

Automated lung segmentation from CT images of normal and COVID-19 pneumonia patients

Faeze Gholamiankhah¹⁺, Samaneh Mostafapour²⁺, Nouraddin Abdi Goushbolagh¹, Seyedjafar Shojaerazavi³, Parvaneh Layegh⁴, Seyyed Mohammad Tabatabaei^{5,6}, Hossein Arabi⁷

¹Department of Medical Physics, Faculty of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

²Department of Radiology Technology, Faculty of Paramedical Sciences, Mashhad University of Medical Sciences, Mashhad, Iran

³Department of Cardiology, Ghaem Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

⁴Department of Radiology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

⁵Department of Medical Informatics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

⁶Clinical Research Development Unit, Imam Reza Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

⁷Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva 4, Switzerland

Abstract

Objective:

Automated semantic image segmentation is an essential step in quantitative image analysis and disease diagnosis. This study investigates the performance of a deep learning-based model for lung segmentation from CT images for normal and COVID-19 patients.

Methods:

Chest CT images and corresponding lung masks of 1200 confirmed COVID-19 cases were used for training a residual neural network. The reference lung masks were generated through semi-automated/manual segmentation of the CT images. The performance of the model was evaluated on two distinct external test datasets including 120 normal and COVID-19 subjects, and the results of these groups were compared to each other. Different evaluation metrics such as dice coefficient (DSC), mean absolute error (MAE), relative mean HU difference, and relative volume difference were calculated to assess the accuracy of the predicted lung masks.

Results:

The proposed deep learning method achieved DSC of 0.980 and 0.971 for normal and COVID-19 subjects, respectively, demonstrating significant overlap between predicted and reference lung masks. Moreover, MAEs of 0.037 HU and 0.061 HU, relative mean HU difference of -2.679% and -4.403%, and relative volume difference of 2.405% and 5.928% were obtained for normal and COVID-19 subjects, respectively. The comparable performance in lung segmentation of the normal and COVID-19 patients indicates the accuracy of the model for the identification of the lung tissue in the presence of the COVID-19 induced infections (though slightly better performance was observed for normal patients).

Conclusion:

The promising results achieved by the proposed deep learning-based model demonstrated its reliability in COVID-19 lung segmentation. This prerequisite step would lead to a more efficient and robust pneumonia lesion analysis.

Keywords: COVID-19, CT, segmentation, deep learning

Introduction

The novel coronavirus named SARS-CoV-2 was first broke out in Wuhan China in December 2019 and has invaded most countries around the globe [1, 2]. Through infecting the respiratory tracts, this virus causes respiratory syndromes [3], where the real-time polymerase chain reaction (RT-PCR) is known as the common/standard method for the diagnosis of COVID-19. However, some concerns such as a high false-negative rate when the viral load is low in the test specimen have limited its application [1, 4].

On the contrary, chest X-ray or computed tomography (CT) imaging are considered as faster and complementary procedures that facilitate the early screening of COVID-19 infections [5, 6]. However, CT imaging outperforms X-ray radiography in providing more structural/anatomical details of the lung [7-13]. Chest CT images are reported to have a sensitivity of 0.97 in the diagnosis of COVID-19 [14], enabling to detect the radiological patterns like bilateral and peripheral ground-glass opacities and patchy consolidations in the lung of infected patients [2, 15]. Moreover, quantitative analysis of CT images provides key information about the size of lesions and severity of the disease [16] wherein reliable lung CT image segmentation is a critical prerequisite step in this regard [17, 18].

Different approaches for lung segmentation have been adopted including manual segmentation, rule-based, atlas-based, machine learning-based as well as hybrid techniques [19-21]. Manual segmentation is a time-consuming and labor-intensive task particularly in situations that the health system is overloaded [6]. Other conventional methods such as atlas-based or intensity-based algorithms lead to acceptable results in normal cases or mild disease, however, their implementation in diseases like COVID-19 that infection alters the common pattern/structure of the lung is inefficient [19, 22, 23]. To address this challenge, recent researches have evaluated the use of deep learning models for lung and lesion segmentation and demonstrated the promising performance of convolutional neural networks in distinguishing lung from the chest wall [24, 25].

There are a number of studies that have employed the common deep learning-based image segmentation architectures such as U-Net, 3D U-Net, U-Net++, and V-Net for COVID-19 lung segmentation [6, 26, 27]. Furthermore, some studies developed and evaluated state-of-the-art algorithms for the cases with insufficient annotated datasets using transfer learning or weakly annotated datasets [28-30]. Due to the presence of considerable abnormalities in the lung caused by the COVID-19 infection, segmentation of the lung in COVID-19 patients faces the challenge of lung boundary and infection discrimination compared to the normal patients which bear distinct contrast between the lung tissue and chest wall [31].

