CT to observe the changes and severity extent of pulmonary pneumonia, which had caused a tremendous burden to professional medical staffs; their severe shortage was also a major difficult to conquer the epidemic. Chest CT, especially high-resolution CT (HRCT) could detect small areas of ground-glass opacitity (GGO). In the initial stage of COVID-19 pneumonia, lung findings on chest CT may present with small, subpleural and peripheral GGO [16], which would consume much more time than the larger-involved and diffusive GGO and/or consolidation patterns. At the current situation, any missed cases would continue to cause COVID-19 spread. So, it has posed great challenge to our radiologists with such a tremendous amount of work as well as high diagnostic accuracy. Moreover, radiologists' visual fatigue would heighten the potential risks of missed diagnosis for some small lesions. Thus, developing an artificial intelligence (AI) method for computer-aided COVID-19 diagnosis was very helpful to radiologists.

Deep learning, as the core technology of the rising AI in recent years, has been reported with significantly diagnostic accuracy in medical imaging for automatic detection of lung diseases [17]–[19]. It surpassed human-level performance on the ImageNet classification task with one million images for training in 2015 [20], showed dermatologist-level performance on classifying skin lesions in 2017 [21] and obtained very impressive results for lung cancer screening in 2019 [17]. However, most deep learning based methods for disease diagnosis requires to annotate the lesions, especially for disease diagnosis in CT volumes. Annotating lesions of COVID-19 costs a huge amount of efforts for radiologists, which is not acceptable when COVID-19 is spreading fast and there are great shortages for radiologists. Thus, performing COVID-19 diagnosis in a weakly-supervised manner is of great importance. One of the simplest labels for COVID-19 diagnosis is the patient-level label, i.e., indicating the patient is COVID-19 positive or negative. Therefore, the aim of current study was to investigate the potential of a deep learning model for automatic COVID-19 diagnosis on chest CT volumes using the weak patient-level label. Technically, we obtain a high performance COVID-19 diagnosis system by training a lung segmentation network using ground-truth masks obtained via an unsupervised method and designing an effective lightweight 3D residual network (ResNet) with a progressive classifier for COVID-19 classification and weakly-supervised lesion localization. Our deep learning solution requires minimal expert annotation and is easy to train, which is very helpful to rapidly develop AI software for COVID-19 diagnosis at this critical situation to counter this outbreak globally.

In [22], a comprehensive review of AI for COVID-19 is presented. Compared with existing COVID-19 methods, we have the following advantages. First, we trained the lung segmentation model using masks generated from an unsupervised learning method. Second, we proposed a weakly-supervised COVID-19 lesion detection by combining deep learning activation regions and unsupervised connected component activation regions, which is the first work performs weakly-supervised COVID-19 lesion localization. Third, the proposed COVID-19 classification network is lightweight and effective; in experiments, we had compared our classifier with different sophisticated deep learning classifiers and our classifier performed significantly better than the others in terms of both computation cost and classification performance.

II. Material and Methods

A. Patients

This retrospective study was approved by Huazhong University of Science and Technology ethics committee, patient consent was waived due to the retrospective nature of this study.

Between Dec. 13, 2019 to Feb. 6, 2020, we searched unenhanced chest CT scans of patients with suspected COVID-19 from the picture archiving and communication system (PACS) of radiology department (Union Hospital, Tongji Medical College, Huazhong University of Science and Technology). Finally, 540 patients (mean age, 42.5 $\pm$ 16.1 years; range, 3-81 years, male 226, female 314) were enrolled into this study, including 313 patients (mean age, 50.7 $\pm$ 14.7 years; range, 8-81 years; male 138, female 175) with clinical diagnosed COVID-19 (COVID-positive group) and 229 patients (mean age, 31.2 $\pm$ 10.0 years; range, 3-69 years; male 88, female 141) without COVID-19 (COVID-negative group). There was no significant difference in sex between the two groups ($\chi^2 = 1.744$; $P = 0.187$), age in COVID-positive

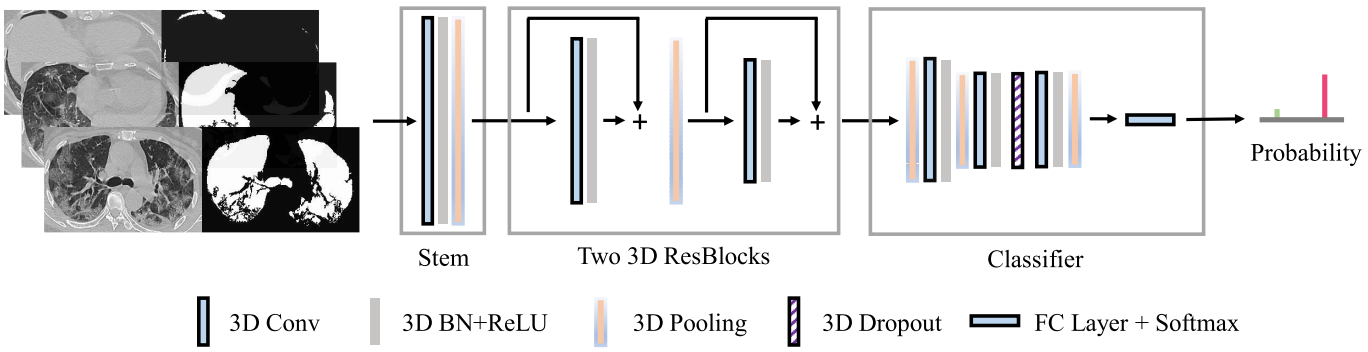

Fig. 1. Architecture of the proposed DeCoVNet. The network took a CT volume with its 3D lung mask as input and directly output the probabilities of COVID-positive and COVID-negative.

group significantly higher than that of COVID-negative group ($t = 17.09$; $P < 0.001$). The main clinical symptoms for these patients were fever, cough, fatigue, and diarrhea. Of all the patients, two were included by both groups due to the first and second follow-up CT scans. The first case (female, year 66) was diagnosed as COVID-19 negative on Jan 24, 2020, then changed into COVID-positive on Feb 6, 2020; the second case (female, year 23) was diagnosed as COVID-19 positive on Jan 24, 2020, then changed into COVID-negative on Feb 3, 2020. All the CT volumes scanned on and before Jan 23, 2020, were assigned for deep learning training, and all the CT volumes scanned after Jan 23, 2020, were assigned for deep learning testing.

B. Image Acquisition

The CT scanning of all the enrolled patients was performed on a gemstone CT scanner (GE Discovery CT750HD; GE Healthcare, Milwaukee, WI), and were positioned in a head-first supine position, with their bilateral arms raised and placed beside bilateral ears. All the patients underwent CT scans during the end-inspiration without the administration of contrast material. Related parameters for chest CT scanning were listed as follows: field of view (FOV), 36 cm; tube voltage, 100 kV; tube current, 350 mA; noise index, 13; helical mode; section thickness, 5 mm; slice interval, 5 mm; pitch, 1.375; collimation 64×0.625 mm; gantry rotation speed, 0.7 s; matrix, 512×512; the reconstruction slice thickness 1 mm with an interval of 0.8 mm; scan rage from apex to lung base; the mediastinal window: window width of 200 HU with a window level of 35 HU, and the lung window: window width of 1500 HU with a window level of -700 HU.

