Segmentation of COVID-19 pneumonia lesions: A deep learning approach

Zahra Ghomi¹, Reza Mirshahi*², Arash Khameh Bagheri³, Ali Fattahpour³, Saeed Mohammadiun⁴, Abdorreza Alavi Ghahrebagh³, Abtin Djavadifar⁴, Hossein Arabalibeik⁵, Rehan Sadig⁶, Kasun Hewage⁴

Received: 13 Sep 2020  Published: 22 Dec 2020

Abstract

Background: Lung CT scan has a pivotal role in diagnosis and monitoring of COVID-19 patients, and with growing number of affected individuals, the need for artificial intelligence (AI)-based systems for interpretation of CT images is emerging. In current investigation we introduce a new deep learning-based automatic segmentation model for localization of COVID-19 pulmonary lesions.

Methods: A total of 2469 CT scan slices, containing 1402 manually segmented abnormal and 1067 normal slices from 55 COVID-19 patients and 41 healthy individuals, were used to train a deep convolutional neural network (CNN) model based on Detectron2, an open-source modular object detection library. A dataset, including 1224 CT slices of 18 COVID-19 patients and 9 healthy individuals, was used to test the model.

Results: The accuracy, sensitivity, and specificity of the trained model in marking a single image slice with COVID-19 lesion were 0.954, 0.928, and 0.961, respectively. Considering a threshold of 0.4% for percentage of lung involvement, the model was capable of diagnosing the patients with COVID-19 pneumonia, with a sensitivity of 0.982% and a specificity of 88.5%. Furthermore, the mean Intersection over Union (IoU) index for the test dataset was 0.865.

Conclusion: The deep learning-based automatic segmentation method provides an acceptable accuracy in delineation and localization of COVID-19 lesions, assisting the clinicians and researchers for quantification of abnormal findings in chest CT scans. Moreover, instance segmentation is capable of monitoring longitudinal changes of the lesions, which could be beneficial to patients’ follow-up.

Keywords: COVID-19, Artificial intelligence, Deep learning, Tomography, X-ray, Pneumonia

Introduction

Since the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as the cause of “unknown viral pneumonia” in December 2019, the virus has spread to more than 200 countries (1). Due to the lack of specific treatment or vaccination, it is of great importance to diagnose the patients in the early stages of the disease for effective isolation. Reverse transcription polymerase chain reaction (RT-PCR) for respiratory specimens is proposed by Chinese guidelines for definitive diagnosis of the disease; however, it is restricted by high false negative at the initial presentation of the disease (2). Recently, it is shown that a lung CT scan has a pivotal role in diagnosis and management of COVID-19 patients so the current SARS-CoV-2 pandemic has imposed great work burden over radiologists.

Artificial intelligence systems help clinicians in detection and localization of pulmonary lesions.

What this article adds:

• Our study introduces a new deep learning-based automatic segmentation method for delineation and localization of COVID-19 pneumonia lesions.

What is “already known” in this topic:

• Chest computed tomography has an important role in diagnosis and management of COVID-19 patients so the current SARS-CoV-2 pandemic has imposed great work burden over radiologists.

*This work has been published under CC BY-NC-SA 1.0 license.

Copyright© Iran University of Medical Sciences

Automatic detection of COVID-19 lesions

high sensitivity in the detection of COVID-19 pneumonia. In fact, a lung CT scan has high value (more than 88% sensitivity) in the detection of suspected patients for COVID-19 with negative RT-PCT at the initial presentation (3).

Moreover, a CT scan is advantageous in the management of COVID-19 pneumonia patients, as it can demonstrate the course of the disease. Changes in pulmonary lesions can lead the physician in a patient’s treatment (4). It is also shown to be a useful criterion for termination of isolation. Patients with a persistent abnormality on CT scan or positive RT-PCR results are advised to restrict their physical social contact (5).

Considering the pivotal role of lung CT scan in diagnosis and management of COVID-19 pneumonia and the worldwide growing number of cases with the disease, considerable demand is created for interpretation of CT scan images by radiologists (6). In fact, radiologists play an important role in the management of COVID-19 patients. Other than early detection of COVID-19 pneumonia lesions, radiologists can monitor disease progression and treatment response by delineating each lesion and evaluating its changes on follow-up scans (7), which corresponds closely with prognosis. Therefore, the current pandemic has imposed great diagnostic burden over radiologists as well as human resources (8).

Application of an artificial intelligence (AI)-based system to detect COVID-19 pneumonia lesions can relieve the increasing work pressure of the radiologists and maintain diagnostic accuracy (9). While delineating discrete lesions for further follow-up in CT scans can be a time-consuming and tedious process for radiologists, it can be readily accomplished by the use of AI-based models. In current study, we aimed to introduce a new deep-learning system for automatic segmentation of COVID-19 lesions.