This study sought to assess the efficiency of the deep learning approach in automated lung segmentation from CT images of patients with COVID-19 in comparison with normal patients. Automated lung segmentation would assist quantitative analysis and segmentation of infections by removing unnecessary regions in the chest images. To conduct a meticulous investigation, a dataset of normal patients was also employed in addition to CT scans of infected patients to study the impact of lung abnormalities caused by COVID-19 infection on the lung segmentation.

Material and Methods

Dataset

The dataset used in this study consists of chest CT images from 1200 patients with RT-PCR confirmed COVID-19 and 120 normal patients without any lung abnormalities. CT image acquisition was performed on a Siemens Somatom Spirit Dual Slice CT scanner with tube energy of 130 kVp, tube current of 48 mAs, rotation time (TI) of 0.8 s, and slice thickness of 5 mm. For generating ground truth lung masks, CT images were segmented semi-automatically using Pulmonary Toolkit (PTK) software [32] and the resulting binary masks were manually corrected to avoid any noticeable errors. Prior to the training of the network, all images were cropped to eliminate areas outside of the lung volume and resized to a matrix size of 296×216 voxels by a linear interpolation algorithm. Thereafter, Hounsfield units were scaled to an intensity range between 0 and 1.

Implementation detail

The ResNet model implemented in NiftyNet was utilized for the implementation of the automated lung segmentation. NiftyNet is an open-source platform built upon TensorFlow that consists of common convolutional neural networks used in medical imaging [33]. The ResNet architecture, as shown in figure 1, comprises 20 convolutional layers wherein every two layers are connected together by residual connections. In this network, dilation factors of one, two, and four are applied on the convolutional kernels to extract low-level, mid-level, and high-level features from the input images, respectively. Also, a fully connected softmax layer is embedded as the last layer of the network [34, 35].

From the total number of subjects included in this study, 1080 CT images and their corresponding masks were randomly selected for the training of the network, and the remaining 120 subjects for validation (external validation). To make sure that there is no risk of overfitting, 5% of the training subjects were employed for validation of the model within the training phase. The investigations revealed no considerable difference between training and evaluation losses.

The training of the deep learning model was performed on two-dimensional slices using the following settings: learning rate = 0.02, optimizer = Adam, loss function = Dice_NS, decay = 0.0001, batch size = 17, and weights regression type = L2norm.

Evaluation Metrics

To evaluate the performance of the deep learning model, predicted and ground truth lung segmentations were compared on the external test dataset including 120 COVID-19 patients and 120 normal subjects. The assessment was performed via calculating the dice similarity coefficient (DSC) (Eq. 1), Jaccard index (JC) (Eq. 2), mean error (ME) (Eq. 3), mean absolute error (MAE) (Eq. 4) within the estimated lung region.

$$DSC = \frac{2 |I_r \cap I_p|}{|I_r| + |I_p|} \quad (1)$$

$$JC = \frac{|I_r \cap I_p|}{|I_r \cup I_p|} \quad (2)$$

$$ME = \frac{1}{N} \sum_{i=1}^N (I_p(i) - I_r(i)) \quad (3)$$

$$MAE = \frac{1}{V} \sum_{i=1}^V |I_p(i) - I_r(i)| \quad (4)$$

Here, I_r and I_p denote the reference and predicted lung masks. V and i indicate the total number of voxels in the lung area and the index of voxels in I_r and I_p images, respectively.

Moreover, the false-positive ratio (Eq. 5) and false-negative ratio (Eq. 6) were estimated as follow:

$$False\ Positive\ Ratio = \frac{FP}{FP + TN} \quad (5)$$

$$False\ Negative\ Ratio = \frac{FN}{FN + TP} \quad (6)$$

In Eqs. 5 and 6, FP is the number of false positives, TN is the number of true negatives, FN is the number of false negatives and TP is the number of true positives for the voxels residing in the lung area.

Furthermore, relative mean CT number (Hounsfield Unit (HU)) difference, absolute relative mean HU difference, relative volume difference, and absolute relative volume difference metrics were calculated between the reference and predicted lung volumes.

Results

Representative results of the lung segmentation for normal and COVID-19 subjects are presented in figures 2 and 3. Figure 2 shows a good match between the reference and predicted masks for both normal and infected lung tissues, which indicates the promising performance of the deep learning model in lung border detection. Figure 3 depicts minor segmentation errors for two cases in which the model was not successful to define an accurate margin for the lung and excluding bronchus from the segmented area (outlier report). The miss-segmentation error is more noticeable in COVID-19 patients due to the similar intensity of the chest wall and severe infections, which have rendered the accurate identification of the lung boundary very challenging.