C. Ground-Truth Label

In the diagnosis and treatment protocols of pneumonia caused by a novel coronavirus (trial version 5) [23] which was released by National Health Commission of the People's Republic of China on Feb 4, 2020, suspected cases with characteristic radiological manifestations of COVID-19 has been regarded as the standard for clinical diagnostic cases in severely affected areas only in Hubei Province, indicating that chest CT is fundamental for COVID-19 identification of clinically diagnosed cases.

Typical CT findings for COVID-19 are also listed: multifocal small patchy shadowing and interstitial abnormalities in the early stage, especially for the peripheral area of the bilateral lungs. In the progressive period, the lesions could increase in range and in number; it could develop into multiple GGO with further infiltration into the bilateral lungs. In severe cases, pulmonary diffuse consolidation may occur and pleural effusion is rarely shown.

The combination of epidemiologic features (travel or contact history), clinical signs and symptoms, chest CT, laboratory findings and real-time RT-PCR for SARS-CoV-2 nucleic acid testing is used for the final identification of COVID-19. The medical CT reports were acquired via the electronic medical record of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology. According to the CT reports, if a CT scan was COVID-positive, its ground-truth label was 1; otherwise, the label was 0. The dataset does not contain other pneumonia and all negative cases are healthy patients.

To evaluate the performance of our algorithm for COVID-19 lesion localization, the bounding boxes of COVID-19 lesions in testing CT scans were manually annotated by a professional radiologist with 15 years of experience working in chest CT.

D. The Proposed DeCoVNet

We proposed a 3D deep convolutional neural Network to Detect COVID-19 (DeCoVNet) from CT volumes. As shown in Fig. 1, DeCoVNet took a CT volume and its 3D lung mask as input. The 3D lung mask was generated by a pre-trained UNet [24]. DeCoVNet was divided into three stages for a clear illustration in Table. I. The first stage was the network stem, which consisted of a vanilla 3D convolution with a kernel size of $5 \times 7 \times 7$, a batchnorm layer and a pooling layer. The setting of the kernel size of $5 \times 7 \times 7$ follows AlexNet [25] and ResNet [26], which is helpful to preserve rich local visual information. The second stage was composed of two 3D residual blocks (ResBlocks). In each ResBlock, a 3D feature map was passed into both a 3D convolution with a batchnorm layer and a shortcut connection containing a 3D convolution that was omitted in Fig. 1 for dimension alignment. The resulted feature maps were added in an element-wise manner. The third stage was a progressive classifier (ProClf), which contained

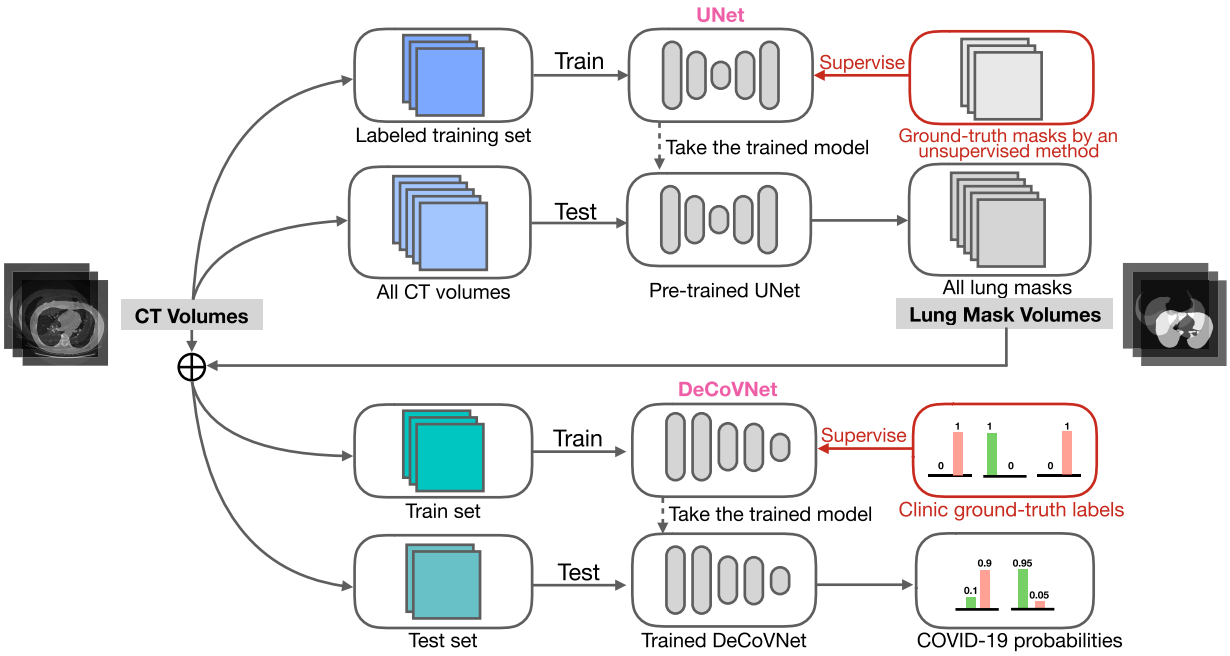

Fig. 2. Training and testing procedures. A UNet for lung region segmentation was first trained on the labeled training set using the ground-truth lung masks generated by an unsupervised learning method. Then, all CT volumes were tested by the pre-trained UNet to obtain all lung masks. Each CT volume was concatenated with its lung mask volume as the input of DeCoVNet. DeCoVNet was trained under the supervision of clinical ground-truth labels (COVID-positive and COVID-negative). Lastly, the trained DeCoVNet made predictions on the testing set.

TABLE I

DETAILED STRUCTURE OF THE PROPOSED DECOVNET. THE NUMBER AFTER THE SYMBOL "@", E.G., $5 \times 7 \times 7$, DENOTES THE KERNEL SIZE OF THE CONVOLUTION LAYER OR THE RESIDUAL BLOCK. "&" MEANS THAT THERE ARE TWO TYPES OF KERNEL SIZE IN THE RESIDUAL BLOCK. "T" DENOTES THE LENGTH OF THE INPUT CT VOLUME. THE NUMBER IN "OUTPUT SIZE" IS IN THE ORDER OF "CHANNEL, LENGTH, HEIGHT, WIDTH". THE INPUT SIZE IS $2 \times T \times 224 \times 336$

Stages	Layers	Output size
Stem	Conv3d(2, 16)@5×7×7+BN+ReLU	16×T×56×84
ResBlocks	ResBlock(16, 64)@3×1×1&1×3×3	64×T×56×84
	MaxPool3d	64×T/2×56×84
	ResBlock(64, 128)@3×1×1&1×3×3	128×T/2×28×42
Progressive classifier	AdaptiveMaxPool3d	128×16×28×42
	Conv3d(128, 64)@3×3×3+ReLU	64×16×28×42
	MaxPool3d	64×4×14×21
	Conv3d(64, 32) @3×3×3+ReLU	32×4×14×21
	Dropout3d(p=0.5)	32×4×14×21
	Conv3d(32, 32)@3×3×3+ReLU	32×4×14×21
	GlobalMaxPool3d	32×1×1×1
	FullyConnected(32, 2)	2

three 3D convolution layers and a fully-connected (FC) layer with the softmax activation function. ProClf progressively abstracts the information in the CT volumes by 3D max-pooling and finally directly output the probabilities of being COVID-positive and COVID-negative.

