

Deep Learning-Based Approach for Diagnosis of COVID-19 Patients from CT Lung Images

*Thi-Thao Tran, Tien-Thanh Tran, Quoc-Cuong Ninh, Van-Truong Pham **

Hanoi University of Science and Technology, Hanoi, Vietnam

**Email: truong.phamvan@hust.edu.vn*

Abstract

COVID-19 is currently one of the most life-threatening problems around the world. The fast and accurate detection of the COVID-19 infection plays an important role in identifying, making better decisions as well as ensuring treatment for the patients. In this study, we propose an approach to identify COVID-19 patients via deep learning methods. The proposed approach includes two steps: segmentation of lung and classification of patients via analyzing the segmented images. Especially, in the first step, the attention gate is integrated into Unet neural network structure, which can help increase the accuracy of the segmentation tasks. In the second step, the EfficientNet-B4 is applied to classify COVID-19 patients. By using EfficientNet, the performance of the classification process as well as the size of the model are improved. We apply the proposed approach for the database including 130 patients. Experimental results show the desired performances of the proposed approach.

Keywords: COVID-19, deep learning, segmentation, computed tomography, artificial intelligence.

1. Introduction

Novel coronavirus disease 2019 (COVID-19) is caused by the new Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). COVID-19 is an ongoing pandemic and one of the most serious pandemics in human history. So far, there have been 192 million confirmed cases and more than 4 million confirmed deaths all around the world [1]. Early detected COVID-19 infected patients play an important role in the treatment and slow down the spread of this disease. In addition, in epidemic areas, insufficient medical resources also become a big challenge [2]. Hence, finding high-risk patients with a worse prognosis for special care as well as prior medical resources is crucial in the treatment of COVID-19. Today, the gold standard to confirm COVID-19 patients relies on the real-time polymerase chain reaction (RT-PCR), which comprises detecting viral RNA from nasopharyngeal swab or sputum [3]. Although it is the gold standard, the RT-PCR has some drawbacks such as it is time-consuming (normally, it takes from several hours to several days to get the results). In addition, the cost of the equipment to perform RT-PCR method is high and the equipment, normally, not available in hospitals.

Another approach that might speed up the screening process is basing on the results of analysing lung CT images from the suspected patients [4]. From the lung CT images analysis, the doctors and experts can confirm the patients are infected or not infected

with the COVID-19. However, this approach requires the experiences of the doctors. In fact, these doctors are normally work in the main hospitals in the cities. Moreover, when the spread of the disease is in a large range, the doctors need to pay more time for confirmed patients and the time for analysing CT lungs of the suspected patients is really limited. Therefore, automatic diagnosis of lesions from lung CT images to find the COVID-19 patients is a promising approach [5]. The computer-aid diagnosis (CAD) for diagnosis COVID-19 can help to save time, reduce overload for the medical system. Even in some cases, the accuracy of the results detected by the CAD system is similar to ones obtained by experienced experts. However, the analysing of CT Lung images is a challenging task since the CT lung images are normally with inhomogeneity, the size and shape of infected regions are various, and the intensity of infected regions are similar to normal ones.

On the other hand, nowadays artificial intelligence (AI) and deep learning (DL) have been widely applied in many fields [6]. In medical imaging, the DL approach can help to obtain promising results via analysing CT lung images [7,8]. With the ability in learning the features, DL can automatically discover and find the features related to diseases from CT lung images. Especially, a fully convolutional neural network (FCN) has shown the ability in features extraction and learning the features from input images. FCN has been used to recognise the tumors in the lung via analysing the CT images [9,10]. In analysing

images of the COVID-19 patients, there are many approaches that apply FCN to analysis of the CT lung such as the works in [8]. In the current study, we propose a supervised approach that applies DL to the diagnosis of COVID-19 patients via analysing the lung CT images. The proposed approach includes two steps which in the first step the lung regions in CT are segmented by performing an encoder-decoder deep-learning model. In the second step, from the segmented regions, a classification-based DL algorithm has been performed to classify the patients with or without COVID-19 infection. We then apply the proposed approach for the database including 130 patients to evaluate the performances of the proposed approach.

2. The Proposed Approach

In this study, we propose an approach based on deep learning for the diagnosis of COVID 19 from CT images. The proposed method includes two main stages, image segmentation, and Classification /Prediction. The main blocks of the method are presented in Fig. 1. The data after pre-processing is fed into the EfficientNet [11] and classifier for class prediction and diagnosis.