Table 1 summarizes the minimum, maximum, mean, and standard deviation of the quantitative metrics for lung segmentation including DSC, JC, ME, MAE, False Positive Ratio, False Negative ratio, mean HU difference within the lung mask, and volume difference calculated for external validation datasets. Overall, the proposed model showed better performance in lung segmentation of normal subjects compared to COVID-19 patients due to the high-density infections residing close to the chest wall in COVID-19 subjects.

The mean DSC and JC values in normal group were 0.980 ± 0.003 and 0.962 ± 0.007 , and in COVID-19 group were 0.971 ± 0.017 and 0.938 ± 0.040 , respectively. The deep learning model achieved ME of -0.015 ± 0.009 HU and -0.024 ± 0.042 HU, and MAE of 0.037 ± 0.007 HU and 0.061 ± 0.040 HU within the lung mask for the normal and COVID-19 subjects, respectively. Moreover, comparable measures of false positive and false negative ratios were obtained for the two test datasets. The quantitative assessment revealed relative Mean HU Differences of $(-2.679 \pm 0.382\%$ and $-4.403 \pm 4.097\%)$ and relative volume differences of $(2.405 \pm 7.359\%$ and $5.928 \pm 17.261\%)$ calculated for the lung tissue in the normal and COVID-19 subjects, respectively.

Figure 4 illustrates boxplots of DSC, JC, ME, MAE, Relative Mean HU Difference, and Relative Volume Difference metrics calculated for the two test datasets. Overall, considering the entire metrics, larger standard deviation and outliers were observed in the COVID-19 patients compared to the normal subjects.

Discussion

Recent studies suggest that chest CT imaging findings play a significant role in COVID-19 diagnosis and management [36, 37]. Accurate lung segmentation is a crucial step for the calculation of the quantitative indices, measurement of lung engagement, and disease severity [38, 39]. The existing segmentation methods, that have shown satisfactory performance in normal or mild lung diseases, are either time-

consuming or face serious challenges in the segmentation of COVID-19 infected lung tissue due to the close similarity between infections and normal tissues [7]. Deep learning-based models have been widely adopted lately by researchers as a dependable solution to assist clinicians in fast and efficient COVID-19 lung and lesion segmentation [40, 41]. This study set out to investigate the performance of a state-of-the-art deep learning approach in lung segmentation. The network was evaluated on two external test datasets including normal and COVID-19 subjects. Multiple evaluation metrics were used to perform a comprehensive performance assessment of the model.

Gerard et al. [22] developed a segmentation algorithm called LobeNet aiming at the prediction of the left and right lung regions in the presence of diffuse opacities and consolidations. The quantitative assessment of LobeNet performance on 87 patients with COVID-19 showed an average Dice coefficient of 0.985. Moreover, Ma et al. [38] reported a DSC score of 0.973 and 0.977 obtained from a 3D U-Net for the left and right lung segmentation from COVID-19 CT images. Tilborghs et al. [40] compared the performance of various deep learning algorithms using a multicenter COVID-19 dataset. They concluded that combining different methods could improve the segmentation performance and increased the Dice coefficient up to 0.987. The proposed ResNet model implemented in this study achieved average Dice coefficients of 0.980 ± 0.003 and 0.971 ± 0.017 in normal and COVID-19 external test data set, respectively. This indicates a very good segmentation overlap between predicted and ground truth lung masks.

Yan et al. [6] established a new network called COVID-SegNet for the segmentation of CT images with COVID-19 infection. They compared their proposed method against other state-of-the-art models including FCN, UNet, VNet, and UNet++. The COVID-SegNet achieved DSC, sensitivity, and precision of 0.865, 0.986, and 0.983, respectively. In another study conducted by Trivizakis et al. [42], the U-Net architecture, one of the most common image segmentation architectures, was implemented for COVID-19 lung segmentation. They reported a DSC of 0.950, sensitivity of 0.920, and specificity of 0.875. Muller et al. [43] performed a similar analysis using 3D U-Net for lung segmentation using a cross-validation scheme to reduce/avoid the risk of overfitting. Their approach led to lung segmentation with DSC of 0.956, sensitivity of 0.956, and specificity of 0.998. Compared to the results reported in the literature, the false-positive ratio of 0.005 and 0.007, and false-negative ratio of 0.026 and 0.044 obtained for the normal and COVID-19 patients in this study demonstrated the promising performance of the proposed deep learning approach.