The 3D lung mask of an input chest CT volume helped to reduce background information and better classify COVID-19. Detecting the 3D lung mask was a well-studied issue. In this study, we trained a simple 2D UNet using the CT images in our training set. To obtain the ground-truth lung masks, we segmented the lung regions using an unsupervised learning method [27], removed the failure cases manually, and the rest segmentation results were taken as ground-truth masks. The 3D lung mask of each CT volume was obtained by testing the trained 2D UNet frame-by-frame without using any temporal information. The overall training and testing procedures of UNet and DeCoVNet for COVID-19 classification were illustrated in Fig. 2.

E. Weakly-Supervised Lesion Localization

Our idea of weakly-supervised COVID-19 lesion localization was to combine the activation regions produced by the deep classification network (i.e., DeCoVNet) and the unsupervised lung segmentation method. The method is illustrated in Fig. 3. In the right part, we inferred a few candidate lesion regions from DeCoVNet by applying the class activation mapping (CAM) method proposed in [28]. The DeCoVNet activation regions had a good recall, but they made many false positive predictions. In the left part of Fig. 3, we extracted potential COVID-19 lesion regions from the unsupervised lung segmentation results. After applying the 3d connected component (3DCC) method [27] to the CT scan, we found the lesion regions were sensitive the 3DCC algorithm, which could be utilized for lesion localization. To get the response map, we calculated the variance (including the standard deviation and the number of connected components) in a 7×7 window for each pixel as the 3DCC activation. Then, the 3DCC activation region with the largest size was selected and termed as R_{3dcc}. Lastly, the CAM activation region that had the largest overlap with R_{3dcc} was selected as the final COVID-19 lesion localization result.

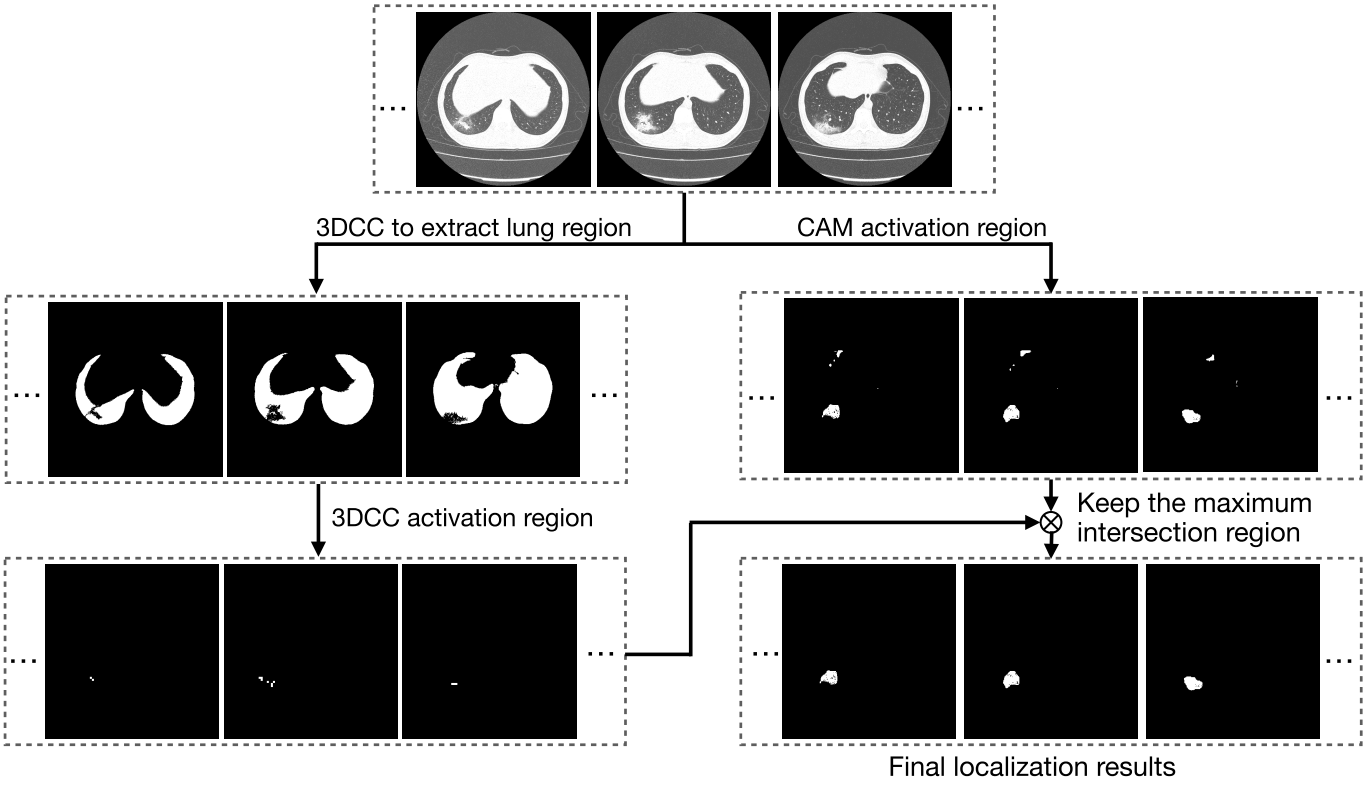

Fig. 3. The pipeline of weakly-supervised lesion localization. "3DCC" denotes the 3D connected components algorithm in [27] and "CAM" denotes the class activation mapping algorithm in [28].

F. Data Preprocessing and Data Augmentation

1) Preprocessing of 2D UNet: All the CT volumes were preprocessed in a unified manner before training the 2D UNet for lung segmentation. First, the unit of measurement was converted to the Hounsfield Unit (HU) and the value was linearly normalized from 16-bit to 8-bit (i.e., 0-255) after determining the threshold of a HU window (e.g., -1 200-600 HU). After that, all the CT volumes were resampled into a same spatial resolution (e.g., 368×368), by which the CT volumes could be aligned without the influence of the cylindrical scanning bounds of CT scanners. This step was applied to the obtained ground-truth lung masks as well.

2) Preprocessing of DeCoVNet: For each CT volume, the lung masks produced by the trained UNet formed a mask volume, then the CT volume was concatenated with the mask volume to obtain a CT-Mask volume. Finally, the CT-Mask volume was resampled into a fixed spatial resolution (e.g., 224×336) without changing the number of slices for DeCoVNet training and testing. The number of slices in the whole dataset was 141 ± 16 ranging from 73 to 250.

