JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation

Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Chao-Wei Zhao, and Ming-Ming Cheng

***Abstract*—Recently, the novel coronavirus 2019 (COVID-19) has caused a pandemic disease over 200 countries, influencing billions of humans. To control the infection, the first and key step is to identify and separate the infected people. But due to the lack of Reverse Transcription Polymerase Chain Reaction (RT-PCR) tests, it is essential to discover suspected COVID-19 patients via CT scan analysis by radiologists. However, CT scan analysis is usually time-consuming, requiring at least 15 minutes per case. In this paper, we develop a novel Joint Classification and Segmentation (*JCS*) system to perform real-time and explainable COVID-19 diagnosis. To train our *JCS* system, we construct a large scale COVID-19 Classification and Segmentation (*COVID-CS*) dataset, with 144,167 CT images of 400 COVID-19 patients and 350 uninfected cases. 3,855 CT images of 200 patients are annotated with fine-grained pixel-level labels, lesion counts, infected areas and locations, benefiting various diagnosis aspects. Extensive experiments demonstrate that, the proposed *JCS* diagnosis system is very efficient for COVID-19 classification and segmentation. It obtains an average sensitivity of 95.0% and a specificity of 93.0% on the classification test set, and 78.3% Dice score on the segmentation test set, of our *COVID-CS* dataset. The online demo of our *JCS* diagnosis system will be available soon.**

***Index Terms*—COVID-19, Joint Diagnosis, CT Classification, CT Segmentation, COVID-19 Dataset.**

I. Introduction

CORONAVIRUS disease 2019, or COVID-19, is an epidemic disease caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). It outbreaks around the world in a short period of time, and has caused 1,439,516 confirmed cases and 85,711 confirmed deaths as of April 10th 2020. COVID-19 pushes the health systems of over 200 countries to the brink of collapse due to the lack of medical supplies and staffs, and thus has been declared as a pandemic by the World Health Organization [1]. Current *golden standard* diagnostic method for COVID-19 cases is via viral nucleic acid detection using Reverse Transcription Polymerase Chain Reaction (RT-PCR) [2]. However, the shortage of RT-PCR test kits around the world [3] makes this *golden standard* test indeed as precious as *gold*. Besides, this process needs cumbersome operations in highly controlled environment, usually taking about 4 hours [4] to receive the test results, limiting its spread popularization [5]. What's more, the false negative cases of RT-PCR tests are the potential thread to public wellness.

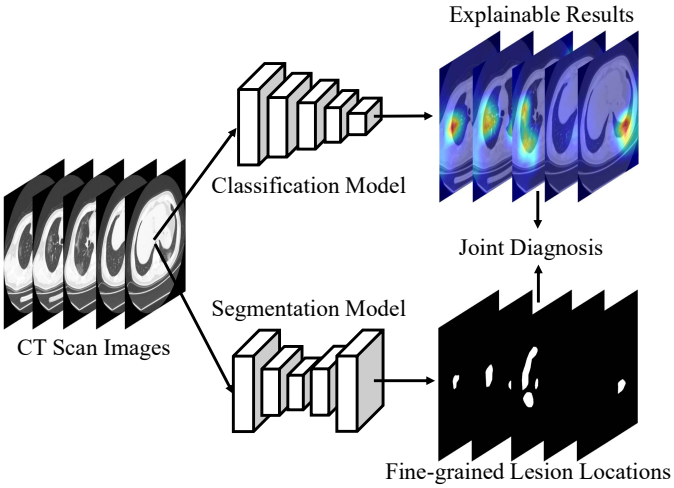


Figure 1. **Illustration of our JCS diagnosis system for COVID-19**. Our JCS system will perform the segmentation diagnosis only if the classification model reports positive COVID-19 predictions.

To hinder the terrific infection of COVID-19, medical radiology imaging is employed as an ultra-fast alternative for discovering the rapidly growing suspected or asymptomatic cases. This is based on the fact that the clinical signs of chest X-rays for most COVID-19 patients suffer from lung infection [6]. The work of [7] demonstrates that CT scan tests exhibit higher sensitivity than the RT-PCR ones. This point is further validated by [8], in which CT scans and RT-PCR tests obtain the sensitivity of 98% and 71%, respectively. However, the diagnosis duration is still the major limitation of CT scan tests: even experienced radiologists need about 21.5 minutes [9] to analyze the test results of each case.

Thanks to the powerful discriminative ability of deep convolutional neural networks (CNNs), artificial intelligence (AI) technologies are reforming the medical imaging community. Deep CNNs are usually trained on large scale datasets to demonstrate their capability. However, most of existing CT scan datasets for COVID-19 [5], [10]–[12] could not meet this demand, as they contain at most hundreds of CT images from tens of cases. Besides, most of the current COVID-19 datasets only provide the patient-level labels (*i.e.*, class labels) of indicating whether the person is infected or not, and lack of fine-grained pixel-level annotations. Thus, CNN models trained with these datasets usually neglect the valuable information for explaining the final predictions. Despite several CT scan

This work was supported in part by the Major Project for New Generation of AI under Grant No. 2018AAA0100400, NSFC (61922046), and Tianjin Natural Science Foundation (18ZXZNGX00110). (Corresponding author: M.-M. Cheng)

Y.-H. Wu is with the TKLNDST, College of Computer Science, Nankai University, and also with the InferVision. (E-mail: wuyuhuan@mail.nankai.edu.cn)

S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, and M.-M. Cheng are with the TKLNDST, College of Computer Science, Nankai University. (E-mail: shgao@mail.nankai.edu.cn, meijie0507@gmail.com, csjunxu@nankai.edu.cn, dengpfan@gmail.com, cmm@nankai.edu.cn)

C.-W. Zhao is with the InferVision. (E-mail: zchaowei@infervision.com)

diagnosis systems [8], [13]–[17] have been established for testing the suspected COVID-19 cases, most of them suffer from two drawbacks: 1) they are trained on small scale datasets and thus not robust enough for versatile COVID-19 infections; 2) they perform classification based on the black box CNNs, while lacking the explainable transparency to assist the doctors during the medical diagnosis.

To largely alleviate the above-mentioned drawbacks, in this work, we 1) construct a large scale *COVID-CS* dataset with both patient-level and pixel-level annotations and 2) propose a *J*oint *C*lassification and *S*egmentation (*JCS*) based diagnosis system, to provide explainable diagnosis results for medical staffs fighting with COVID-19. Specifically, we utilize the collected *COVID-CS* dataset that contains thousands of CT images from hundreds of COVID-19 cases to train our *JCS* system for better diagnosis performance. As illustrated in Figure 1, our *JCS* diagnosis system firstly identifies the suspected COVID-19 patients by a classification model, and provide the diagnosis explanations via activation mapping techniques [18]. Then, our system is feasible to discover the locations and areas of the COVID-19 infection in lung radiography via fine-grained image segmentation techniques. With the explainable classification results and corresponding fine-grained lesion segmentation, our *JCS* system largely simplifies and accelerates the diagnosis process for radiologists or other medical staffs. As shown in Table II, our *JCS* system needs only 19.0 seconds for each case, much faster than the RT-PCR tests and CT scan analysis by experienced radiologists.

In summary, our contributions are mainly three-folds:

- **We construct a new large scale COVID-19 dataset**, called *COVID-CS*, which contains 3,855 fine-grained pixel-level labeled CT images from 200 patients, 64,771 patient-level annotated CT images from 200 COVID-19 patients and 75,541 CT images of 350 uninfected cases.
- **We develop a novel COVID-19 diagnosis system** to perform Joint explainable Classification and accurate lesion Segmentation (*JCS*), showing clear superiority over previous systems.
- On our *COVID-CS* dataset, **our *JCS* system achieves 95.0% sensitivity and 93.0% specificity on COVID-19 classification, and 78.3% Dice score on segmentation**, surpassing previous state-of-the-art segmentation methods.

