Journal of Engineering Research

Volume 6 | Issue 5 Article 21

2022

DeepLab V3+ Based Semantic Segmentation of COVID -19 Lesions in Computed Tomography Images

Merihan Mohamed Abdalwahab Eissa Abdalwahab, Sameh Napoleon, Amira S. Ashour

Vol. 6, No. 5 – 2022 Journal of Engineering Research (ERJ)

DeepLab V3+ Based Semantic Segmentation of COVID -19 Lesions in Computed Tomography Images

Merihan M. Eissa*, Sameh A. Napoleon, Amira S. Ashour

Department of Electronics and Electrical communication Engineering, Faculty of Engineering, Tanta University, Egypt
email: merihan.eissa@f-eng.tanta.edu.eg

***Abstract*- Coronavirus 2019 spreads rapidly worldwide causing a global epidemic. Early detection and diagnosis of COVID-19 is critical for treatment as it causes respiratory syndrome appears in the chest medical images, such as computed tomography (CT) images, and X-ray images. The CT images are more sensitive and have more details compared to the X-ray images. Thus, automated segmentation plays an imperative role in detecting, diagnosing, and determining the spreading of COVID-19. In this paper, the DeepLabV3+ combined with MobileNet-V2 model was implemented. To validate this combination, we conducted a comparative study between the DeepLabV3+ variants by its combination with MobileNet-V2 against DeepLabV3+ combined with different CNN, namely ResNet-18, and ResNet50. Also, a comparative study with the basic traditional U-Net and modified Alex for segmentation was carried out. The experimental results showed the superiority of the using DeepLabV3+ combined with MobileNet-V2 for COVID-19 segmentation by achieving 97.5% mean accuracy, 95.2% sensitivity, 99.7% specificity, 99.7% precision, 99.3 % weighted Jaccard coefficient, and 97.5% weighted dice coefficient.**

***Keywords*- COVID-19, semantic segmentation, DeepLabV3+, MobileNet-V2, depth-wise separable convolution layer.**

1.INTRODUCTION

Coronavirus spreads rapidly causing death and negatively affecting the public health as it is considered a global pandemic. The diagnosis and screening of COVID-19 reduce its spread and improve the ability of treatment [1] Medical imaging techniques like X-ray and computed tomography (CT) scans are considered robust solutions to diagnosis and determine the severity of the COVID-19 infection[2,3]. The CT imaging provides a 3D scan which provides more details compared to X-ray imaging. Also, CT scans are sensitive and able to detect irregularities region in lung and have a high spatial resolution [1,4]. Accordingly, several studies have been conducted to develop automatic segmentation models for detecting COVID-19. One of these models is using semantic segmentation which is based on pixel-wise image classification, where each pixel in the image is classified into one of the predetermined classes. Its architecture is divided into two main parts, namely the encoder for features extraction, and the decoder part for generating the predicted segmented image from the extracted information [5].

The DeepLabV3+ network is based on redesigned convolutional neural network (CNN) model, such as pretrained ResNet-18, ResNet-50, and MobileNet-V2 as encoder part in the semantic segmentation [4]. On the other hand, MobileNetV2 is a CNN network that consists of inverted residuals, and linear bottlenecks blocks based on depth-wise separable convolution layer [6]. The depth-wise separable convolutional layer is used instead of the standard convolutional layer as it consists of depth-wise convolution layer and point-wise convolution layer that reduces the number of parameters for faster network [8].

Several researches applied deep learning networks for segmentation using DeepLabV3+, for example in [5] it was designed based on Res-Net 50, and Atrous Spatial Pyramid Pooling (ASPP) module with minor modification in its dilated rate. Small dilated rates, including 4, 8, and 12, were used instead of 6, 12, and 18, to capture small morphological details. This model was used to segment three classes: background, lung, covid-19 lesions in CT chest images, which achieved 88.1%, 80%, 90%, and 97.8% dice coefficient, Jaccard coefficient, sensitivity, and specificity, respectively. In addition, a 3D CU-Net was designed by Zheng *et al*.[7] for CT covid-19 segmentation based on the standard 3D U-Net. Skip connections were placed between the encoders and their corresponding decoders using attention gate to improve the feature map representation. Moreover, a pyramid fusion module with dilated convolutions was used to create multiscale information.

The Tversky loss function was applied to resolve irregular distribution and the size of lesions. This 3D CU-Net realized 96.3%, and 77.8% dice similarity coefficients in lung, and COVID-19 lesions, respectively. Double U-Net was implemented by Jha *et al*. [8], which consists of two consecutive U-Nets. The architecture of the first stage U-Net used the pretrained VGG-19 network as an encoder to capture efficient information, then a ASPP module was applied between encoder and decoder to capture high resolution feature maps. The input of the second stage U-Net is the multiplication of the original image and the predicted output of the first network. The decoder of the second network used a skip connection from encoders of the two networks. This DoubleU-Net achieved superior performance compared to single U-Net especially in the case of the CVC-Clinic DB as realized 92.39% Dice coefficient, and 86.11% mean Jaccard.