3) Data Augmentation: To avoid the overfitting problem since the number of training CT volumes was limited, online data augmentation strategies were applied including random affine transformation and color jittering. The affine transformation was composed of rotation ($0° \pm 10°$), horizontal and vertical translations ($0\% \pm 10\%$), scaling ($0\% \pm 20\%$) and shearing in the width dimension ($0° \pm 10°$). The color jittering adjusted brightness ($0\% \pm 50\%$) and contrast ($0\% \pm 30\%$). For each training sample, the parameters were randomly generated and the augmentation was identically applied for each slice in the sampled CT volume.

G. Training and Testing Procedures

The DeCoVNet software was developed based on the PyTorch framework [29]. Our proposed DeCoVNet was trained in an end-to-end manner, which meant that the CT volumes were provided as input and only the final output was supervised without any manual intervention. The network was trained for 100 epochs using Adam optimizer [30] with a constant learning rate of 1e-5. Because the length of CT volume of each patient was not fixed, the batch size was set to 1. The binary cross-entropy loss function was used to calculate the loss between predictions and ground-truth labels.

During the procedure of testing, data augmentation strategies were not applied. The trained DeCoVNet took the preprocessed CT-Mask volume of each patient and output the COVID-positive probability as well as COVID-negative probability. Then the predicted probabilities of all patients and their corresponding ground-truth labels were collected for statistical analysis.

The cohort for studying the COVID-19 classification and weakly-supervised COVID-19 lesion detection contained 630 CT scans collected from Dec 13, 2019 to Feb 6, 2020. To simulate the process of applying the proposed DeCoVNet for clinical computer-aided diagnosis (i.e., prospective clinical trials), we used the 499 CT scans collected from Dec 13, 2019 to Jan 23, 2020 for training and used the rest 131 CT volumes

collected from Jan 24, 2020 to Feb. 06, 2020 for testing. Of the training volumes, 15% were randomly selected for hyperparameter tuning during the training stage.

H. Statistical Analysis

COVID-19 classification results were reported and analyzed using receiver operating characteristic (ROC) and precision-recall (PR) curves. The area under the ROC curve (ROC AUC) and the area under the precision-recall curve (PR AUC) were calculated. Besides, multiple operating points were chosen on the ROC curve, e.g., the points with approximately 0.95 sensitivity (high sensitivity point) and with approximately 0.95 specificity (high specificity point). ROC AUC, PR AUC, and some key operating points were used to assess the deep learning algorithm.

To quantitatively analyse the performance of our weakly-supervised lesion localization algorithm, we followed the evaluation metric in [31] to calculate the lesion hit rate as follows. For each of the CT scans predicted as positive by DeCoVNet, we took the most confident 3D lesion mask predicted by the proposed weakly-supervised lesion localization algorithm; if the center of predicted 3D lesion mask was inside any one of the annotated boxes, it was a successful hit; otherwise, it failed to hit; finally, we calculated the hit rate by dividing the number of successful hits over all the number of true positives.

III. Experimental Results

The code for COVID-19 classification with an online web app as well as the results are available at https://github.com/sydney0zq/covid-19-detection. Training DeCoVNet on the training set which consisted of 499 CT volumes took about 20 hours (11 hours for UNet and 9 hours for DeCoVNet) and testing a CT volume costed an average of 1.93 seconds (1.80 seconds for UNet and 0.13 seconds for DeCoVNet) on an NVIDIA Titan Xp GPU.

A. COVID-19 Classification Results

For every testing CT scan, we used the trained DeCoVNet to predict its probability of COVID-19. By comparing with their binary ground-truth labels, we plotted ROC and PR curves as shown in Fig. 4 and Fig. 5 respectively. In the ROC, we obtained a ROC AUC value of 0.959. When true positive rate (TPR, i.e., sensitivity) was approximately 0.95, our model obtained a true negative rate (TNR, i.e., specificity) of 0.786; when TNR was approximately 0.95, our model obtained a TPR of 0.880; there was another operating showed that our algorithm obtained both TPR and FPR larger than 0.9, i.e., sensitivity $=$ 0.907 and specificity $=$ 0.911. On the PR curve, our model obtained a PR AUC of 0.975.

When using the threshold of 0.5 to make COVID-19 classification prediction (i.e., if the probability of COVID-19 was larger than 0.5, the patient was classified as COVID-positive, and vice versa), the algorithm obtained an accuracy of 0.901 with a positive predictive value (PPV) of 0.840 and a negative predictive value (NPV) of 0.982. By varying the probability threshold, we obtained a series of COVID-19 classification accuracy, PPV and NPV in Table II.

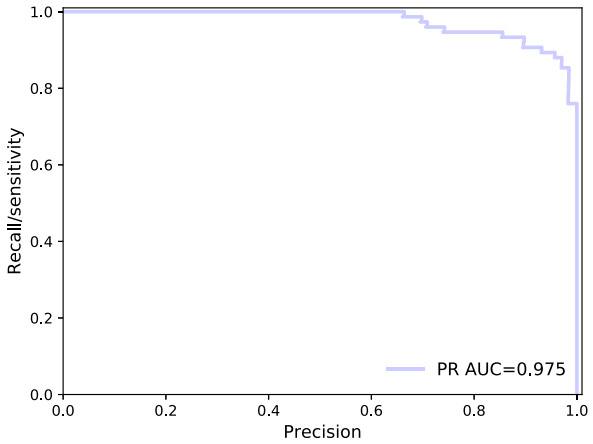

Fig. 4. COVID-19 classification results evaluated using the receiver operating characteristic curve.

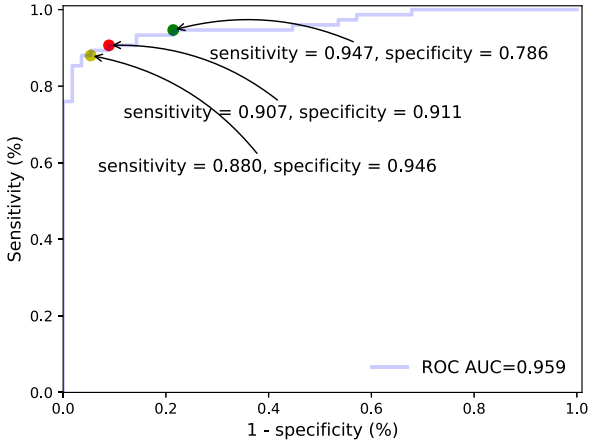

Fig. 5. COVID-19 classification results evaluated using the precision-recall curve.