Methods

The current study was approved by local review board and adherents to tenets of the declaration of Helsinki.

Data acquisition

In this prospective study, lung CT scans of 73 cases of clinically definitive COVID-19 were retrieved from the local hospital. Patients with negative RT-PCR results were excluded from the study. All CT images were performed using a 16-detector scanner (Siemens Medical Systems) without the IV injection of contrast media. The CT protocol was as follows: 120 kV; automatic tube current; rotation time, 0.35 second; section thickness, 1.5 mm; collimation, 0.75 mm; pitch, 1–1.2; matrix, 512×512; and inspiration breath hold. Both lung (window width, 1000–1500 HU; window level, −700 HU) and mediastinal (window width, 350 HU; window level, 35–40 HU) were obtained for each patient; however, only the anonymized lung window DICOM files of the patients were imported for further analysis.

Three expert radiologists (ZG, AF, AK) reviewed the CT scan images. All areas of active pulmonary lesions were meticulously marked and segmented (Fig. 1) using ITK-SNAP (v 3.8.0, Yushkevich et al; http://www.itksnap.org) by a single physician (ZG), and 2 other radiologists (AF and AK) reviewed the segmentations and corrected the margins, if necessary. In 12 patients, all lesions were masked. In other 61 sets of images, sections of lung were sorted by area and each 10-slice was assessed for presence and marking of the lesions. Both main lesions of COVID-19 (ground glass and consolidation) were present in our cases and used for segmentation; however, a single mask was used for both of them.

For discrimination of the pathologic features, CT scan slides of 50 healthy individuals were obtained for comparison. Normal CT scans were acquired from those individuals with nonspecific symptoms who were being screened for COVID-19, but their scans were completely normal (based on consensus by the radiologists) and their laboratory tests were negative.

Model training

To reduce noise in CT scan images, a previously trained model has been implemented to segment right and left lungs (10). Afterwards, regions of interests (ROIs) marked
by the radiologists were applied to segmented images.

A total of 2469 CT scan slices (1402 abnormal and 1067 normal slices) were included for training. The images were randomly selected from 55 COIVD-19 patients, including all slices of 10 fully labelled set of images, and 41 healthy individuals. A total number of 6480 annotations were available in the training dataset.

For image segmentation, Detectron2, an open-source modular object detection library developed by the Facebook AI Research (FAIR) team was used. Detectron2 provides powerful tools for detection and semantic/instance segmentation tasks using various models, including panoptic segmentation, densepose, Cascade/Mask R-CNN (11), and rotated bounding boxes (12).

In this research, a Panoptic Feature Pyramid Network (PFPN) was used to perform semantic and instance segmentation. PFPN architecture includes 3 distinct blocks, feature extraction and region proposal block as a backbone, and 2 blocks in network head for semantic and instance segmentation. The first block works based on the Feature Pyramid Network (FPN) previously introduced by FAIR (13) added to a ResNet-50 backbone. FPN is a standard network extracting features at predefined spatial resolutions, which is used to construct a feature pyramid, parallel to selected feature maps in forward layers of related CNN. In the next block, the PFPN uses Cascade/Mask R-CNN on top of FPN, for instance, segmentation. Before applying Mask R-CNN to each pyramid level separately, the proposed regions of interest undergo an operation called Region of Interest Align (RoIAlign), which eliminates the quantization error of RoI pooling (11). Finally, to obtain the semantic segmentation in the third block, a lightweight dense prediction branch, consisting of a range of bilinear up-sampling and convolution operators, have been used on top of the same FPN features to merge different layers into a pixel-wise output. These operators combine features, results, and information of all formerly developed FPN spatial resolutions and generate a single semantically segmented image. The conceptual framework of the image segmentation used in this research is illustrated in Figure 2.

The utilized PFPN model is pretrained on the COCO dataset (14) trained in the present study to segment lesions on lung CT image slices. The details of model implementation are illustrated in Table 1. To avoid overfitting, data augmentation strategies were used, including random rotation, random flip, adjusting brightness, and contrast. The model was run on a cloud computing service (Google Colab®) in a total time of approximately 20 minutes. ROIs with a detection confidence threshold greater than 0.3 are used for segmentation of the lesions. This threshold determines the confidence of the model (0-1) required for making a prediction. The threshold of 0.3 was set empirically based on maximum resulting IoU in test dataset from different values of threshold.

Model metrics

A dataset, including 1224 CT slices of 18 COVID-19’s patients and 9 healthy individuals was used for the test. The total number of abnormal slices was 250, along with 974 normal images. To assess the efficacy of the model, 2 sets of evaluation were selected.