In the first stage, a database, called Lung dataset including CT images and the corresponding CT masks are applied to the Attention Unet for image segmentation task. It is noted that the Lung dataset is only used once to get the Segment weight. After training by the Attention Unet, the optimal weight is saved and called the Segment weight. Next, the target dataset for COVID 19 diagnosis, different from the Lung dataset, is used for the next stages.

The target dataset is shuffled into the training and

testing sets. The input images from the target dataset for COVID 19 diagnosis including the training and testing sets are segmented using the saved Segment weight. The purpose of this segmentation step is to remove undesired regions, which are outside the lung from the images. The outputs of the Attention Unet are predicted binary masks. From these masks, we have CT image segmentations and can crop the region nearby the lung, which are more useful for classification and diagnosis. This stage gives cropped images from both training and testing sets for the next stage. The cropped images of the training set are fed into the EfficientNet to learn and get the weight that contains optimal parameters for image classification. Finally, the cropped images in the testing sets are applied to the EfficientNet using this weight and get the classification and prediction results.

2.1. CT Lung Image Segmentation

The Lung dataset for image segmentation includes CT images provided by [12]. Each image pair includes an image and the corresponding mask. The masks are segmented by radiologists to determine a pixel group belonging to the lung or not [13]. Some examples of this CT lung image database are provided in Fig. 2. As we can see from this figure, each mask is separated into two regions, the area in white refers to the lung, and the black region belongs to the background. The role of image segmentation is to assign each pixel to the white or the black and then used it for further processing. For this image segmentation stage, we propose using the Unet [14] incorporated with the attention gate [15] network. For training the neural network, we use the Focal Tversky loss [16] for efficient CT lung image segmentation.

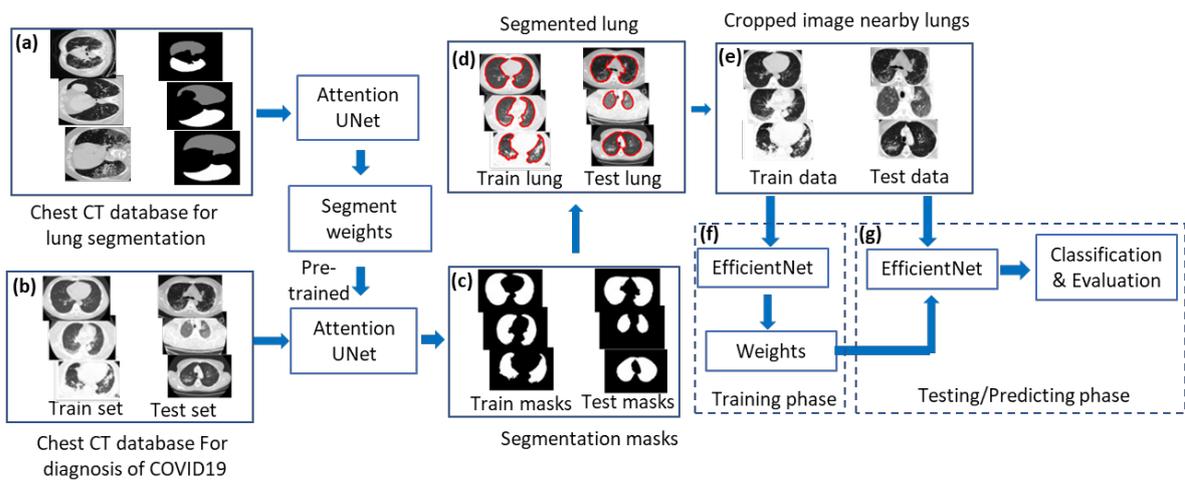


Fig. 1. Pipeline of the proposed approach for COVID-19 patient diagnosis. (a) The dataset used for training lung segmentation. (b) Train and Test sets for COVID19 diagnosis. (c) Predicted masks of Attention Unet. (d) Lung segmentation results. (e) Cropped image region nearby lungs. (f) Training phase in which images from training set are fed into EfficientNet to get classified weights. (g) Testing/Predicting phase.