TABLE II
COVID-19 Classification Statistics by Varying the Probability Thresholds. (PPV: Positive Prediction Value. NPV: Negative Prediction Value)

Threshold	Accuracy	PPV	NPV
0.1	0.893	0.933	0.839
0.2	0.908	0.880	0.946
0.3	0.908	0.867	0.964
0.4	0.901	0.840	0.982
0.5	0.901	0.840	0.982
0.6	0.878	0.800	0.982
0.7	0.870	0.787	0.982
0.8	0.855	0.760	0.982
0.9	0.847	0.733	1.000

Our data showed that the COVID-19 prediction accuracy obtained by the DeCoVNet algorithm was higher than 0.9 when the threshold ranged from 0.2 to 0.5. At the threshold setting of 0.5, there were 12 false positive predictions in total and only one false positive prediction by the algorithm in our study, indicating that the algorithm to have a very high negative predictive value.

B. Comparison to Different Classification Networks

We compared the proposed 3D DeCoVNet with different deep classifiers in Table III. Firstly, a 2D COVID-19 classification network (2DClfNet) designed following the way of [32].

TABLE III
COVID-19 CLASSIFICATION RESULTS COMPARED BETWEEN DECOVNET AND OTHER DEEP LEARNING CLASSIFIERS. "2DCLFNET" DENOTES A 2D CLASSIFICATION NETWORK DESIGNED FOLLOWING [32]. "DECOVNET W/ DROPOUT" AND "DECOVNET-FCCLF" DENOTES DECOVNET WITHOUT DROPOUT LAYER OPERATIONS AND THE MODIFIED DECOVNET BY REPLACING THE PROGRESSIVE CLASSIFIER WITH A FULLY-CONNECTED CLASSIFIER RESPECTIVELY

Method	Params (M)	FLOPs (G)	ROC AUC
2DClfNet	0.600	378.6	0.900
DeCoVNet w/ dropout	0.347	28.87	0.924
DeCoVNet-FCClf	43.1	25.81	0.893
DeCoVNet	0.347	28.87	0.959

As shown in Fig. 8, the 2DClfNet method used 2D convolutions to extract features for each slice in CT, and the slice-level features were aggregated into CT-level feature for COVID-19 via a max-pooling layer. For a fair comparison, we kept the number of parameters and the number of network layers almost the same to our DeCoVNet. However, due to computations were not shared among different slices, its computation cost was much higher than our DeCoVNet. Evaluated by the floating-point operations (FLOPs) metric, it had 378.6G FLOPs, while our DeCoVNet only had 28.87G FLOPs. Our COVID-19 classification result was also much better than 2DClfNet (0.959 ROC AUC versus 0.900 ROC AUC) since our DeCoVNet can capture the information among different slices. The results confirmed the effectiveness of utilizing the 3D backbone of COVID-19 CT classification. Secondly, we performed an ablative study to verify the effectiveness of the dropout operations in DeCoVNet. The results showed that the dropout operations provided 3.5% ROC AUC performance gain. Finally, we compared the proposed progressive classifier with a fully-connected classifier (FCClf), as shown in Fig. 9. Our progressive classifier contained an adaptive max-pooing layer, a strided max-pooling layer, a global max-pooling layer, 3 convolution layers, and a fully-connected layer, while the compared FCClf used 3 fully-connected layers for COVID-19 classification according to traditional classification network design methods. The results showed that our progressive classifier saved about 3G FLOPs computation cost and 6.6% ROC AUC performance gain.

C. Comparison With Other Methods

To prove the effectiveness of our proposed method, we compared our DeCoVNet with other methods as shown in Table IV. The gray-scale histogram feature inside the obtained lung mask volume was provided for different classifiers. After adjusting the number of bins and parameters of each classifier, the best results were kept. The proposed DeCoVNet surpassed the traditional methods by at least 10% in accuracy. Without UNet which provided the obtained lung mask volume, the performance dropped about 8%. Although our DeCoVNet performed not better than human expert, it was still promising to assist in improving efficiency.

TABLE IV
COVID-19 CLASSIFICATION RESULT COMPARISON AMONG DECOVNET, OTHER METHODS AND HUMAN EXPERT

Method	ACC	Time(s)
32 bins hist. + RBF svm	72.5%	1.96
64 bins hist. + linear svm	75.6%	1.96
32 bins hist. + random forest	78.6%	1.96
DeCoVNet w/o UNet	81.7%	0.13
DeCoVNet (ours)	90.1%	1.93
Human expert[1]	97.4%	300

TABLE V
THE RESULTS OF WEAKLY-SUPERVISED LESION LOCALIZATION

3DCC activation	CAM	NormGrad	Hit Rate
	✓		35.6%
		✓	34.2%
✓			65.7%
✓	✓		68.5%
✓		✓	43.8%

D. Visualization of Classification Results

The accurate predictions (a true positive and a true negative) were presented in Fig. 6 (A-B), and erroneous predictions in Fig. 6 (C-F). In images corresponding to the true positive and the false negative, the lesions of COVID-19 were annotated by red arrows. As shown in Fig. 6 (D, E, F), the false negative predictions were made by the algorithm, and Fig. 6 (C) showed the only false positive prediction, in which the respiratory artifact had been mistaken as a COVID-19 lesion by the DeCoVNet algorithm.

To get a deeper understanding of our DeCoVNet, we visualized the learned attention region (in red color) as shown in Fig. 7. For each CT volume, we applied CAM [28] on the deep feature before the progressive classifier and the weights in the final fully connected layer, then we selected several representative CT images. It could be observed that our model indeed learned where the pneumonia occurred.

E. Weakly-Supervised Lesion Localization Results

The results of weakly-supervised lesion localization are presented in Table. V. The results were evaluated via the hit rate metric described Sec II-H. The CAM method only obtained a hit rate of 35.6%. The 3DCC activation method obtained a hit rate of 65.7%; this is a good hit rate but the region discovered by 3DCC activation is too small, which is not meaningful to radiologists. By combining CAM with the unsupervised 3DCC activation method, we obtained a hit rate of 68.5%, which was non-trivial achievement since no lesion annotations were used in our approach. We also explored a very recent weakly-supervised deep learning method, i.e., NormGrad [33]. The results of NormGrad were worst than CAM in this task.