The remaining paper is organized as follows. In §II, we briefly summarize the related works. In §III, we present our *COVID-CS* dataset with our labeling procedures in detail. In §IV, we introduce the developed diagnosis system for recognizing and analysing the COVID-19 cases. Extensive experiments are conducted in §V to evaluate the performance of our system on COVID-19 recognition, with in-depth analysis. §VI concludes this work.

II. Related Works

A. Existing Accessible COVID-19 Datasets

Currently, over a million people are infected by COVID-19. But their CT scans are usually private and could not be publicly accessed. To facilitate the development of diagnosis systems, several COVID-19 related datasets are publicly released by researchers around the world. A summary of the these datasets is provided in Table I.

Table I
SUMMARY OF DIFFERENT DATASETS (UPDATED ON 2020/4/10).

Dataset	Date	Link	Type	#Images	#Cases
PLXR [11]	2020/03/23	Link	X-rays	98	70
8023Dataset [10]	2020/03/25	Link	X-rays	229*	-
CTSeg [12]	2020/03/28	Link	CT	110	60
COVID-CT [5]	2020/03/30	Link	CT	746*	-
***COVID-CS* (Ours)**	2020/04/12	-	CT	>**144K**†	**750**

*: The number is reported from the authors' GitHub repository.
†: Among our dataset, 3,855 images of 200 positive cases are pixel-level annotated, 64,771 images of the other 200 positive cases are patient-level annotated, and the rest 75,541 images are from the 350 negative cases.

Table II
AVERAGE TIME OF COVID-19 DIAGNOSIS BY DIFFERENT METHODS.

Method	RT-PCR	CT Radiologist	*JCS*
Time	∼4 h [4]	21.5 min [9]	19.0 s

One X-ray dataset from Cohen *et al.* [10] contains overall 122 frontal view X-rays, including 100 images of COVID-19 cases, 11 SARS images and 11 other pneumonia images. The COVID-CT dataset from [5] has 746 CT scan images, with 349 images from COVID-19 patients and 397 from non-COVID-19 cases. All the images in these datasets are collected from public websites and/or COVID-19 related papers on medRxiv, bioRxiv, and journals, *etc*. CTs containing COVID-19 abnormalities are selected by reading the figure captions in the papers. Some other resources of COVID-19 dataset are PLXR [11] and CTSeg [12], which contains 98 and 110 CT scan images cases, respectively. These datasets are in a small scale and lack of diversity, since they often contain at most hundreds of images from tens of cases. To fully exploit the power of deep CNNs, it is extremely essential to construct a large scale dataset for the training of deep CNNs in accurate and robust COVID-19 systems.

B. COVID-19 Diagnosis Systems

Most of current medical imaging systems are developed for common diseases that exist for many years, *e.g.*, tuberculosis [19]. These developed systems can be directly modified to attenuate the COVID-19 outbreak. The doctors find that the chest X-rays of COVID-19 patients exhibiting certain abnormalities in the radiography. Based on ResNet-50 [20], COVID-ResNet [21] is proposed to differentiate three different types COVID-19 infections from the normal pneumonia individuals. On 1531 chest X-ray images, Zhang *et al.* proposed a deep anomaly detection system for COVID-19 screening, achieving 96.0% sensitivity and 70.65% specificity. Yang *et al.* [22] proposed a system to evaluate the images of 102 volunteers, with a sensitivity of 83.3% and specificity of 94.0%. The system developed by Li *et al.* [23] identifies 78 COVID-19 patients with a sensitivity of 82.6% and a specificity of 100.0% by using the axial and coronal-view of lung CT severity index (CTSI). Chung *et al.* [14] confirmed via collected from 21 patients that, visual inspection helps to identify the COVID-19

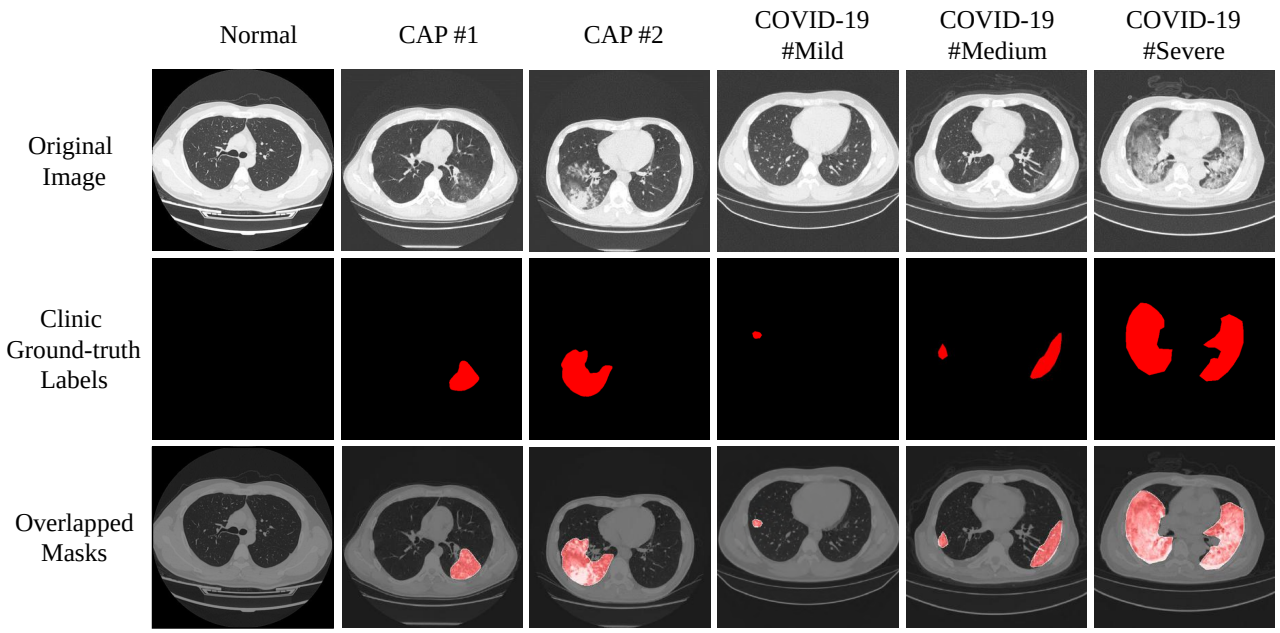


Figure 2. **Examples of our *COVID-CS* dataset**, including CT scan images and labels of a normal person (1^{st} column), two community acquired pneumonia (CAP) cases (2^{nd} and 3^{rd} columns), and three COVID-19 patients from mild to severe ($4^{th} \sim 6^{th}$ columns).

cases and predict the severity via the overall lung total severity score (LTSS). Bernheim *et al.* [15] analyzed the 121 COVID-19 patients, and carried on a visual check by experienced radiologist to categorize them as early, intermediate and late cases. Wang *et al.* [16] found that the COVID-19 disease will be severe during 6-11 days from the infection, based on a study on 366 CT scans of 90 patients. Shi *et al.* [17] developed an imaging assisted diagnosis procedure to detect the COVID-19 caused pneumonia. Fang *et al.* [8] examined 81 patients by procedure based on the CTSI, and obtained a sensitivity of 98.0%, contrast to the sensitivity of 71.0% by RT-PCR. Zhou *et al.* [24] implemented the examination using the non-contrast CTSI of 62 COVID-19 patients, confirming that the CT assisted evaluation shows better detection accuracy in progressive stage confirmed to the early stage. Despite their success on small set of samples, these COVID-19 diagnosis systems have not been tested by large scale samples, and could not provide useful diagnostic evidence during their diagnostic inference.