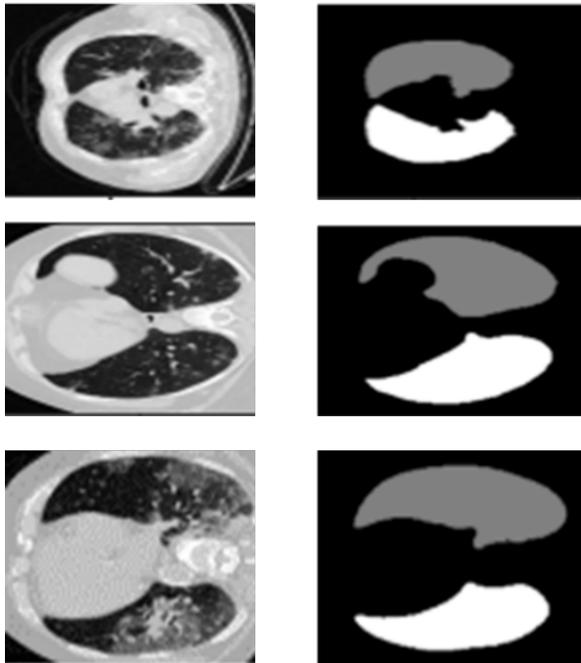


Fig. 2. Input CT images and corresponding masks of left and right lungs

Unet is an encoder-decoder symmetric architecture including multiple convolutional layers [14], activation, and pooling layers as shown in Fig. 3. The main purpose of the encoder is to extract the

image features while the decoder enables precise localization of objects using transposed convolutions. In particular, in the encoder, after some convolution operations, a max-pooling layer is followed to reduce the size of feature maps. There are total 5 max-pooling layers in the Unet baseline architecture, as shown in the left-right of Fig. 3. For the decoder, 5 corresponding unpooling layers are used to preserve the spatial dimension of the feature maps.

In Unet architecture, to maintain image spatial information that might be lost while doing the pooling operation, the authors [11] used the skip connection mechanism to fuse feature maps in the encoder to the decoder layer of the same size. Besides, to deal with varying shapes and sizes of objects, we use the Unet variant with the use of attention gate [15] as shown in Fig. 4. By this way, the model can be more effective since it can focus on important target structure from the feature map. The raw information extracted from the encoder is then fed to the attention gate to eliminate noise and the less important information before fusing with the corresponding features in the decoder.

The feature map after the last layer of decoder is applied to the sigmoid function to classify all pixels of the feature map into two classes, the lung or the background. The output of the classification is a binary mask whose size, height, and width are the same as the size of the input image.