The investigation of the deep learning model showed high accuracy in terms of Relative Mean HU Difference (-2.679% and -4.403%), and Relative Volume Difference (2.405% and 5.928%) in normal and COVID-19 cases, respectively. Therefore, the performance of the ResNet model in COVID-19 patients,

wherein the infection has changed the intensity of the lung, is comparable with its performance in normal subjects.

The investigation of the framework proposed in this work revealed promising results in lung segmentation of patients with COVID-19. However, there are some limitations associated with this study that require to be addressed in future studies. One of these limitations would be the lack of a multicenter dataset to investigate the sensitivity of the developed model to the variation of the image quality and acquisition parameters across different scanners/centers. Moreover, in order to develop a comprehensive framework which is suitable for clinical practice, the performance of the ResNet model in lesion segmentation should be evaluated. This would enable distinguishing COVID-19 lesions from other pneumonia, and diagnosing subtypes of COVID-19 pneumonia.

Conclusion

In this work, deep learning-based automated lung segmentation from chest CT images in patients with COVID-19 was investigated. The proposed method achieved very promising results with a dice coefficient of 0.980 and 0.971 in normal and COVID-19 external test datasets, respectively. The reliable lung segmentation would facilitate lesion segmentation and will lead to more accurate quantitative analysis, diagnosis, and treatment of the COVID-19 patients.

References

- [1] Jin C, Chen W, Cao Y, Xu Z, Zhang X, Deng L, et al. Development and Evaluation of an AI System for COVID-19 Diagnosis. medRxiv. 2020.
- [2] Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Lung infection quantification of covid-19 in ct images with deep learning. arXiv preprint arXiv:200304655. 2020.
- [3] Shiri I, Akhavanallaf A, Sanaat A, Salimi Y, Askari D, Mansouri Z, et al. Ultra-low-dose chest CT imaging of COVID-19 patients using a deep residual neural network. European radiology. 2020;1-12.
- [4] Zhou L, Li Z, Zhou J, Li H, Chen Y, Huang Y, et al. A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis. IEEE transactions on medical imaging. 2020;39:2638-52.
- [5] Dai W-c, Zhang H-w, Yu J, Xu H-j, Chen H, Luo S-p, et al. CT imaging and differential diagnosis of COVID-19. Canadian Association of Radiologists Journal. 2020;71:195-200.
- [6] Yan Q, Wang B, Gong D, Luo C, Zhao W, Shen J, et al. COVID-19 Chest CT Image Segmentation--A Deep Convolutional Neural Network Solution. arXiv preprint arXiv:200410987. 2020.
- [7] Fan D-P, Zhou T, Ji G-P, Zhou Y, Chen G, Fu H, et al. Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. IEEE Transactions on Medical Imaging. 2020.
- [8] Minaee S, Kafieh R, Sonka M, Yazdani S, Soufi GJ. Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. arXiv preprint arXiv:200409363. 2020.
- [9] Ghane B, Karimian A, Mostafapour S, Gholamiankhak F, Shojaerazavi S, Arabi H. Quantitative analysis of image quality in low-dose CT imaging for Covid-19 patients. arXiv preprint arXiv:210208128. 2021.
- [10] Arabi H, Asl ARK. Feasibility study of a new approach for reducing of partial volume averaging artifact in CT scanner. 2010 17th Iranian Conference of Biomedical Engineering (ICBME)2010. p. 1-4.
- [11] Arabi H, Kamaliasl AR, Aghamiri SM. THE EFFECT OF FOCAL SPOT SIZE ON THE SPATIAL RESOLUTION OF VARIABLE RESOLUTION X-RAY CT SCANNER. INTERNATIONAL JOURNAL OF RADIATION RESEARCH. 2010;8:37-43.
- [12] Lohrabian V, Kamali-Asl A, Arabi H, Mamashi F, Hemmati HR, Zaidi H. Design and construction of a variable resolution cone-beam small animal mini-CT prototype for in vivo studies. Radiation Physics and Chemistry. 2019;162:199-207.
- [13] Lohrabian V, Kamali-Asl A, Harvani HG, Aghdam SRH, Arabi H. Comparison of the x-ray tube spectrum measurement using BGO, NaI, LYSO, and HPGe detectors in a preclinical mini-CT scanner: Monte Carlo simulation and practical experiment. arXiv preprint arXiv:210205760. 2021.
- [14] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020;296:E32-E40.
- [15] Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. Deep learning-based detection for COVID-19 from chest CT using weak label. MedRxiv. 2020.
- [16] Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. Cell. 2020;181:1423-33. e11.
- [17] Lessmann N, Sánchez CI, Beenen L, Boulogne LH, Brink M, Calli E, et al. Automated assessment of CO-RADS and chest CT severity scores in patients with suspected COVID-19 using artificial intelligence. Radiology. 2020.
- [18] Wu Q, Wang S, Li L, Wu Q, Qian W, Hu Y, et al. Radiomics Analysis of Computed Tomography helps predict poor prognostic outcome in COVID-19. Theranostics. 2020;10:7231.
- [19] Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. European Radiology Experimental. 2020;4:1-13.