As far as we know, the work of [25] is the only one that extracts infected region via pixel-level segmentation. But the segmentation is performed via the watershed transform techniques [26] with coarse results and limited accuracy. In this work, we propose a diagnosis system by integrating learning based classification and segmentation networks, to provide explainable diagnostic evidence for doctors and improve the user-interactive aspects of the diagnosis process.

C. Deep Classification and Segmentation Methods

Ever since the release of ImageNet dataset [27], deep convolutional neural networks (CNNs) are becoming the workhorse for image classification tasks with improving performance. Representative deep classifiers, *e.g.*, AlexNet [28], VGGNet [29], ResNet [20], DenseNet [30], and Res2Net [31], have been widely employed as the feature extractors for other computer vision tasks, such as image segmentation [32], salient object detection [33], face recognition [34], aerial images analysis [35], feature matching [36], and image restoration [37], *etc*. Despite the impressive representation ability of these classifiers, the classification process is in a black box, providing no explanation of the predicted results.

Image segmentation tackles the problem of pixel-level predictions. Semantic segmentation aims to distinguish the stuffs from each other [38]. Representative work in this area include the DeepLab [39]–[41] and the MobileNet [42]. Instance segmentation focuses on discriminating foreground objects in the image [31]. Panoptic segmentation [43] integrates the semantic-level and instance-level segmentation, and considers both stuff-level and object-level predictions. UNet [44] is a widely employed network for medical image segmentation analysis. It is further extended to 3D U-Net [45], TernausNet [46], and UNet++ [47] with promising performance on versatile image segmentation tasks. In this work, we develop a novel COVID-19 diagnosis system by integrating deep-based image classification and segmentation techniques.

III. Our *COVID-CS* Dataset

Data acts as a basis role in the deep-based AI diagnosis system. Currently, there are few publicly available COVID-19 datasets with either large scale samples or fine-grained pixel-level labeling. To fill in this gap, we construct a new COVID-19 Classification and Segmentation (***COVID-CS***) dataset. In this section, we present the data collection, professional labeling and statistics of our dataset. Fig. 2 shows some examples of our *COVID-CS* dataset. Fig. 4 presents diverse information in the segmentation set of our *COVID-CS* dataset.

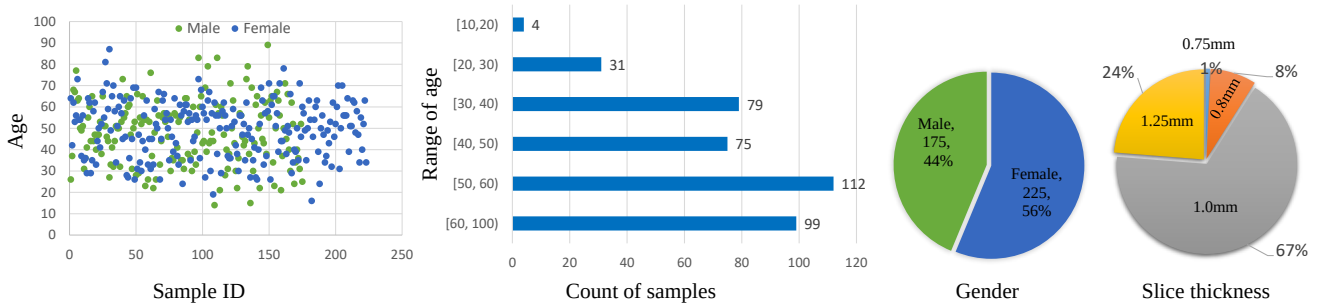


Figure 3. **The age, gender, and slice thickness distribution** of the COVID-19 patients in our *COVID-CS* dataset. Zoom in for details.

Table III
THE CT SCANNERS AND NUMBERS OF POSITIVE CASES.

Manufacturer	Product Name	#Cases
GE Medical Systems	Revolution CT	1
GE Medical Systems	LightSpeed VCT	6
GE Medical Systems	Discovery CT750 HD	12
GE Medical Systems	BrightSpeed	12
Toshiba	Aquilion ONE	33
GE Medical Systems	LightSpeed16	64
United Imaging Healthcare	uCT 780	272

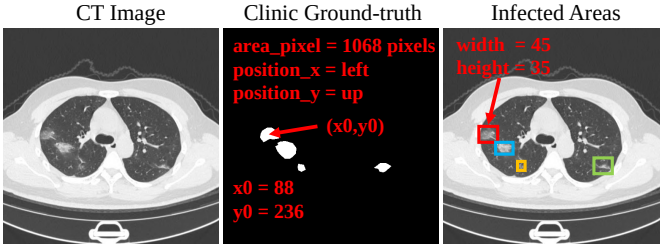


Figure 4. **Illustration of diverse information** about infected areas (in pixels), location (x0,y0), position (left, up), and width/height of infected areas in our *COVID-CS* dataset.

A. Data Collection

To protect the patients' privacy, we omit their personal information in our dataset construction. We collected 144,167 CT scan images from 750 cases, 400 of which are positive cases of COVID-19 and the other 350 cases are negative, all confirmed by RT-PCR tests. As previous studies [48] did, we do not take the community acquired pneumonia (CAP) patients (see Fig. 2) into consideration. All involved patients underwent standard chest CT scans. The CT scanners include BrightSpeed, Discovery CT750 HD, LightSpeed VCT, LightSpeed16, Revolution CT from GE Medical Systems, Aquilion ONE from Toshiba, and uCT 780 from United Imaging Healthcare. The numbers of cases from different scanners are summarized in Table III. The thickness of reconstructed CT slices ranges from 0.75mm to 1.25mm (percentage from 1.0% to 67.0%, please refer to Fig. 3 for more details).

B. Professional Labeling

We provide two aspects of labels for the collected CT scan images in our *COVID-CS* dataset, so as to implement joint classification and segmentation tasks. As mentioned above, our dataset is divided into 400 COVID-19 cases and 350 uninfected cases. For the segmentation task, we perform professional labeling through the following strategies:

- In order to save their labeling time, the radiologists only select at most 30 discrete CT scan images for each patient, in which the infections are observed for further annotation. In this step, our goal is to label every infected area with pixel-level annotations.
- To generate high-quality annotations, we first invite a radiologist to mark as many infected areas as possible based on his/her clinical experience. Then we invite another senior radiologist to refine the labeling marks several times for cross-validation. Some inaccurate labels are fixed after this step.

By implementing the above annotation procedures, we finally obtain 3,855 pixel-level labeled CT scan images of 200 COVID-19 patients with a resolution of 512×512. 64,771 CT images of the other 200 COVID-19 patients are without pixel-level annotation due to the shortage of radiologists, but such data will be used in classification test. As can be seen in the last three columns of Fig. 2, our *COVID-CS* dataset covers different levels, *i.e.*, mild, medium, and severe, of COVID-19 cases.

C. Dataset Statistics

Age. The 400 COVID-19 patients (175 males and 225 females) range from 14 to 89 years, with an average age of 48.9 years. Fig. 3 shows the distribution of ages, the counts of samples in age ranges, and the gender percentages.

Lesion count. As shown in Fig. 5 (a), we illustrate the distribution of lesion counts. We observe that the lesion count distributes from 1 to 10 in each CT scan image.

Infected areas. We plot the widths and heights of the infected areas in Fig. 5 (b). The ranges of width and height are $7 \sim 191$ and $8 \sim 271$, respectively, showing diverse distributions.

Location. We also show the relationship between each infected area and the corresponding central location (x0, y0) in Fig. 5 (c). As can be seen, the normalized infected areas range from the smallest size (35/28452 pixels) to the largest size (28452/28452 pixels). It also shows that, in our *COVID-CS* dataset, the infected areas are evenly distributed in diverse locations, which are also evenly distributed in lungs.