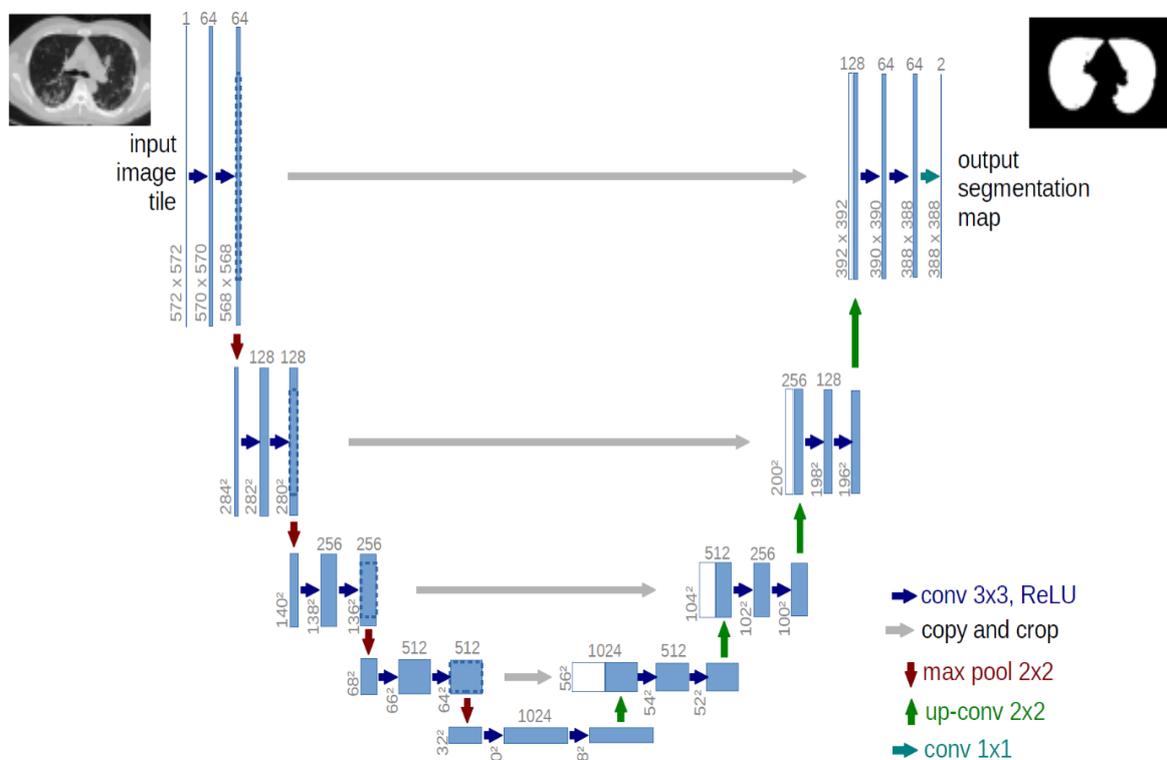


Fig. 3. The Unet structure and illustration of application on CT image segmentation

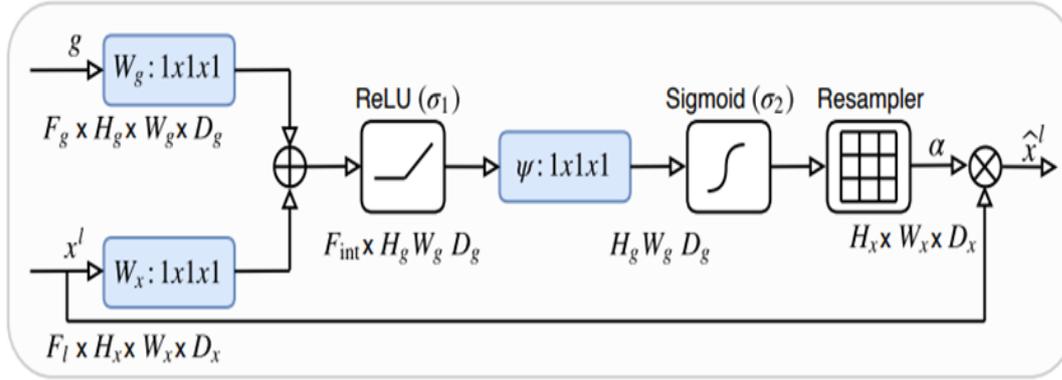


Fig. 4. Structure of the attention gate

For training the network, in this study, we use the Focal Tversky loss [16]. The use of the Tversky loss is originated from the notion of Tversky index, defined as

$$TI_c = \frac{\sum_{i=1}^N p_{ic} g_{ic} + \varepsilon}{\sum_{i=1}^N p_{ic} g_{ic} + \alpha \sum_{i=1}^N p_{i\bar{c}} g_{i\bar{c}} + \beta \sum_{i=1}^N p_{ic} g_{i\bar{c}} + \varepsilon} \quad (1)$$

where p_{ic} is the probability that a pixel I has been predicted to belong to class c (lung), $p_{i\bar{c}}$ is the probability that a pixel I not belong to c . Similarly, g_{ic} and $g_{i\bar{c}}$ are probabilities that the pixel belongs to ground truth or not. The hyperparameters α , β are chosen as 0.7 and 0.3, respectively. From the Tversky index, the Focal Tversky loss is defined as

$$FTL_c = \sum_c (1 - TI_c)^{\gamma} \quad (2)$$

with $\gamma = 1.3$. Using the focal Tversky helps the model focus more on the loss while using the exponential. It is also useful while considering the False Positive and False Negative, rather than focusing only on the True Positive and True Negative as in other losses.

Segmentation Evaluations

In this study, the training for image segmentation is performed on 600 training image pairs. To evaluate the performance of the segmentation network, we use another set of 100 test images whose corresponding masks (labels) are available but blinded during training. The segmentation training is trained with 50 epochs using Adam optimizer.

To evaluate the performance of the segmentation stage, we use the Dice similarity score (DSC). The DSC measures the overlap between the automatic segmentation (attention Unet architecture, in this study), and the ground truth (masks) segmented by experts. The Dice metric is defined as

$$DSC = \frac{\sum_{i=1}^N p_{ic} g_{ic} + \varepsilon}{\sum_{i=1}^N p_{ic} + g_{ic} + \varepsilon} \quad (3)$$

Using the DSC metric, we obtained the score of DSC up to 0.98 for the image segmentation on the test set of 100 images mentioned above.