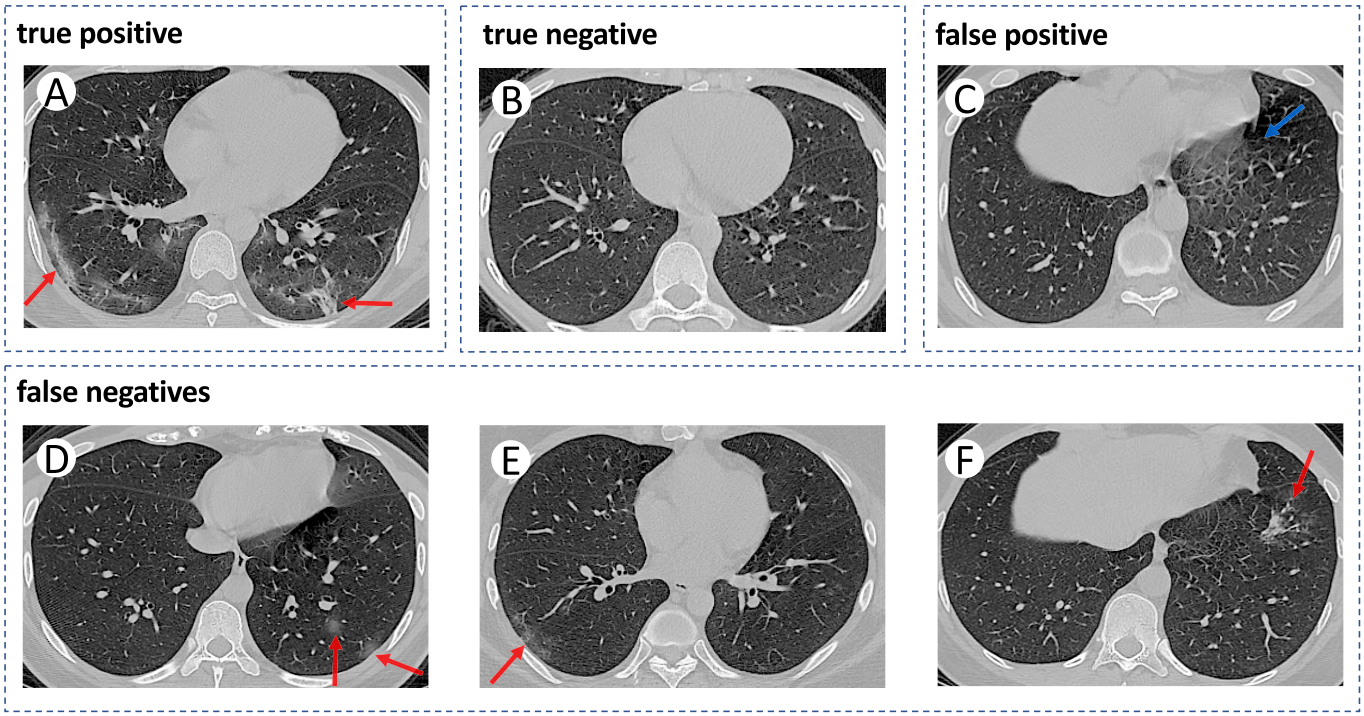

Fig. 6. Some accurate and erroneous predictions of the proposed DeCoVNet.

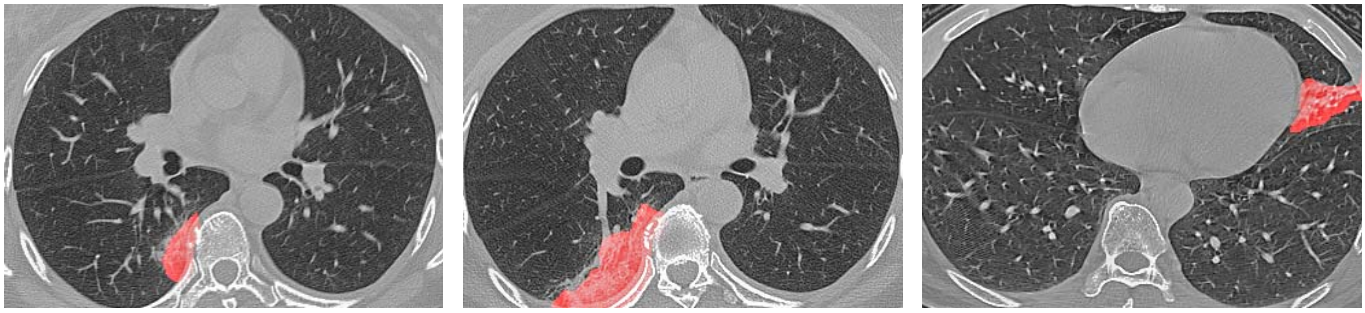

Fig. 7. Some visualizations of the learned attention region where the pneumonia occurs.

IV. Discussion

This study performs computer-aided COVID-19 diagnosis using a large number of CT volumes from the frontline hospital and very weak labels. By designing an effective weakly-supervised deep learning-based algorithm and training it on CT volumes collected before Jan 23, 2020 with only patient-level labels, the testing results on 131 CT scans collected from Jan 24, 2020, to Feb 6, 2020, were very impressive, e.g., the PR AUC value was 0.975. On the ROC curve, the algorithm obtained sensitivity and specificity values larger than 0.9, which were both clinically applicable. Compared with the concurrent work [32], we have the following advantages: (1) we propose a lung segmentation network trained using ground-truth masks generated by an unsupervised method without expert annotation. (2) Our DeCoVNet is light-weight 3D CNN which is more efficient than the CoVNet in [32] that performs ResNet-50 classification for each slides in CT. (3) We have more testing CTs (131 v.s. 68). With similar ROC AUC values for COVID-19 classification, more testing CTs illustrates more robust performance. (4) We have presented a weakly-supervised COVID-19 lesion localization algorithm.

The motivation of this study was to utilize AI to alleviate the problem of shortage of professional interpretations for CT images when the epidemic is still fast spreading. Though there were many effective applications of medical AI in previous studies [17], [34], developing AI for automatic COVID-19 diagnosis was still a challenging task. Firstly, in the current emergency situation, the number of enrolled patients is relatively smaller compared with that used in previous studies [17], [34]; and patients enrolled in our study were clinically diagnosed cases with COVID-19, because the majority of them did not undergo the nucleic acid testing due to the sudden outbreak and limited medical resource in such a short time period. Secondly, the lesions of COVID-19 in CT volumes were not labeled by radiologists and only patient-level labels (i.e., COVID-positive or COVID-negative) were utilized for training the AI algorithm in our study. Thirdly, some small infected areas of COVID-19 have the potential to be missed even by professional radiologists,

and whether it is feasible to be detected by deep learning-based 3D DCNN model remains unclear. We hypothesized to solve these problems by proposing a delicate 3D DCNN, i.e., DeCoVNet. It solved the first problem by applying extensive data augmentation on training CT volumes to obtain more training examples. The second problem was solved by regarding the COVID-19 lesion localization problem as a weakly-supervised learning problem [35]. The class activation mapping algorithm and the connected component algorithm were used for weakly-supervised lesion localization. The third problem was addressed by taking the advantages of deep learning and utilizing a pre-trained UNet for providing the lung masks to guide the learning of DeCoVNet.

The deep learning-based COVID-19 diagnostic algorithm used in our study is effective compared to recent deep learning-based computer-aided diagnosis methods. On the task of predicting the risk of lung cancer [17], the deep learning model was trained on 42290 CT cases from 14851 patients and obtained 0.944 ROC AUC. On the task of critical findings from head CT [34], the deep learning model was trained on 310055 head CT scans and obtained ROC AUC of 0.920. In our study, only 499 scans were used for training, but the obtained ROC AUC was 0.959. By comparing the data between them, it was able to find that the task of COVID-19 classification may be easier and the proposed deep learning algorithm was very powerful. As for the erroneous 12 false negative predictions in our results, the most possible explanations after we rechecked the original CT images were listed as follows: those lesions were slightly increased in CT densities, and images of those ground-glass opacities were very faint without consolidation.