IV. OUR COVID-19 DIAGNOSTIC SYSTEM

Our *JCS* system consists of an explainable classification model to identify the COVID-19 infected cases and a seg-

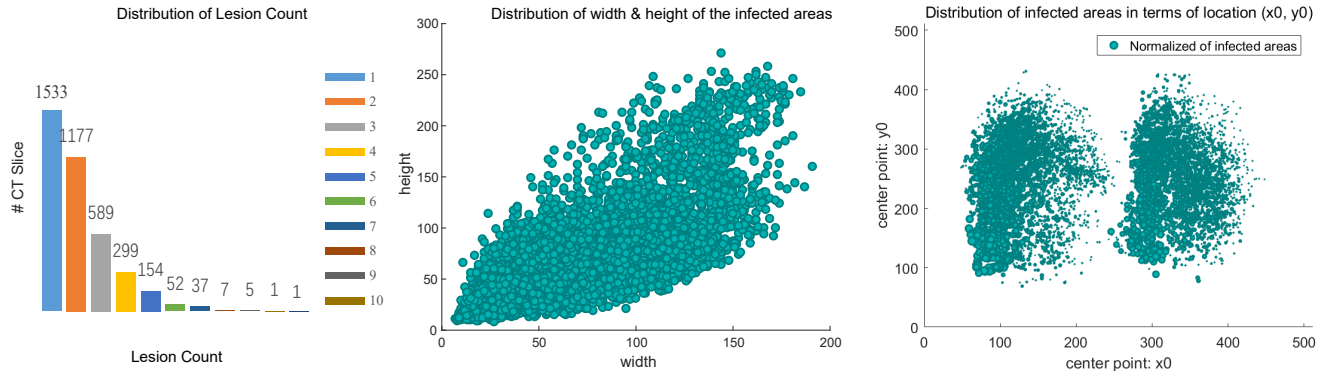


Figure 5. **Statistics of the segmentation set (200 COVID-19 cases) in our *COVID-CS* dataset**. (a) Lesion count distribution. (b) The distribution of width & height of the infected areas. (c) The relationship between the infected areas and their locations.

mentation model to discover the infected areas. The classifier is trained on large amount of images with low-cost patient-level annotations. And the segmentation model is trained with accurately annotated CT images, performing fine-grained lesion segmentation. By integrating the two models, our *JCS* system provides informative diagnosis results for COVID-19.

A. Explainable Classification Model

Owing to the strong representation ability of CNNs, the COVID-19 infections can be predicted through only patient-level supervised training. To this end, we propose a classification model to endow our *JCS* diagnosis system the capability of discriminating the COVID-19 patients.

1) Diagnosing COVID-19 via Classification: Predicting whether the suspected patient is COVID-19 positive or not is basically a binary classification task based on his/her CT scan images. The designing of novel classification model is not our focus, here we build our classifier based on the Res2Net network [31]. As a powerful network, it has the capability of multi-scale representation. The last layer is modified as a fully-connected layer with two channels, to output the probability of COVID-19 infection or not. If the probability of infected channel is larger than that of the uninfected one, the patient is diagnosed as COVID-19 positive, and vice versa. For each patient, the CT images are sent to classification model one by one. If the number of infected CT images is larger than a threshold, the patient is diagnosed as COVID-19 positive.

2) Explanation by Activation Mapping: As the diagnosis process of CNN classification is in a black box, we employ the activation mapping [18] to increase the explainable transparency of our COVID-19 diagnosis system on its predictions. The last convolutional layer of the classification network is followed by a global average pooling (GAP) layer and a fully-connected layer. Through the GAP layer, our classification model down-samples the feature size from (H, W) to $(1, 1)$, and thus lost the spatial representation ability. Through activation mapping [18], our system finds the response region of the prediction result, through the hypothesis that the gradient of regions in features before GAP layer is consistent with the evidence for prediction. The feature map before GAP layer contains both high-level semantic and location information. Each channel corresponds to the activation for different semantic cues. The activation mapping is obtained through the gradients of the predicted probability to the feature map. Specifically, given the prediction of COVID-19 branch y_p and the feature map X before GAP, the weight for the k-th channel of X is calculated as:

$$w_k = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} \frac{\partial y_p}{\partial X_{i,j}^k}, \tag{1}$$

where $X_{i,j}^k$ is the value at position (i, j) in the k-th channel of feature map X. Larger gradients in Eqn. (1) produce larger weight of the activation mapping for a certain channel. The activation mapping for a COVID-19 case is computed as:

$$AM_p = \sum_k ReLU(w_k X^k). \tag{2}$$

As shown in Fig. 8, the activation mapping accurately locates the infected areas of COVID-19 patients, providing explainable results for the prediction of our *JCS* system.

3) Alleviating Data Bias by Image Mixing: By utilizing our explainable classification model, our system can be trained only with patient-level annotation. However, since CT images are from multiple sources, the classifier may be possibly trained to overfit unwanted areas (e.g., the area outside the lesion), as been observed via the activation mapping. Therefore, we propose to utilize the image *mixing* technique [49] and help the classifier focus on the lesion areas of COVID-19 cases. The CT images from different sources and the corresponding patient-level annotations are mixed during training. Specifically, for two randomly sampled CT images x_i and x_j $(i \neq j)$ and corresponding labels $\hat{y}_i$ and $\hat{y}_j$, the newly mixed sample and the corresponding label are written as:

$$\begin{aligned} x_{ij}^m &= \lambda x_i + (1 - \lambda) x_j, \\ \hat{y}_{ij}^m &= \lambda \hat{y}_i + (1 - \lambda) \hat{y}_j, \end{aligned} \tag{3}$$

where $\lambda \in [0, 1]$ is a random number generated in Beta distribution, i.e., $\lambda \sim \text{Beta}(\alpha, \alpha)$. With mixed samples, our classification model is trained to focus more on the decisive lesion areas of COVID-19 cases, rather than the bias in data source. Also, the mixing process weakens the confidence of labels, and thus alleviating our system from overfitting.

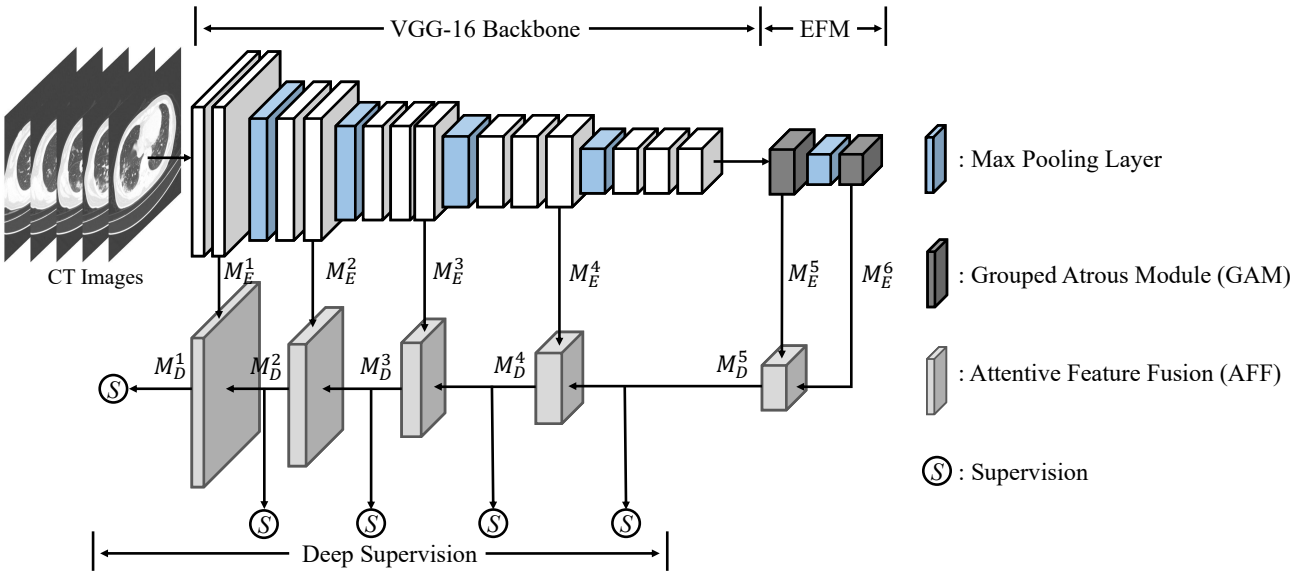


Figure 6. **Architecture of our segmentation model**. EFM indicates the Enhanced Feature Module (§IV-B2). AFF refers to the Attentive Feature Fusion strategy (§IV-B3). We apply the deep supervision to train our segmentation model.