Some representative results of image segmentation are given in Fig. 5. From this figure, we can observe that the segmented results by the Attention Unet are similar to the ground truths.

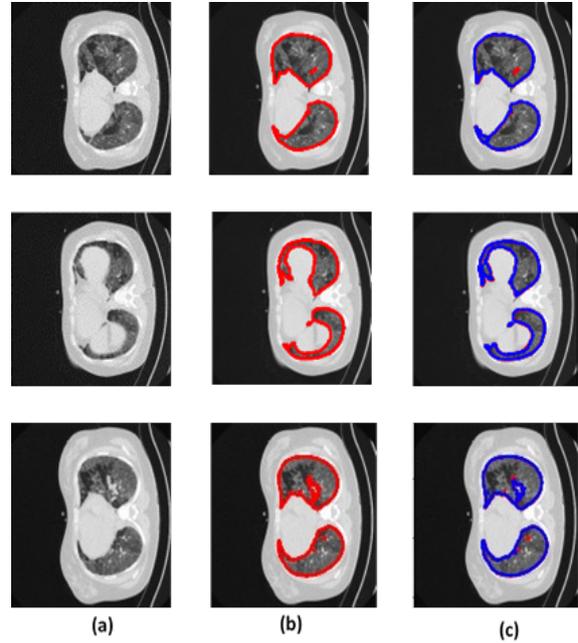


Fig. 5. Illustration of CT lung image segmentation. (a) input CT images; (b) Ground truths; (c) Predicted maps resulted from Attention Unet are drawn on the input images.

2.2 CT Image Segmentation Postprocessing

From prediction maps obtained from the image processing step, in the patient's CT image, we determine the regions belonging to the lung and those outside the lung. There have been several methods to process the pixels outside the lung. One of the most commonly used is to set all pixels outside the lung to white (corresponding to the value of 255 if we use 8-bit format) and set all pixels inside the lung to black (pixel value is zero). We also can compute the mean intensity value of pixels outside the lung and set these pixels to the mean value and keep the pixels outside the lung unchanged.

In this study, we directly crop the image nearby the lung relying on the binary mask obtained from the preceding image segmentation. The detailed steps are described as the following:

- i. From the mask of the corresponding input CT image, determine the boundary positions of pixels belonging to the lung ((x_{\min}, y_{\min}) , (x_{\max}, y_{\max})).
- ii. Crop the CT image bounded by the determined positions ($(x_{\min} + r, y_{\min} + r)$, $(x_{\max} - r, y_{\max} - r)$), with r is a hyperparameter, chosen as 5.
- iii. Resize the input image obtained from (ii) to the original resolution using bilinear interpolation.

Some examples showing the input images, corresponding masks, and the cropped versions are shown in Fig.6.

2.3. Classification and COVID-19 diagnosis

After segmentation and postprocessing, we use a deep convolutional neural network with a classifier to classify the CT images. The process is given as follows.

Dataset: The CT dataset acquired from 143 patients provided by Zhao et al. in [5] is used in this study. The raw images are acquired from different devices and conditions so the images are with different contrast, resolution, and quality. The smallest, average, and largest heights of these images are respectively 153, 491, and 1853. Meanwhile, the smallest, average, and largest widths of these images are respectively 124,383, and 1485. Some examples of this database are shown in Fig.7.

Preprocessing: Applying the first stage of the proposed approach, the CT images are segmented by Attention Unet architecture, then the unwanted background outside the lung is removed. After that, the cropped images are resized to the original resolutions.

Data augmentation: a simple but effective approach to help deep neural networks avoid overfitting during training. By using random affine transformation on the input image, we can obtain additional images with different poses, sizes, and positions of objects, without changing the labels. Nevertheless, choosing suitable data augmentation methods for medical images in general, and CT, in particular, is needed to take into consideration. These types of images are more difficult to augment compared to non-medical images since they need knowledge about medical image modality as well as the experience to avoid losing useful image information and features for diagnosis.

In this study, the authors used several bilinear interpolations on the input image to ensure the labels do not vary. In particular, the images are performed with some transformations such as random scale with the weight of 0.1, random rotation below 5 degrees, and horizontal flip.