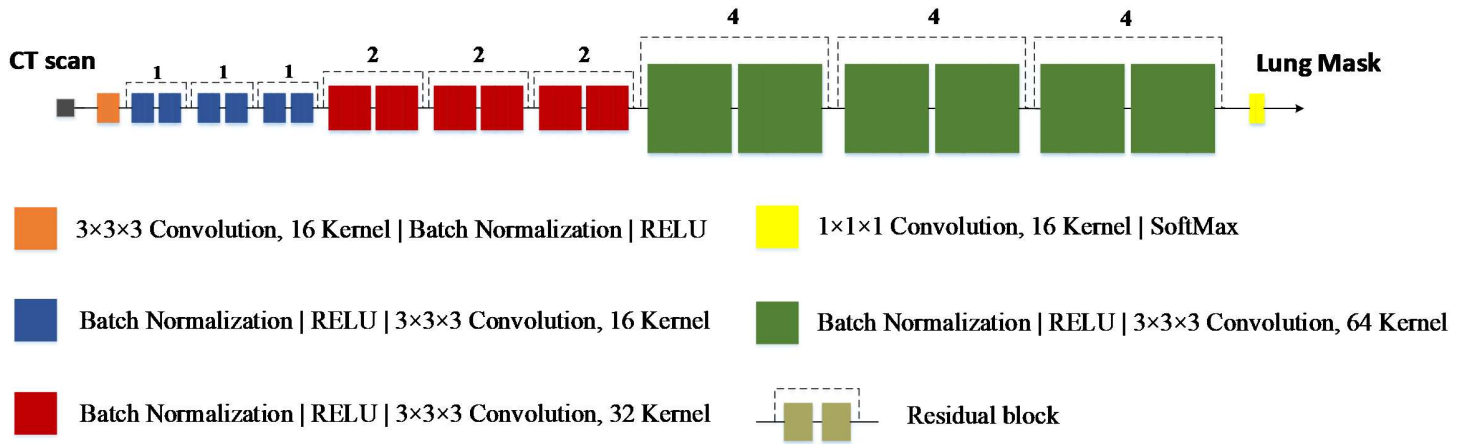
- [20] Arabi H, Zaidi H. Comparison of atlas-based techniques for whole-body bone segmentation. *Medical image analysis*. 2017;36:98-112.
- [21] Arabi H, Zaidi H. Whole-body bone segmentation from MRI for PET/MRI attenuation correction using shape-based averaging. *Medical physics*. 2016;43:5848.
- [22] Gerard SE, Herrmann J, Xin Y, Martin KT, Rezoagli E, Ippolito D, et al. CT image segmentation for inflamed and fibrotic lungs using a multi-resolution convolutional neural network. *Scientific reports*. 2021;11:1-12.
- [23] Arabi H, Zaidi H. Deep learning-based metal artefact reduction in PET/CT imaging. *European radiology*. 2021.
- [24] Skourt BA, El Hassani A, Majda A. Lung CT image segmentation using deep neural networks. *Procedia Computer Science*. 2018;127:109-13.
- [25] Arabi H, Zaidi H. Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy. *Euro J Hybrid Imaging*. 2020;4:17.
- [26] Zhou T, Canu S, Ruan S. An automatic covid-19 ct segmentation network using spatial and channel attention mechanism. *arXiv preprint arXiv:200406673*. 2020.
- [27] Bahrami A, Karimian A, Fatemizadeh E, Arabi H, Zaidi H. A new deep convolutional neural network design with efficient learning capability: Application to CT image synthesis from MRI. *Medical physics*. 2020;47:5158-71.
- [28] Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE transactions on medical imaging*. 2020;39:2531-40.
- [29] Chen S, Ma K, Zheng Y. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:190400625*. 2019.
- [30] Karimi D, Dou H, Warfield SK, Gholipour A. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical image analysis*. 2020;65:101759.
- [31] Arabi H, Zaidi H. Magnetic resonance imaging-guided attenuation correction in whole-body PET/MRI using a sorted atlas approach. *Medical image analysis*. 2016;31:1-15.
- [32] Alnaser A, Gong B, Moeller K. Evaluation of open-source software for the lung segmentation. *Current Directions in Biomedical Engineering*. 2016;2:515-8.
- [33] Arabi H, Zaidi H. One registration multi-atlas-based pseudo-CT generation for attenuation correction in PET/MRI. *European journal of nuclear medicine and molecular imaging*. 2016;43:2021-35.
- [34] Ladefoged CN, Benoit D, Law I, Holm S, Kjær A, Højgaard L, et al. Region specific optimization of continuous linear attenuation coefficients based on UTE (RESOLUTE): application to PET/MR brain imaging. *Physics in Medicine & Biology*. 2015;60:8047.
- [35] Arabi H, Zaidi H. Deep learning-guided estimation of attenuation correction factors from time-of-flight PET emission data. *Medical image analysis*. 2020;64:101718.
- [36] Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raouf S, et al. The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society. *Chest*. 2020;158:106-16.
- [37] Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*. 2020;296:E115-E7.
- [38] Ma J, Wang Y, An X, Ge C, Yu Z, Chen J, et al. Towards Data-Efficient Learning: A Benchmark for COVID-19 CT Lung and Infection Segmentation. *Medical physics*. 2020.
- [39] Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE reviews in biomedical engineering*. 2020.