B. Accurate Segmentation Model

Our segmentation model aims at discovering the exact lesion areas from the CT images of COVID-19 patients. Fig. 6 shows the architecture of our segmentation model.

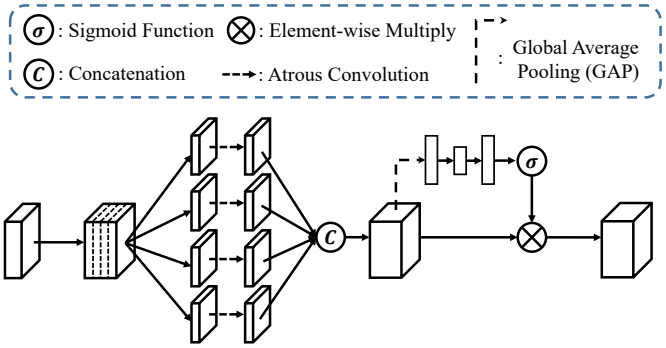
1) Encoder-Decoder Architecture: Our segmentation model consists of an encoder and a decoder.

Encoder. The encoder is based on the VGG-16 [29] backbone, without the last two fully-connected layers. It has five VGG blocks defined as $\{E_1, E_2, E_3, E_4, E_5\}$, respectively. The VGG-16 backbone is first fed with the CT images, and produces multi-scale feature maps from the last layers of the five VGG blocks. To downsize the input feature map by half, the front of each block (except the first one) is a $max\ pooling$ function with a stride of 2. The feature map produced by the block E_1 contains the finest features with the highest resolution, while the feature map by the block E_5 is coarsest with lowest resolution. To achieve better performance, we propose an Enhanced Feature Module (EFM, will be introduced in §IV-B2) for our encoder to improve its representational power. The EFM module is added after the last layer $conv5\_3$ in the block E_5. It consists of two Grouped Atrous Modules (GAM) to extract stronger feature maps with larger receptive fields. The GAM module generates an extra smaller feature map, which is of half size compared to the coarsest feature map of the VGG-16 backbone. It also enhances the representational power of the feature map produced by the block E_5. Hence, our encoder produces six levels of feature maps $\{M_E^1, M_E^2, M_E^3, M_E^4, M_E^5, M_E^6\}$, with strides of $\{1, 2, 4, 8, 16, 32\}$, respectively. As we employ a U-shape encoder-decoder architecture [50], all these six feature maps will be used in the decoder, as will be introduced later.

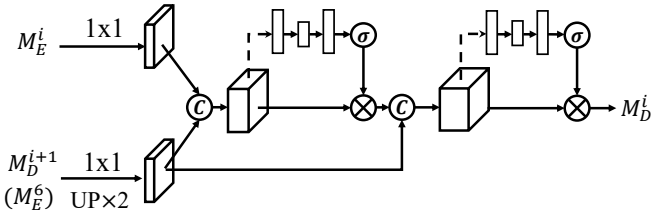
Decoder. Our decoder has five side-outputs with 5 different sizes. Here, we do not predict the side-output from the coarsest feature map with stride of 32, and thus no side-output matches the size of the coarsest feature map M_E^6. In our decoder, we propose an Attentive Feature Fusion (AFF, will be introduced in §IV-B3) strategy to aggregate the feature maps from different stages and predict the side-output of each stage. Our AFF emphasizes the significance of the top-level feature map, and utilizes the attention mechanism to filter useful features from the bottom feature map. The last output with the same resolution of the CT image input will be chosen as the final prediction.

2) Enhanced Feature Module: The proposed EFM module is added after the last layer of E_5 in the VGG-16 encoder. It consists of two sequential GAM modules, and a $max\ pooling$ function between them. As shown in Fig. 7 (a), the first layer of the GAM module is a 1×1 convolution layer to expand the channels of the feature map. Then the feature map is equally divided into 4 groups. Different from the trivial group convolution, we deploy the atrous convolution with different atrous rates to the 4 groups so as to derive a more abundant feature map with various receptive fields. To fully exploit useful features, we adopt the Squeeze-Excitation (SE) unit [51] in our network. That is, each channel of the feature map is multiplied a channel factor calculated by a SE block, which consists of two linear layers followed by a sigmoid function. We set the reduction rate in the SE block as 4. To reduce the output channels by half, we add an 1×1 convolution layer after the SE block. At last, we use a 3×3 convolution layer, in which the number of channels equals to that of input feature map, as the transition layer to the next module.

3) Attentive Feature Fusion: Traditional fusion strategy of top-down decoders [50], [52] treats the input feature maps equally. To better aggregate the feature maps, we propose an Attentive Feature Fusion (AFF) strategy. In our AFF fusion strategy, the feature map with smaller size is more valued. As shown in Fig. 7 (b), the input feature maps M_E^i and M_D^{i+1} in current stage are reduced to half size via 1×1 convolution layers. Then the reduced M_D^{i+1} is up-sampled by bilinear interpolation to output a double sized feature map. We concatenate the two outputs together, and apply the SE



(a) Grouped Atrous Module (GAM)



(b) Attentive Feature Fusion (AFF)

Figure 7. **Proposed (a) GAM and (b) AFF for the segmentation network**. In AFF, M_D^{i+1} will be replaced with M_E^6 if $i = 5$. Cubes represent three dimensional feature maps, while rectangles mean feature vectors.

block (also used in GAM) to produce an enhanced feature map. This feature map is then concatenated with the feature map of doubly up-sampled output in previous stage. After the concatenation we use another SE block to enhance the feature map again. After each SE block we use a 3×3 convolution layer, with the same number of channels as the input, as the transition layer. An 1×1 convolution layer with a single neuron will be used to predict one feature map as the side-output of the current stage.

4) Deep Supervision Loss: Although the final prediction is only from the last side-output, we apply the deep supervision strategy [53] to all side-outputs with different sizes. For each side-output, we up-sample it to the size of the ground-truth map, and compute the sum of the standard binary cross-entropy loss and the Dice loss [54] as follows:

$$\mathcal{L} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} [p_{i,j} \log(p_{i,j})] + \frac{\mathbf{P} \odot \mathbf{G}}{\|\mathbf{P}\|_1 + \|\mathbf{G}\|_1}, \quad (4)$$

where the binary cross-entropy loss is averaged among all $H \times W$ pixels, $p_{i,j}$ is the confidence score at pixel (i, j) calculated by a $sigmoid$ function, and $\odot$ means the element-wise production. $\mathbf{P}$ and $\mathbf{G}$ are predicted map and ground-truth map, respectively, while $\|\mathbf{P}\|_1$ and $\|\mathbf{G}\|_1$ denote the corresponding ℓ_1 norms.