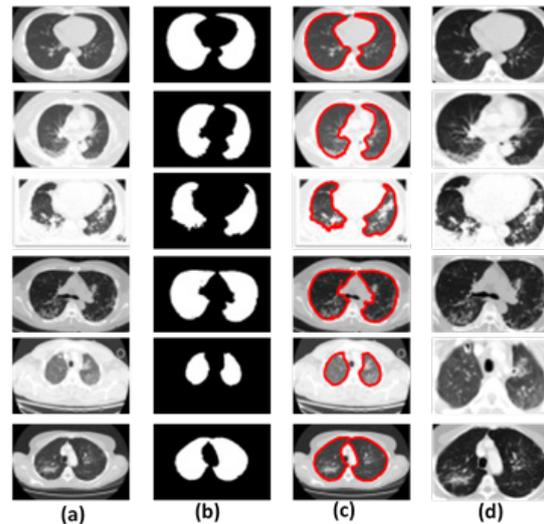


Fig. 6. Illustration of image segmentation and postprocessing. (a) Input CT images. (b) Prediction masks by Attention Unet. (c) Segmentation contour drawn on input images. (d). CT images after being cropped to eliminate the unwanted regions outside the lung.

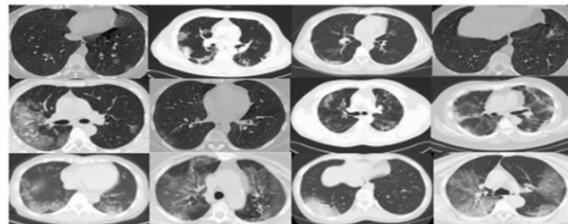


Fig. 7. Some examples of images in the dataset of COVID-19 patients

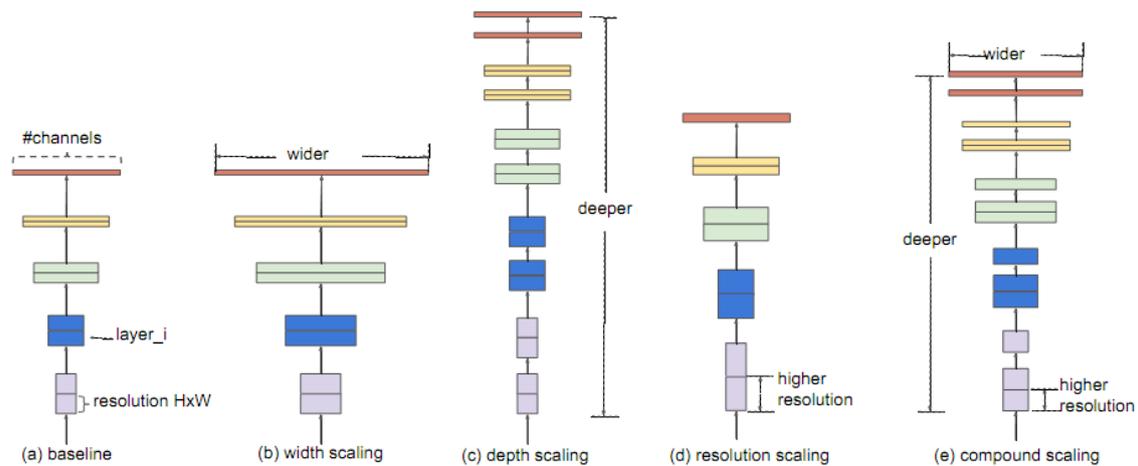


Fig. 8. Comparison between conventional scaling methods (b-d) that arbitrary scale a single network dimension (a), and the EfficientNet [11] that uniformly scales up all dimensions (e).

Deep neural networks: There have been many deep convolution neural networks developed and proposed for image classification in recent years. Some well-known and first models proposed during the revolution can be listed as AlexNet, VGG, GoogleNet. Inspired from pioneering works, succeeding models are proposed with new architecture and mechanism to improve the accuracy and performance of the classification problems such as Resnet and DenseNet. Resnet and Densenet are presented with the novelty of using “skip connection” could help the models of huge trainable parameters avoid common problems in neural networks such as overfitting, vanishing gradient.