- [40] Tilborghs S, Dirks I, Fidon L, Willems S, Eelbode T, Bertels J, et al. Comparative study of deep learning methods for the automatic segmentation of lung, lesion and lesion type in CT scans of COVID-19 patients. arXiv preprint arXiv:200715546. 2020.
- [41] Loey M, Smarandache F, M Khalifa NE. Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. *Symmetry*. 2020;12:651.
- [42] Trivizakis E, Tsiknakis N, Vassalou EE, Papadakis GZ, Spandidos DA, Sarigiannis D, et al. Advancing COVID-19 differentiation with a robust preprocessing and integration of multi-institutional open-repository computer tomography datasets for deep learning analysis. *Experimental and therapeutic medicine*. 2020;20:1-.
- [43] Müller D, Rey IS, Kramer F. Automated Chest CT Image Segmentation of COVID-19 Lung Infection based on 3D U-Net. arXiv preprint arXiv:200704774. 2020.

Table 1. Statistics of quantitative metrics including ME, MAE, RMSE, RE, RVD, Dice, JC, Sensitivity, SSIM, and PSNR calculated between the reference and predicted lung masks in normal and COVID-19 test datasets.

Parameter	Normal			COVID-19		
	Min	Max	Mean \pm SD	Min	Max	Mean \pm SD
Dice Coefficient	0.970	0.985	0.980 \pm 0.003	0.903	0.986	0.971 \pm 0.017
Jaccard Index	0.942	0.971	0.962 \pm 0.007	0.823	0.973	0.938 \pm 0.040
ME	-0.031	0.015	-0.015 \pm 0.009	-0.155	0.042	-0.024 \pm 0.042
MAE	0.028	0.057	0.037 \pm 0.007	0.026	0.176	0.061 \pm 0.040
False Positive Ratio	0.001	0.022	0.005 \pm 0.004	0.003	0.028	0.007 \pm 0.004
False Negative Ratio	0.020	0.036	0.026 \pm 0.003	0.011	0.167	0.044 \pm 0.040
Relative Mean HU Diff (%)	-3.673	-2.020	-2.679 \pm 0.382	-16.799	-1.183	-4.403 \pm 4.097
Absolute Relative Mean HU Diff (%)	0	0.374	0.828 \pm 1.318	0.921	10.894	3.253 \pm 3.106
Relative Volume Diff (%)	-4.726	29.524	2.405 \pm 7.359	-12.666	90.561	5.928 \pm 17.261
Absolute Relative Volume Diff (%)	1.960	33.164	7.875 \pm 6.548	2.199	91.486	12.743 \pm 16.384

Figure 1. The architecture of the ResNet model.