Our study provided a typical and successful solution for developing medical AI for emerging diseases, such as COVID-19. While we were developing this AI, doctors in Wuhan were still extremely busy with treating a huge number of COVID-19 patients and it may be impossible for them to annotate the lesions in CT volumes in the current austere fight against this epidemic. Thanks to the weakly-supervised algorithm in this study, locations of pulmonary lesions in CT volumes are not necessary to be annotated, and radiologists' annotating efforts can be minimized, i.e., only providing patient-level labels. Therefore, developing a helpful AI tool swiftly has become possible and available in the clinical application. In the future, the burden of AI experts could be lifted significantly by automatic machine learning (AutoML) [36].

Limitations of this study: There are still several limitations in this study. First, network design and training may be further improved. For example, the UNet model for lung segmentation did not utilize temporal information and it was trained using imperfect ground-truth masks, which could be improved by using 3D segmentation networks and adopting precise ground-truth annotated by experts. Second, the data used in this study came from a single hospital and cross-center validations were not performed. Third, since this study was performed during the outbreak of COVID-19 in Wuhan and there were a great shortage of medical staff at that time, the CT data of community-acquired pneumonia (CAP) were not collected in the experiments; nevertheless, adding a new class for CAP prediction would not be difficult for an advanced deep learning classifier, which had been confirmed in [32]. Fourth, when diagnosing COVID-19, the algorithm worked in a black-box manner, since the algorithm was based on deep learning and its explainability was still at an early stage. Related work of all limitations mentioned above will be addressed in our further studies.

TABLE VI
DETAILED STRUCTURE OF THE COMPARED 2DCLFNET. THE NUMBER AFTER THE SYMBOL "@", E.G., 7 × 7, DENOTES THE KERNEL SIZE OF THE CONVOLUTION LAYER OR THE RESIDUAL BLOCK. "T" DENOTES THE LENGTH OF THE INPUT CT VOLUME. THE NUMBER IN "OUTPUT SIZE" IS IN THE ORDER OF "CHANNEL, HEIGHT, WIDTH". THE INPUT SIZE IS 2 × 224 × 336

Stages	Layers	Output size
2D Stem	Conv2d(2, 64)@7×7+BN+ReLU	64×56×84
2D ResBlocks	ResBlock(64, 128)@3×3	128×56×84
	MaxPool2d	128×28×42
	ResBlock(128, 128)@3×3	128×28×42
2D ConvLayers	Conv2d(128, 128)@3×3+ReLU	128×28×42
	Conv2d(128, 128)@3×3+ReLU	128×28×42
	GlobalMaxPool2d	128×1×1
Classifier	StackSlices	T×128×1×1
	SliceMaxPool	128×1
	FullyConnected(128, 2)	2

V. CONCLUSION

In conclusion, without the need for annotating the COVID-19 lesions in CT volumes for training, our weakly-supervised deep learning framework obtained strong COVID-19 classification performance and good lesion localization results. Therefore, our algorithm has great potential to be applied in clinical application for accurate and rapid COVID-19 diagnosis, which is of great help for the front-line medical staff and is also vital to control this epidemic worldwide.

CONTRIBUTORS

XW, XD, QF and QZ contributed equally to this study and are considered as joint first authors. CZ, XD, XW, QZ, QF and LW conceived and designed the study. QZ, XW and JF developed the algorithms with the help of clinical input from CZ, XD and QF. CZ, XD and HM collected, anonymized, and prepared the data from Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, China. CZ, XD and QF contributed to the protocol of the study. XW and QF did the statistical analysis. XW, QF and QZ wrote the initial draft. All authors subsequently critically edited the report. XD manually annotated the data. All authors read and approved the final report. CZ, XD, QF, HM, QZ and XW had full access to all data in the study.

APPENDIX
DETAILS OF THE COMPARED CLASSIFIERS

To clarify the compared deep learning based classification networks in Sec. III-B, we illustrate the 2DClfNet and

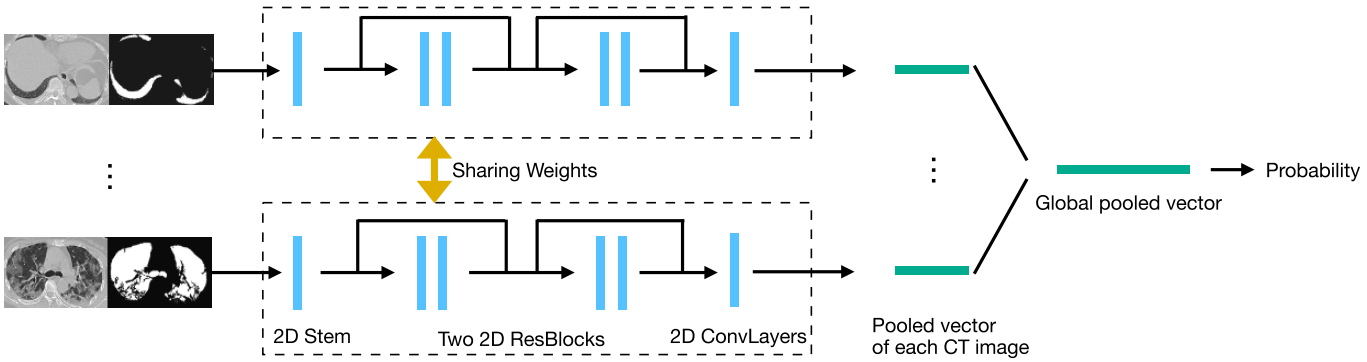

Fig. 8. Architecture of 2DClfNet for COVID-19 classification. It extracted features for each slice in CT using a 2D residual network, and the slice-level features were aggregated into CT-level feature for COVID-19 via a max-pooling layer. For the detailed structure, please refer to Table. VI.

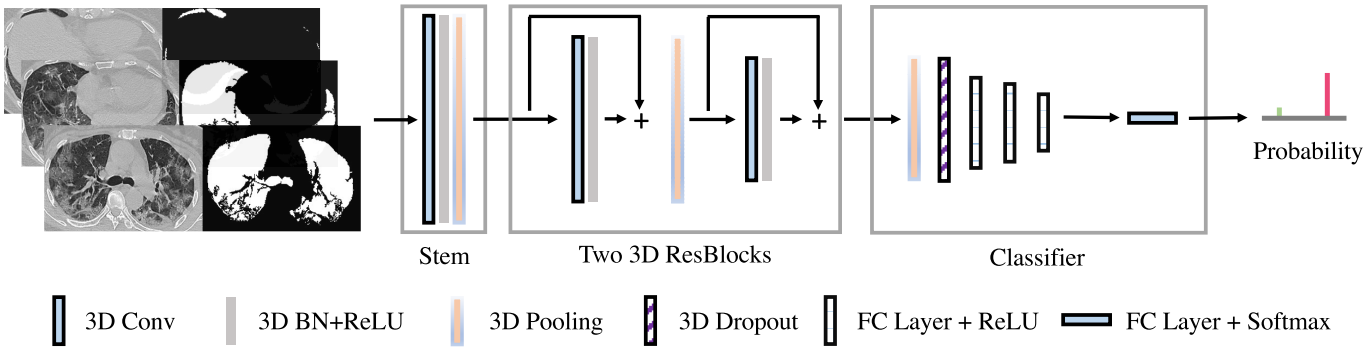

Fig. 9. Architecture of DeCoVNet-FCClf. Different from the standard DeCoVNet, DeCoVNet-FCClf used fully-connected layers for COVID-19 classification. For the detailed structure, please refer to Table. VII.