The development of deep neural network tends to increase the “deep” of the neural network. The increase of the deep normally leads to the increases of network parameter numbers input image sizes to improve the performance of the classification. However, these approaches generally focus on only one of the following factors: increasing the depths of networks (deeper), increasing the input image size (higher resolution), or increasing the channel numbers of feature maps (wider). More recently, a new neural network model has been introduced, EfficientNet [11] and showed a breakthrough in the field, that has taken all three above factors into consideration. In the EfficientNet, the three metrics, network depth, input resolution, and channel number of features maps, are mutually related. In more detail, in the EfficientNet, the authors proposed the compound coefficient to scale up CNNs in a more structured manner. Unlike conventional network that arbitrarily scale width, depth and resolution, the EfficientNet uniformly scales each dimension with a fixed set of scaling coefficients. An illustration of comparison between conventional scaling methods that arbitrary scale a single network dimension, and the EfficientNet that uniformly scales up all dimensions is shown in Fig.8.

It has been shown that the tuning of these factors in the network is practical and could help improve the performance in terms of accuracy and improved resolution. In the work [11], the authors proposed 8 different models ranging from EfficientNet-B0 to EfficientNet-B7. The base EfficientNet-B0 network is based on the mobile inverted bottleneck convolution (MBCConv) of MobileNetV2, and the squeeze-and-excitation blocks. While compared with other existing CNNs on ImageNet, the EfficientNet models achieve both higher accuracy and better efficiency, and also reduce parameter size. For instance, the EfficientNet-B7 reaches state-of-the-art 84.4% top-1 / 97.1% top-5 accuracy on ImageNet; the EfficientNet-B4 improves the top-1 accuracy from 76.3% of ResNet-50 to 82.6% (+6.3%) while compared with the widely used ResNet.

In this study, after intensive investigation and evaluation on models of the EfficientNet family, we chose the EfficientNet-B4 for classification problem of COVID-19 in this study, based on the size of the data.

Training protocol: The model is trained using the GPU hardware provided by GoogleColab, with 100 epochs. The binary cross-entropy loss and Adam optimization algorithm with the learning rate of 0.0001 are used. During training, the weight of the best validation performance is stored and then used for the inference step with the test images.

3. Results on COVID 19 Classification

The proposed approach is validated on the CT image database provided by [5] for COVID-19 diagnosis. To evaluate the performance of image classification, we use some common metrics, including the Accuracy, Recall, F1. In addition, we also provide the ROC (receiver operating characteristic) curve along with AUC (area under the curve) value.

The experiments are conducted with the following assessments:

- The diagnosis results in the case of using the image segmentation step, and the case of not using the segmentation step, presented in Table 1. We also plot ROC curves in such two cases in Fig. 9.
- The comparison between the results by the proposed approach with those by Zhao et al. [5], given in Table 2.

Table 1. Results by the proposed approach using the dataset provided by Zhao *et al.* [5]

Metrics	Proposed approach without segmentation	Proposed approach with segmentation
Accuracy	83.5	87.2
Recall	83.7	91.9
F1	88.7	91.5
AUC	90.9	92.6

Table 2. Comparison of results by the proposed approach and the work of Zhao *et al.* [5]

Metrics	Zhao et al. [5]	Proposed approach
Accuracy	84.7	87.2
Recall	76.2	91.9
F1	85.3	91.5
AUC	82.4	92.6

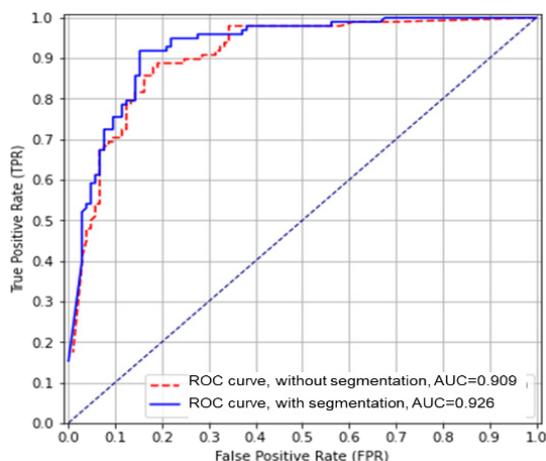


Fig. 9. ROC curves of the proposed approach in the case of without using the image segmentation step (red dot line), and the case of using image segmentation step (blue line)

From the above experimental results and evaluation, we can perceive that:

When using all steps by the proposed approach including the image segmentation stage, we obtain higher evaluation metrics, as in Table 1. In addition, the ROC in the case including the segmentation step lies above the ROC of the case without segmentation stage, as shown in Fig. 9. This evaluates the effectiveness of the approach in this study.