C. Joint Diagnosis

An explainable classifier or accurate segmentation model itself could not fully implement comprehensive functions for COVID-19 diagnosis. Comparing to the segmentation model, our classifier is trained with CT images from both COVID-19 infected and uninfected cases, benefiting from more training data with lower annotation costs. Although our classifier can provide explainable lesion location of COVID-19 through activation mapping techniques, it cannot perform accurate and complete lesion segmentation. To this end, our segmentation model further provides complementary analysis by discovering the complete lesions in lung and estimate the severity of the COVID-19 patients. But annotating vast segmentation labels by experienced radiologists is prohibitively expensive. To integrate their advantages for better application, we develop a diagnosis system for COVID-19 via joint explainable classification and segmentation models. In practice, our classification model will first predict whether the CT images of a suspected case to be COVID-19 positive or not. If the prediction is positive, the suspected case is very likely to be infected by COVID-19. Then our segmentation model will be performed on the CT images for in-depth analysis, and discover the whole infected areas in each CT image.

D. Implementation Details

In our *JCS* system, the classification and segmentation models are trained separately. For the classification model, we train it with a batch-size of 256 on 4 GPUs. The CT images are resized into 224×224 for computational efficiency. We adopt the SGD optimizer with the initial learning rate of 0.1, which is divided by 10 in every 30 epochs. The classifier is trained with 100 epochs. For data augmentation, we use the random horizontal flip and random crop, and the image mixing technique [49] to alleviate the data bias. The α in Beta distribution of image mixing is set as 0.5.

For the segmentation model, the number of CT images in each mini-batch is always 4, and the size of the input CT images is unchanged as 512×512. The backbone of our segmentation model is pretrained on ImageNet [27]. The atrous rates of four atrous convolutions in two sequential GAMs are $\{1, 3, 6, 9\}$ and $\{1, 2, 3, 4\}$, respectively. The initial learning rate is 2.5×10^{-5}. We adopt the $poly$ learning rate policy that the actual learning rate will be multiplied by a factor $(1 - \frac{cur\_iter}{max\_iter})^{power}$, where the power is 0.9. The segmentation model is trained with 21000 iterations. We employ the Adam [55] optimizer, and set β_1, β_2 as 0.9 and 0.999 respectively. For data augmentation, we use random horizontal flip and random crop.

V. Experiments

A. Experimental Settings

Training/Test Protocol. For the segmentation task, our training set contains 2,794 images from 150 COVID-19 patients and the test set has 1,061 images from other 50 COVID-19 cases. For the classification task, the training set contains the 2,794 images from the 150 COVID-19 infected cases in the segmentation set. In addition, we randomly pick 150 uninfected cases with 7,500 CT images as negative cases for training. The test set contains the 64,711 images of the other randomly selected 200 infected cases and the 68,041 images from 200 uninfected cases.

Evaluation Metrics. For the classification task, we adopt the widely used metrics of specificity and sensitivity as suggested by [19]. For the segmentation task, we use two standard metrics, *i.e.*, Dice score [56] and Intersection over Union (IoU). To

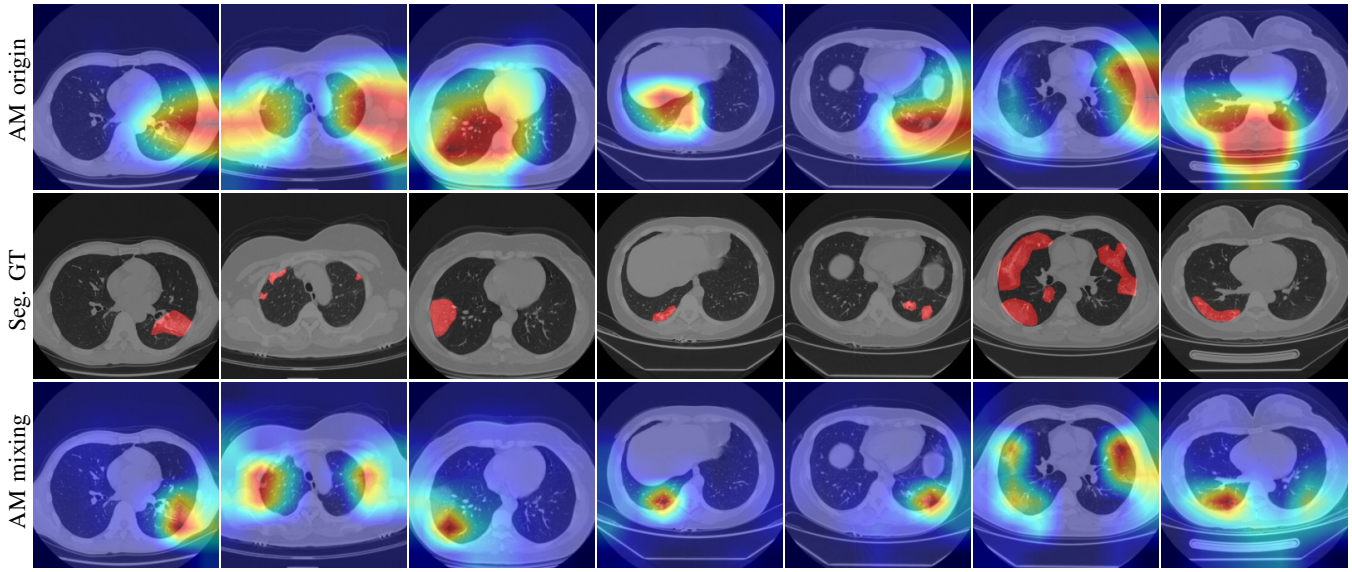


Figure 8. **Visualizations of activation mapping (AM)**. AM origin (mixing) means the AM of models trained without (with) image mixing technique [49].

provide more comprehensive evaluation, we further use the enhanced alignment measure (E_ϕ) [57].

Comparison methods. On classification task, we compare our classification model with or without image mixing technique [49]. On segmentation task, to provide in-depth evaluation of our *JCS* model, we compare it with versatile cutting-edge models, i.e., the UNet [50] for medical imaging and the DSS [33], PoolNet [58], and EGNet [59] for saliency detection.

B. Results

Performance on explainable classification. Fig. 8 shows the visualization of activation mapping of our classification model trained with or without image mixing [49]. The activation mapping (AM) of our classification model trained with random horizontal flip and random crop (i.e., the "AM origin" in Fig. 8) not only covers the lesion areas, but also presents unrelated areas. This indicates that the classification model is biased to non-lesion areas. By introducing the image mixing technique [49], the AM of our classification model provides more accurate locations of the lesion areas (the "AM mixing" in Fig. 8). During the inference, AM assists the medical staffs using our *JCS* system to judge whether the prediction is correct or not. When the number of CT images from a suspected patient is larger than a threshold, the patient is diagnosed as COVID-19 positive. Changing the threshold enables our model to achieve a trade-off between sensitivity and specificity. Table IV shows the results of the classification model under different thresholds on the test set of our *COVID-CS* dataset. One can see that our model is very robust to the changing of thresholds, and achieves a sensitivity of 95.0% and a specificity of 93.0% when the threshold is 25. However, AM could not provide accurate segmentation of lesion areas (if have). Subsequently, we further employ our segmentation model to discover the lesion areas in the CT images of COVID-19 patients.