TABLE VII
DETAILED STRUCTURE OF THE COMPARED DECOVNET-FCCLF. THE NUMBER AFTER THE SYMBOL "@", E.G., $5 \times 7 \times 7$, DENOTES THE KERNEL SIZE OF THE CONVOLUTION LAYER OR THE RESIDUAL BLOCK. "&" MEANS THAT THERE ARE TWO TYPES OF KERNEL SIZE IN THE RESIDUAL BLOCK. "T" DENOTES THE LENGTH OF THE INPUT CT VOLUME. THE NUMBER IN "OUTPUT SIZE" IS IN THE ORDER OF "CHANNEL, LENGTH, HEIGHT, WIDTH". THE INPUT SIZE IS $2 \times T \times 224 \times 336$

Stages	Layers	Output size
Stem	Conv3d(2, 16)@5×7×7+BN+ReLU	16×T×56×84
ResBlocks	ResBlock(16, 64)@3×1×1&1×3×3	64×T×56×84
	MaxPool3d	64×T/2×56×84
	ResBlock(64, 128)@3×1×1&1×3×3	128×T/2×28×42
FullyConnected Classifier	AdaptiveMaxPool3d	64×4×4×6
	Dropout3d(p=0.5)	
	FlattenView	6144×1
	FullyConnected(6144, 3072)+ReLU	3072×1
	FullyConnected(3072, 1536)+ReLU	1536×1
	FullyConnected(1536, 384)+ReLU	384×1
	FullyConnected(384, 2)	2

DeCoVNet-FCClf models by simple diagrams as shown in Fig. 8 and Fig. 9. And their detailed network structures are presented in Table. VI and Table. VII.

REFERENCES

[1] C. Huang *et al.*, "Clinical features of patients infected with 2019 Novel Coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[2] H. Lu, C. W. Stratton, and Y. Tang, "Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle," *J. Med. Virol.*, vol. 92, no. 4, pp. 401–402, Apr. 2020.

[3] N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 Novel Coronavirus pneumonia in wuhan, China: A descriptive study," *Lancet*, vol. 395, no. 10223, pp. 507–513, Feb. 2020.

[4] *International Committee on Taxonomy of Viruses (ICTV)*. Accessed: Feb. 14, 2020. [Online]. Available: https://talk.ictvonline.org/

[5] *World Health Organization (WHO)*. Accessed: Feb. 15, 2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200213-sitrep-24-covid-19.pdf?sfvrsn=9a7406a4_4

[6] Q. Li *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *New England J. Med.*, 2020.

[7] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study," *Lancet*, vol. 395, no. 10225, pp. 689–697, Feb. 2020.

[8] (2020). *National Health Commission of the People's Republic of China*. Accessed: Feb. 14, 2020. [Online]. Available: http://www.nhc.gov.cn/xcs/yqtb/202002/553ff43ca29d4fe88f3837d49d6b6ef1.shtml

[9] *Coronavirus Disease (COVID-19) Situation Dashboard*. Accessed: Mar. 30, 2020. [Online]. Available: https://www.who.int/redirect-pages/page/novel-coronavirus-(covid-19)-situation-dashboard

[10] *World Health Organization (WHO)*. Accessed: Mar. 30, 2020. [Online]. Available: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)

[11] *World Health Organization (WHO)*. Accessed: Mar. 30, 2020. [Online]. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-28-february-2020

[12] W.-J. Guan *et al.*, "Clinical characteristics of 2019 Novel Coronavirus infection in China," *New England J. Med.*, vol. 382, pp. 1708–1720, 2020, doi: 10.1056/NEJMoa2002032.

[13] J. Lei, J. Li, X. Li, and X. Qi, "CT imaging of the 2019 Novel Coronavirus (2019-nCoV) pneumonia," *Radiology*, vol. 295, no. 1, 2020, Art. no. 200236.

[14] F. Song *et al.*, "Emerging 2019 Novel Coronavirus (2019-nCoV) pneumonia," *Radiology*, vol. 295, no. 1, 2020, Art. no. 200274.

[15] M. Chung *et al.*, "CT imaging features of 2019 Novel Coronavirus (2019-nCoV)," *Radiology*, vol. 295, no. 1, 2020, Art. no. 200230.

[16] P. Feng *et al.*, "Time course of lung changes on chest CT during recovery from 2019 Novel Coronavirus (COVID-19) pneumonia," *Radiology*, vol. 295, no. 3, 2020, Art. no. 200370.

[17] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Med.*, vol. 25, no. 6, pp. 954–961, Jun. 2019.

[18] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Phys. Technol.*, vol. 10, no. 3, pp. 257–273, Sep. 2017.

[19] N. Coudray *et al.*, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *Nature Med.*, vol. 24, no. 10, pp. 1559–1567, Oct. 2018.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[21] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[22] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," 2020, *arXiv:2004.02731*. [Online]. Available: http://arxiv.org/abs/2004.02731

[23] National Health Commission of the People's Republic of China. *Diagnosis and Treatment Protocols of Pneumonia Caused by a Novel Coronavirus (Trial Version 5)*. Accessed: Feb. 14, 2020. [Online]. Available: http://www.nhc.gov.cn/yzygj/s7653p/202002/d4b895337e19445f8d728fcaf1e3e13a.shtml

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Munich, Germany: Springer, 2015, pp. 234–241.

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

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[27] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-OR network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3484–3495, Nov. 2019.

[28] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. CVPR*, 2016, pp. 2921–2929.

[29] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.

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

[31] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Is object localization for free?—Weakly-supervised learning with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 685–694. [Online]. Available: https://hal.inria.fr/hal-01015140

[32] L. Li *et al.*, "Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT," *Radiol.*, 2020, Art. no. 200905, doi: 10.1148/radiol.2020200905.

[33] S.-A. Rebuffi, R. Fong, X. Ji, H. Bilen, and A. Vedaldi, "NormGrad: Finding the pixels that matter for training," 2019, *arXiv:1910.08823*. [Online]. Available: http://arxiv.org/abs/1910.08823

[34] S. Chilamkurthy *et al.*, "Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study," *Lancet*, vol. 392, no. 10162, pp. 2388–2396, Dec. 2018.

[35] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Jan. 2018.

[36] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-Weka: Combined selection and hyperparameter optimization of classification algorithms," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 847–855.