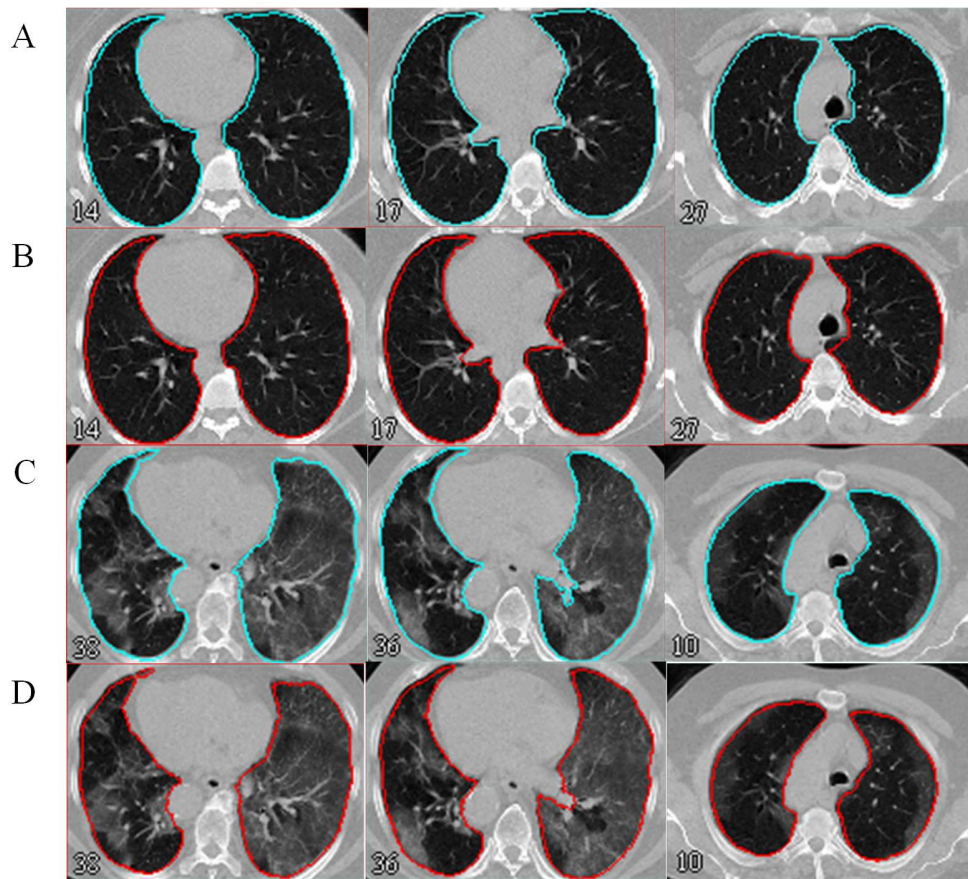


Figure 2. Axial views of A) Ground truth lung masks in normal subjects. B) and the corresponding predicted lung masks. C) Ground truth lung masks in COVID-19 subjects D) and the corresponding predicted lung masks.

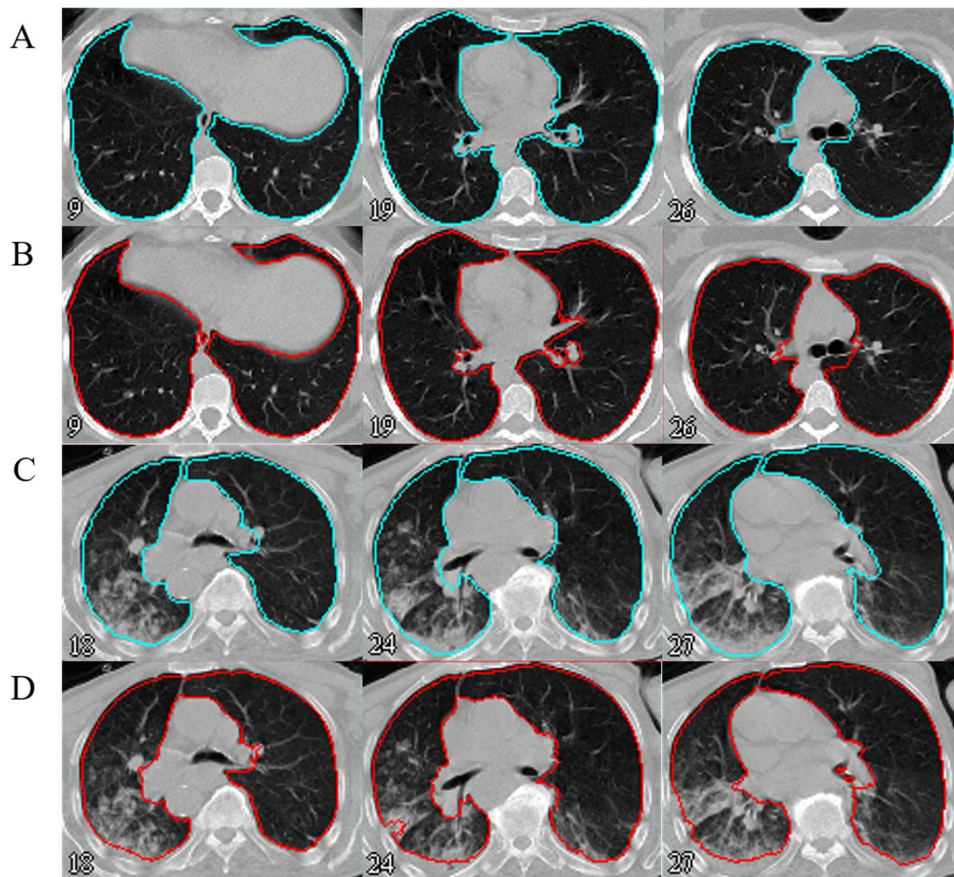


Figure 3. Outlier report. Cases with noticeable errors. Axial views of A) Ground truth lung masks in normal subjects. B) and the corresponding predicted lung masks. C) Ground truth lung masks in COVID-19 subjects D) and the corresponding predicted lung masks.