First, the exported instance masks of the validation dataset were obtained from the model. Afterwards, each image was classified based on the presence or the absence of predicted COVID-19 lesion. The classification of predicted images was compared with ground truth classification to calculate sensitivity, specificity, accuracy, and area under the curve (AUC) for binary classification results as follows:

![Fig. 2. A basic diagram of the image segmentation process](image-url)

**Table 1. Model parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Rate</td>
<td>0.005</td>
</tr>
<tr>
<td>Number of iterations</td>
<td>500</td>
</tr>
<tr>
<td>Batch size per image</td>
<td>512</td>
</tr>
<tr>
<td>Error functions</td>
<td>Loss_mask, loss_rpn_cls, loss_box_reg, loss_rpn_loc</td>
</tr>
<tr>
<td>Final Loss_mask</td>
<td>0.09</td>
</tr>
</tbody>
</table>

http://mjiri.iums.ac.ir
Med J Islam Repub Iran. 2020 (22 Dec); 34.174.
Automatic detection of COVID-19 lesions

Accuracy = \frac{TP+TN}{TP+FP+TN+FN}

Sensitivity = \frac{TP}{TP+FN}

Specificity = \frac{TN}{TN+FP}

With TP, TN, FP, and FN being true positive, true negative, false positive, and false negative, respectively.

Jaccard Index was used to evaluate instant segmentation. This index is defined as the intersection of predicted and ground truth masks divided by the union of these 2:

\[ J(A,B) = \frac{\text{intersection}(A,B)}{\text{union}(A,B)} \]

The Jaccard index or the Intersection over Union (IoU) metric provides a good estimation of overlapping pixels between the predicted and ground truth masks. The IoUs of each pair of images (prediction and ground truth) were calculated; an IoU score of 1 was assigned to instances of healthy individuals, where the model did not detect any lesion successfully, and there was a score of 0 for false positive detection. The total sum of IoUs divided by the number of images was defined as the mean IoU of the model.

Results

We employed an instance segmentation model for localization and quantification of COVID-19 pneumonia lesions. To achieve the best accuracy, different numbers for the confidence of detection were used and a confidence score of 0.3 was selected based on the AUC results.

In image level, an accuracy of 0.924, sensitivity of 0.954, and specificity of 0.884 was obtained on training dataset with an AUC of 0.926. On the test dataset, the AUC of 0.920 was accompanied with accuracy, sensitivity, and specificity of 0.954, 0.928, and 0.961, respectively (Table 2).

To assess the efficacy of the model in detecting COVID-19 patients in the case level (the whole set of CT images for a patient), percentage of lung involvement detected by the model was determined for each patient and healthy participants. Based on ROC curve analysis with a threshold of 0.4% for percentage of lung involvement, the model achieved a sensitivity of 98.7%, a specificity of 88.5%, and AUC of 0.982 in differentiating abnormal lung CT scans from healthy participants.

To assess instance segmentation, IoU was evaluated for each pair of images (prediction and ground truth) separately, and mean IoU across the dataset was calculated. The mean IoU for training and test dataset was 0.704 and 0.865, respectively. An example of COVID-19 lesions localization on one of the test dataset images is illustrated in Figure 3.

Table 2. Results based on training and testing datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training Dataset</th>
<th>Testing Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity*</td>
<td>0.954</td>
<td>0.928</td>
</tr>
<tr>
<td>Specificity*</td>
<td>0.884</td>
<td>0.961</td>
</tr>
<tr>
<td>AUC*</td>
<td>0.926</td>
<td>0.920</td>
</tr>
<tr>
<td>Intersection over Union (IoU)</td>
<td>0.704</td>
<td>0.865</td>
</tr>
</tbody>
</table>

*At image level

Fig. 3. Prediction of COVID-19 lesion: Main CT scan image (A). Ground truth manual segmentation of COVID-19 lesions (B). Binary mask generated from manual segmentations (C). The input of deep learning model (D). Prediction made by the model showing three distinct lesions (E). Binary mask generated from model prediction.
Discussion

Chest CT scan is proved to be a useful tool in both diagnosis and management of patients with COVID-19 pneumonia (15). Previously, it has been shown that the severity and appearance of lung opacification changes during different stages of the disease and it can be used to predict the severity of the clinical disease (16, 17). Multiple studies emphasize the role of CT scan imaging as a prognostic tool. A CT involvement score has been proposed by Xie et al for the classification of lung involvement (3) and assessment of disease extent and severity.

Although radiologists can efficiently diagnose and differentiate COVID-19 from other causes of pneumonia (18), the increasing requests of lung CT scan interpretation during COVID-19 pandemic can cause significant pressure on the radiologists and exhaust human resources. On the other hand, manual segmentation of discrete lesions for assessing the follow-up scans and predicting progression is a time-consuming and difficult task (19), taking up to 5 hours for each patient based on our experience. In this regard, artificial intelligence and deep learning approaches offer a unique tool for quantification of abnormalities.