Comparison on segmentation performance. Table V lists the quantitative comparisons of 4 cutting-edge methods and our model on segmentation. It can be seen that the proposed model achieves the best result on all three metrics. It obtains improvements of 0.078, 0.093 and 0.085 on Dice score, IoU, and E_ϕ over the second best PoolNet [58], respectively. With the multi-scale data augmentation strategy, our boosted *JCS*† obtains further improvements of 0.008, 0.013, and 0.001 on the Dice score, IoU, and E_ϕ, respectively. Besides, PoolNet [58] and EGNet [59] obtains comparable results on the three metrics. U-Net [50] performs better than DSS [33] in terms of IoU, though they are comparable on the Dice score. Fig. 9 shows the qualitative results of the comparison methods. One can see that the other competitors produce inaccurate or even wrong predictions of the lesion areas in the CT images of mild, medium and severe COVID-19 infections. But our segmentation model correctly discovers the whole lesion areas on all levels of COVID-19 infections.

To further study its stability, we perform statistical analysis of our segmentation model on our segmentation test set. Fig. 10

Table IV
SENSITIVITY AND SPECIFICITY OF OUR CLASSIFICATION MODEL UNDER DIFFERENT THRESHOLDS. WE SET THE THRESHOLD AS 25 IN THE FINAL SETTING.

Threshold	Sensitivity	Specificity
15	96.0%	91.5%
20	95.0%	92.0%
25	*95.0%*	*93.0%*
30	94.5%	93.5%

Table V
QUANTITATIVE RESULTS ON OUR SEGMENTATION TEST SET. "†" INDICATES THE MODEL WITH MULTI-SCALE TRAINING.

Methods	Publication	Dice	IoU	E_ϕ
U-Net [50]	MICCAI'15	0.651	0.541	0.797
DSS [33]	TPAMI'19	0.657	0.517	0.799
EGNet [59]	ICCV'19	0.693	0.554	0.836
PoolNet [58]	CVPR'19	0.697	0.559	0.839
JCS (Ours)	Submit'20	0.775	0.652	0.924
JCS† **(Ours)**	Submit'20	**0.783**	**0.665**	**0.925**

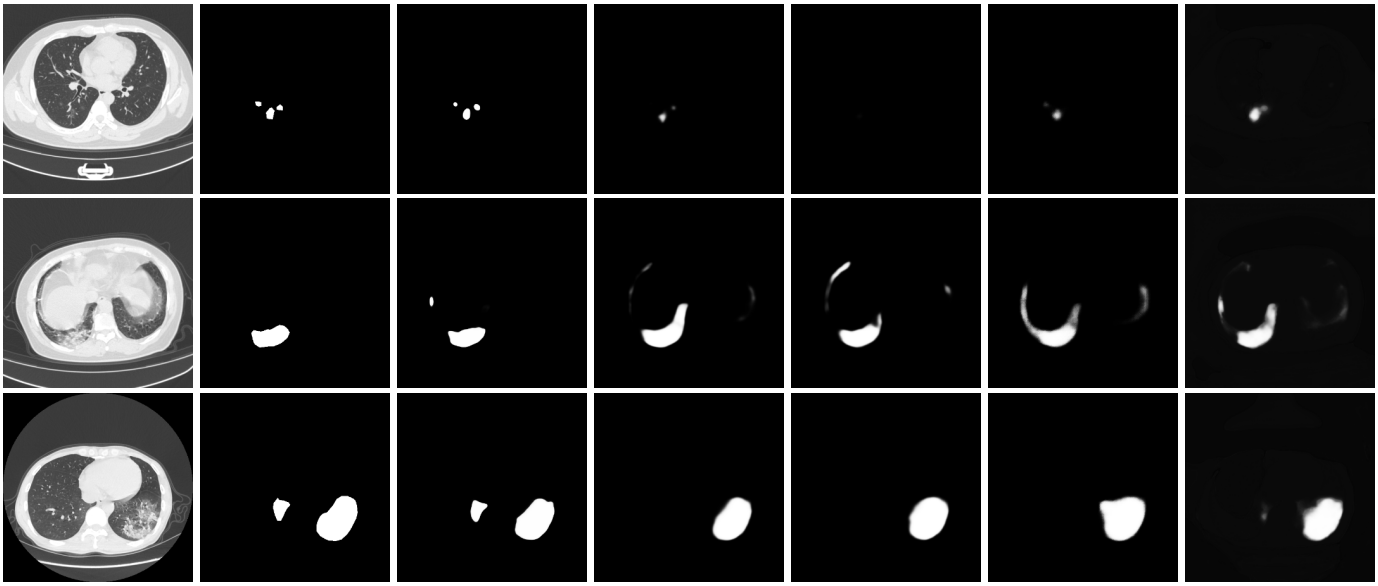


Figure 9. **Qualitative comparisons of different methods on our segmentation test set**. The first, second, and third rows show the comparison results on CT images with different lesion areas, from mild, medium, and severe COVID-19 patients, respectively.

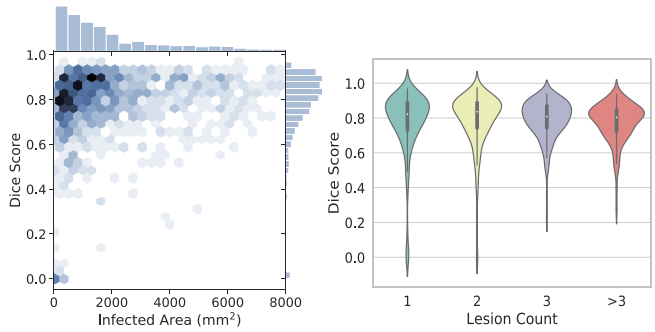


Figure 10. **Statistical analysis for our segmentation model on our segmentation test set**. (a) The relationship between the infected area of each CT image and the corresponding Dice score. (b) The relationship between the lesion count and the corresponding probability distribution of the Dice score.

(a) shows the correlation between the Dice score of our model and the infected areas of CT images. Note that the CT images with infected area $\geq 8000mm^2$ are not plotted in Fig. 10 (a), since they only occupy 1.0% of all CT images in terms of quantity. We observe that 95.0% CT images have the Dice scores in $[0.6, 1]$, while the other 3.3% CT images are with Dice scores between $[0.1, 0.6)$ and recognized as bad cases. Only 1.7% CT images suffer from Dice score of less than 0.1, and they are taken as failure cases. We also explore the relationship between the lesion count of each slice and the Dice score from a different perspective. As shown in Fig. 10 (b), the probability distribution of Dice score is little affected by the number of lesion counts in a CT image. The medium dice score is above 0.8 for 4 different cases of lesion counts, and the 95.0% confidence interval lies in $[0.5, 1]$. We also observe that the lesion count of failure cases is ≤ 2. The consistently promising performance on segmenting lesion areas and the low probability (1.7%) of failure confirm the stability of our segmentation model.

Diagnosis time. The speed test of *JCS* system is on a single RTX 2080Ti. Assuming each suspected case has 300 CT images, the classification model in *JCS* only costs about 1s to ensure whether infected. If infected, The segmentation model will spend 18.0s on fine-grained lesion segmentation. Hence, *JCS* system costs 19s for each case. Note that the complete RT-PCR test and radiologist CT diagnosis cost about 4 hours and 21.5 minutes respectively.

VI. Conclusion

To facilitate the training of strong CNN models for COVID-19 diagnosis, in this paper, we systematically constructed a large scale COVID-19 Classification and Segmentation (*COVID-CS*) dataset. We also developed a Joint Classification and Segmentation (*JCS*) system for COVID-19 diagnosis. In our system, the classification model identified whether the suspected patient is COVID-19 positive or not, along with convincing visual explanations. It obtained a 95.0% sensitivity and 93.0% specificity on the classification test set of our *COVID-CS* dataset. To provide complementary pixel-level prediction, we implemented a segmentation model to discover fine-grained lesion areas in the CT images of COVID-19 patients. Comparing to the competing methods, e.g., PoolNet [58], our segmentation model achieved an improvement of 0.078 on Dice metric. Our *JCS* system is also very stable. On our segmentation test set, it failed only on 1.7% images and obtained Dice scores between $[0.6, 1]$ for 95.0% of images. The online demo on COVID-19 diagnosis will be available soon.