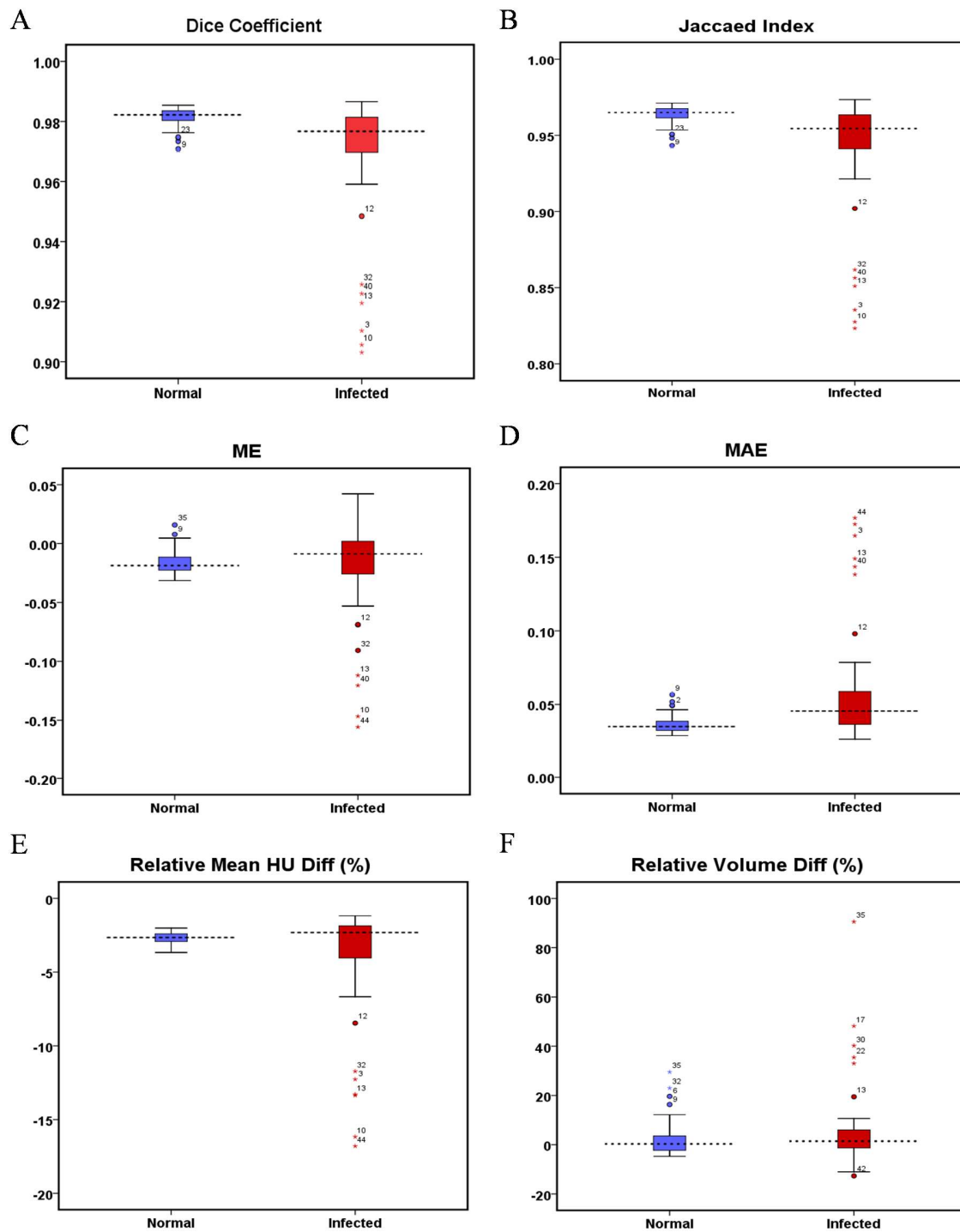


Figure 4. Box plots comparing A) Dice Coefficient, B) Jaccard Index, C) ME, D) MAE, E) Relative Mean HU Difference (%) and F) Relative Volume Difference(%) metrics between normal and COVID-19 test datasets.