The proposed approach obtains higher evaluation measures, including accuracy, AUC, F1 and Recall than the model by Zhao et al. [5], shown in Table 2. Among the evaluation metrics, the improvement of Recall is very important in the case of COVID 19 diagnosis. This stems from the fact that a high Recall shows that the model has a good ability in diagnosing of positive cases of COVID-19. And via versa, if a model gives a low Recall, it might omit more positive cases of COVID-19, leading to bad circumstances/effects.

4. Conclusions

This paper presents a method including two main stages for the diagnosis of COVID 19 from CT images. The proposed approach achieves higher performance compared to the baseline method in terms of various evaluation metrics. More importantly, the Recall measure by the proposed approach is much higher than the comparative model, showing the ability and reliability in diagnosis of patients positive with COVID 19. The proposed approach can be improved to obtain higher classification accuracy by using more effective image segmentation, as well as data augmentation methods. In addition, the weighted loss can be taken into consideration to help the model focus on positive cases of COVID 19. For a real application, our research group plan to develop the proposed method in a server and incorporate with an Internet of Thing system to help COVID 19 diagnosis.

Acknowledgement

This research is supported by the Hanoi University of Science and Technology (HUST) under grant No: T2020 - PC - 017.

References

- [1]. The World Health Organization daily situation report on the COVID-19 pandemic. Available at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- [2]. Holshue, M., DeBolt, C., Lindquist, S., al., e.: First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine* 2020 382(10), 929-936 (2020). <https://doi.org/10.1056/NEJMoa2001191>
- [3]. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., Xia, L.: Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019

- (COVID-19) in China: A Report of 1014 Cases. *Radiology*, 200642 (2020).
<https://doi.org/10.1148/radiol.2020200642>
- [4]. Feng, S., Jun, W., Jun, S., Ziyang W., Qian, W., Zhenyu, T., Kelei, H., Yinghuan, S., Dinggang, S.: Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. in *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 4-15, 2021,
<https://doi.org/10.1109/RBME.2020.2987975>
- [5]. Zhao, J., Zhang, Y., He, X., Xie, P., COVID-CT-Dataset: A CT Scan Dataset about COVID-19. <https://arxiv.org/abs/2003.13865> (2020).
- [6]. Van-Truong Pham, Thi-Thao Tran, Pa-Chun Wang, Po-Yu Chen, Men-Tzung Lo: EAR-UNet: A deep learning-based approach for segmentation of tympanic membranes from otoscopic images. *Artificial Intelligence In Medicine* 115, 1-12 (2021).
<https://doi.org/10.1016/j.artmed.2021.102065>
- [7]. Wang, S., Shi, J., Ye, Z., Dong, D., Yu, D., Zhou, M., Liu, Y., Gevaert, O., Wang, K., Zhu, Y.: Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *European Respiratory Journal* 53(3), 1800986 (2020).
<https://doi.org/10.1183/13993003.00986-2018>
- [8]. Walsh, S., Calandriello, L., Silva, M., Sverzellati, N.: Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *The Lancet Respiratory Medicine* 6(11), 837-845 (2018).
[https://doi.org/10.1016/S2213-2600\(18\)30286-8](https://doi.org/10.1016/S2213-2600(18)30286-8)
- [9]. Choe, J., Lee, S., Do, K., Lee, G., Lee, J., Lee, S., Seo, J., Deep learning-based image conversion of ct reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *radiology* 292(2), 365-373 (2019).
<https://doi.org/10.1148/radiol.2019181960>
- [10]. Wang, P., Xiao, X., Glissen Brown, J., Berzin, T., Tu, M., Xiong, F., ... Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng.* 2(10), 741-748 (2018).
<https://doi.org/10.1038/s41551-018-0301-3>
- [11]. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: *Proc. 36th Int. Conf. Mach. Learn* 2019, pp. 6105–6114
- [12]. COVID-19 CT segmentation dataset. Available at: <http://medicalsegmentation.com/covid19/>.
- [13]. Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem.
<https://arxiv.org/abs/2001.11767> (2020).
<https://doi.org/10.1186/s41747-020-00173-2>
- [14]. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* 2015, pp. 234-241
https://doi.org/10.1007/978-3-319-24574-4_28
- [15]. Schlemper, J., Oktay, O., Schaap, M., M. Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis* 53, 197–207 (2019).
<https://doi.org/10.1016/j.media.2019.01.012>
- [16]. Sadegh, S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: *Proc. Int. Workshop Mach. Learn. Med. Imag.* 2017, pp. 379-387.
https://doi.org/10.1007/978-3-319-67389-9_44