REFERENCES

[1] WHO, "Coronavirus disease (covid-19) outbreak situation," *https://www.who.int/emergencies/diseases/novel-coronavirus-2019*, 2020.

[2] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *medRxiv*, 2020.

[3] K. G. Malone, "Testing backlog linked to shortage of chemicals needed for covid-19 test," *The Canadian Press*, March 25, 2020.

[4] J. Won, S. Lee, M. Park, T. Kim, M. Park, B. Choi, D. Kim, H. Chang, V. Kim, and C. Lee, "Development of a laboratory-safe and low-cost detection protocol for sars-cov-2 of the coronavirus disease 2019 (covid-19)." *Experimental neurobiology*, 2020.

[5] J. Zhao, Y. Zhang, X. He, and P. Xie, "Covid-ct-dataset: a ct scan dataset about covid-19," *arXiv preprint arXiv:2003.13865*, 2020.

[6] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," 2020.

[7] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, 2020.

[8] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest ct for covid-19: comparison to rt-pcr," *Radiology*, 2020.

[9] Z. Huang, S. Zhao, Z. Li, W. Chen, L. Zhao, L. Deng, and B. Song, "The battle against coronavirus disease 2019 (covid-19): Emergency management and infection control in a radiology department," *Journal of the American College of Radiology*, 2020.

[10] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection," *arXiv 2003.11597*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[11] N. Sajid, "Covid-19 patients lungs x ray images 10000," https://www.kaggle.com/nabeelsajid917/covid-19-x-ray-10000-images, accessed 04 10, 2020.

[12] H. B. Jenssen, "Covid-19 ct segmentation dataset," http://medicalsegmentation.com/covid19/, accessed 04 10, 2020.

[13] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, and J. Xia, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, p. 200905, 2020.

[14] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z. A. Fayad, A. Jacobi, K. Li, S. Li, and H. Shan, "Ct imaging features of 2019 novel coronavirus (2019-ncov)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020.

[15] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z. A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, K. Li, S. Li, H. Shan, A. Jacobi, and M. Chung, "Chest ct findings in coronavirus disease-19 (covid-19): Relationship to duration of infection," *Radiology*, p. 200463, 2020.

[16] Y. Wang, C. Dong, Y. Hu, C. Li, Q. Ren, X. Zhang, H. Shi, and M. Zhou, "Temporal changes of ct findings in 90 patients with covid-19 pneumonia: A longitudinal study," *Radiology*, p. 200843, 2020.

[17] H. Shi, X. Han, N. Jiang, Y. Cao, O. Alwalid, J. Gu, Y. Fan, and C. Zheng, "Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study," *The Lancet Infect Disease*, vol. 20, no. 4, pp. 425–434, 2020.

[18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Int. Conf. Comput. Vis.*, Oct 2017.

[19] Y. Liu, Y.-H. Wu, Y. Ban, H. Wang, and M.-M. Cheng, "Rethinking computer-aided tuberculosis diagnosis," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016.

[21] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," *ArXiv*, 2020.

[22] R. Yang, X. Li, H. Liu, Y. Zhen, X. Zhang, Q. Xiong, Y. Luo, C. Gao, and W. Zeng, "Chest ct severity score: An imaging tool for assessing severe covid-19," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200047, 2020.

[23] K. Li, Y. Fang, W. Li, C. Pan, P. Qin, Y. Zhong, X. Liu, M. Huang, Y. Liao, and S. Li, "Ct image visual quantitative evaluation and clinical classification of coronavirus disease (covid-19)," *European Radiology*, 2020.

[24] Z. Zhou, D. Guo, C. Li, Z. Fang, L. Chen, R. Yang, X. Li, and W. Zeng, "Coronavirus disease 2019: initial chest ct findings," *European Radiology*, 2020.

[25] V. Rajinikanth, N. Dey, A. N. J. Raj, A. E. Hassanien, K. C. Santosh, and N. S. M. Raja, "Harmony-search and otsu based system for coronavirus disease (covid-19) detection using lung ct scan images," *arXiv*, 2020.

[26] J. B. Roerdink and A. Meijster, "The watershed transform: Definitions, algorithms and parallelization strategies," *Fundamenta Informaticae*, no. 1,2, pp. 187–228, 2000.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F. Li, "Imagenet: A large-scale hierarchical image database," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 248–255.

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

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. Learn. Represent.*, 2015.

[30] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017.

[31] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2020.

[32] M.-M. Cheng, Y. Liu, Q. Hou, J. Bian, P. Torr, S.-M. Hu, and Z. Tu, "Hfs: Hierarchical feature selection for efficient image segmentation," in *Eur. Conf. Comput. Vis.* Springer, 2016, pp. 867–882.

[33] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply supervised salient object detection with short connections," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, 2019.

[34] K. Zhao, J. Xu, and M.-M. Cheng, "Regularface: Deep face recognition via exclusive regularization," in *IEEE Conf. Comput. Vis. Pattern Recog.*, June 2019, pp. 1136–1144.

[35] Y.-Q. Tan, S.-H. Gao, X.-Y. Li, M.-M. Cheng, and B. Ren, "Vecroad: Point-based iterative graph exploration for road graphs extraction," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020.

[36] J.-W. Bian, Y.-H. Wu, J. Zhao, Y. Liu, L. Zhang, M.-M. Cheng, and I. Reid, "An evaluation of feature matchers for fundamental matrix estimation," in *Brit. Mach. Vis. Conf.*, 2019.

[37] J. Xu, Y. Huang, M.-M. Cheng, L. Liu, F. Zhu, X. Hou, and L. Shao, "Noisy-as-clean: learning unsupervised denoising from the corrupted image," *arXiv preprint arXiv:1906.06878*, 2019.

[38] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," in *Int. Conf. Comput. Vis.*, 2019.

[39] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.

[40] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Eur. Conf. Comput. Vis.*, 2018.

[41] C. Liu, L.-C. Chen, F. Schroff, H. Adam, W. Hua, A. Yuille, and L. Fei-Fei, "Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019.

[42] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 04 2017.

[43] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image. Comput. Comput. Assist. Interv.*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., 2015, pp. 234–241.

[45] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," in *Med. Image. Comput. Comput. Assist. Interv.*, Oct 2016, pp. 424–432.

[46] V. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," *ArXiv e-prints*, 2018.

[47] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, 2019.

[48] F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui, and D. Shen, "Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification," *arXiv preprint arXiv:2003.09860*, 2020.

[49] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopezpaz, "mixup: Beyond empirical risk minimization," in *Int. Conf. Learn. Represent.*, 2018.

[50] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image. Comput. Comput. Assist. Interv.* Springer, 2015, pp. 234–241.

[51] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, pp. 7132–7141.

[52] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 2117–2125.

[53] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Artificial intelligence and statistics*, 2015, pp. 562–570.

[54] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Int. Conf. 3D Vision*. IEEE, 2016, pp. 565–571.

[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learn. Represent.*, 2015.

[56] F. Shan+, Y. Gao+, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Lung infection quantification of covid-19 in ct images with deep learning," *arXiv preprint arXiv:2003.04655*, 2020.

[57] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-alignment Measure for Binary Foreground Map Evaluation," in *Int. Joint Conf. Artif. Intell.*, 2018.

[58] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, "A simple pooling-based design for real-time salient object detection," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019, pp. 3917–3926.

[59] J.-X. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "Egnet: Edge guidance network for salient object detection," in *Int. Conf. Comput. Vis.*, 2019, pp. 8779–8788.