The previously published studies on artificial intelligence and COVID-19 CT scans were primarily focused on differential diagnosis and screening of the patients (20-22). However, the main goal in the current study was developing an accurate method for segmentation of SARS-CoV-2 lesions in lung CT scans. Few studies have investigated the role of AI in segmentation of COVID-19 lesions. Shen et al (23) used a semi-automatic computer assisted method for segmentation of COVID-19 lesions. After segmenting the lung, pulmonary vasculature and air walls were subtracted from the images using a “3D active surface evolution approach” for detection of COVID-19 lesion. In their study, a Pearson correlation of $r = 0.8$ was observed between percentage of lesion detected by expert radiologist and computer tool. In another study by Fan et al (24) automatic segmentation was implemented using semi-supervised Inf-Net model. Based on a parallel partial decoder, the extracted high level features were aggregated to a global map for coarse segmentation of lesions. Afterwards, reverse attention and edge attention modules were added to the model for enhancing the boundaries of localized lesions. In quantitative assessment, a Dice similarity coefficient of 0.739 was obtained with sensitivity and specificity of 0.725 and 0.96, respectively.

In Shan et al study (19), to reduce the time consuming work of manual delineation of the COVID-19 lesions, a human-in-the-loop (HITL) strategy was adopted. In brief, the first batch of manually annotated images were fed to VB-net based network and the DL model updated iteratively with a training dataset from manually corrected segmented results of previous models. A Dice score of 91.6% ± 10.0% in lesion segmentation was obtained by the final model.

The current deep learning model demonstrated a Jaccard index of 0.865 for automatic segmentation of COVID-19 lesions and it has an accuracy of 96.29% for labeling a single slice CT scan image as atypical pneumonia. Furthermore, by analyzing the output of the model for the complete set of chest CT scan images of patients and normal participants, a threshold for number of abnormal slices was defined for the diagnosis of COVID-19 patients with 100% sensitivity.

The admitted patients undergo multiple scans usually with a mean interval of 3.5 days (17, 25). The pulmonary lesions change during the time-course of COVID-19 disease (4). The latest Novel Coronavirus Pneumonia Diagnosis and Treatment Plan (Provisional 7th Edition), issued

![Fig. 4. First row showing CT scan slices of a patient during 10 days follow up. The second row generated a binary mask of the model output of corresponding CT images showing progression and enlargement of the lesion during the follow up. In addition the mean value of masked pixels shows changes from 31.23 (gray scale) toward 39.43 highlighting the transformation of the lesion from ground glass opacity to consolidation. (Note that the instance segmentation of the new emerging lesion in anterior of the left lung is omitted for better presentation).](http://mjiri.iums.ac.ir)
by National Health Commission of the public republic of China (5), has stated that more than 50% progression of lesions on lung imaging render the clinical stage as a severe form of the disease. The instance segmentation of the present research model is capable of monitoring every single lesion in serial scans and with post-processing techniques, the image features indices (eg, mean value of pixels) can be calculated in the ROI. The changes in these indices could potentially demonstrate the transformation of lesions during the course of the disease in addition to the alterations in size and percentage of lung involvement (Fig. 4). Moreover, with the implementation of differential lobe segmentation of an available pretrained model (10), the percentage of COVID-19 involvement can be calculated in each lobe separately. As mentioned in the literature, COVID-19 lesions distribution is more pronounced in lower segments of lungs and the present model could potentially evaluate the preferential location of COVID-19 pneumonia.

There are some limitations to this study. First, the evaluated datasets were collected from the patients of a single center, which might affect the output of implementing the model on CT scan images of the patients from different geographical locations. Second, various slices of the same subject are used and since they are not statistically independent, the estimated sensitivity, specificity and other metrics should be interpreted with caution. Third, the developed model cannot differentiate the commonly described pathologies in COVID-19 CT imaging, including ground glass opacity, crazy paving patterns, reversed halo opacity, and consolidation (26); however, our training dataset included various types of COVID-19 lesions. Another limitation of the study is the lack of incorporation of CT scan of the patients with pulmonary findings similar to COVID-19 (eg, influenza pneumonia) to differentiate SARS-CoV-2 from other viral pathogens. In future studies, we plan to train our model on a more diverse dataset, including different types of lesions and patients with different diagnoses of lung diseases. In addition, the monitoring capability of the model can be used for studies regarding the follow-up of the patients and risk stratification at the time of admission.

**Conclusion**

The current deep learning-based approach in localization and quantification of COVID-19 pneumonia can be a useful tool in classification, follow up, and management of patients with COVID-19 pneumonia. It also paves the way for future studies evaluating the role of baseline and longitudinal characteristics of CT scan imaging in predicting the outcome of the patients and risk stratification.

**Conflict of Interests**

The authors declare that they have no competing interests.

**References**