A MSDC-Net implemented by [9] for COVID-19 lesions segmentation by using U-Net as a backbone, and a multiscale feature capture block to segment different lesions size. Different feature maps from downsampling path were aggregated by multilevel feature aggregate (MLFA) module to improve the performance of the segmentation network. Al-Eiadeh [10] implemented a DeepLabV3 with ResNet-18 for Chest X-ray lung segmentation using ASPP module to provide rich and multi-scale spatial information. This model achieved 96.9%, and 94.1% dice, and Jaccard, respectively. Jabber *et al*. [10] used chest x-ray image to detect COVID-19 patients, Different classification network, namely, Inception V3, ResNet50, Mobile Net, and Xception, were used to

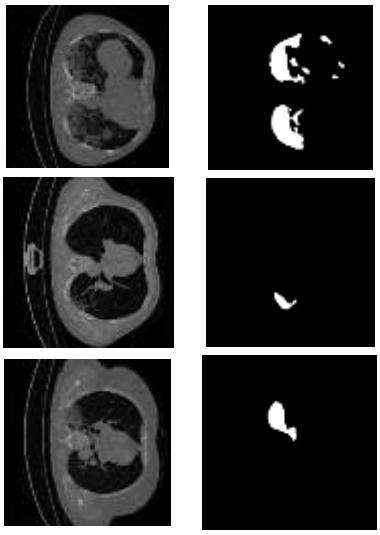
evaluate their performance. Mobile Net achieved the best performance of 98.6% accuracy, 99.3% specificity, and 87.8% precision.

The previous studies showed that the dilated convolution helps to capture information of multi scale and segment lesions of different size. Thus, the ASPP module helps to improve the performance of network, the Mobile Net proved its ability to classified COVID-19 patient which by extracting the significant features. The DeepLabV3+ redesigned with a CNN as an encoder to improve its accuracy. Therefore, in this paper, the DeepLabV3+ was integrated with the MobileNet-V2, then conduced a comparative study between using different CNN, namely ResNet-18, and ResNet-50 with DeepLabV3+ for segmentation. Also, we compared these models with the modified AlexNet for segmentation, and the basic U-Net structure.

2.MATERIAL AND METHODS

A. Material

In this study, a public COVID-19 CT image dataset consists of 20 labelled COVID-19 volumetric CT scan for each patient. Each CT scan has nii extension format. The, 2D images of each patient are extracted from their corresponding volumetric CT form leading to 3436 CT slices [11]. The sizes of the original images area 512*512 or 630*630. The CT images and their ground truth are 256 gray levels images as they are 8-bits images. In this paper, the most significant CT images/slices that show obviously the infection lesions 2D images from the CT scans were selected, which are 120 CT images. These selected images are used for the augmentation process. Fig. 1 shows samples of COVID-19 lung CT images, and their corresponding ground-truth.



(a) (b)

Figure 1. Sample images from the used dataset, where (a) COVID-19 lung CT images, and (b) their corresponding ground-truth of the infected lesion.

B. Methodology

Semantic segmentation divides the image into regions based on the features, such as grayscale, geometric shapes, and colour. The encoding part is responsible of feature extraction, where several CNN architectures can be used. Typically, the CNN includes encoder and decoder parts for segmentation, however, the model's architecture has an effect on extracted features [12]. Also, increasing number of layers for deeper network and meaningful features may cause overfitting problem. In this paper, a DeepLabV3+ integrated with MobileNet-V2 was implemented. In addition, comparisons between this combination against using other different CNNs with DeepLabV3+ were conducted, including DeepLabV3+ with Resnet18, and DeepLabV3+ with Resnet50. Furthermore, a comparison with the traditional AlexNet, and traditional U-Net was conduct to evaluate the superiority of the proposed network for segmenting COVID-19 lesions.

The DeepLabV3+ combined with MobileNet-V2 consists of encoder and decoder parts, where i) the encoder is based on MobileNet-V2 as inverted residuals and linear bottlenecks that consist of depth-wise separable convolutions to extract feature using few number of weights achieving low computational cost, ii) the Atrous Spatial Pyramid Pooling (ASPP) module, and the iii) decoder. The different stages in the implemented model are as follows.

a. Pre-processing stage

An augmentation process was applied on the 120 selected CT images by rotating a set of the original images at different angles, namely 10, 20, 30, 180, -10, -20, and -30 to obtain 960 images, as the seven rotation angles providing 840 augmented images from the 120 CT images. In addition, a resizing to 224×224 of all used images (i.e. CT images, and their corresponding ground-truth ones) was conducted. Besides, the dataset was split randomly into 70% training subset, and 30% (15% validation subset and 15 % testing subset).

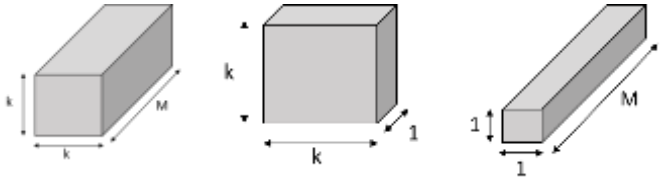
b. The combined DeepLabV3+ and MobileNet-V2 for COVID-19 segmentation

In this section, the main component of the combined DeepLabV3+ and MobileNet-V2 is described to explain its advantages. The combined model architecture consists of inverted residuals and linear bottlenecks as an encoder, ASPP module, and a decoder.

1. Inverted residuals and linear bottlenecks encoder

(a) Depth-wise separable convolutions

The depth-wise separable convolutions are the main components of MobileNet by replacing and factorizing the standard convolution layer by two separated convolution layers, namely, depth-wise convolution layer, and point-wise convolution layer. The depth-wise convolution layer (DW) applies a 2D kernel to filter each channel of the input feature map using number of the required kernels equals to the number of input channels. Then, the point-wise convolution layer (PW) applies a 1×1 convolution layer along the depth (channel) to combine the output C. The main advantage of using the depth-wise separable convolution instead of the standard convolution is the reduction of the computational cost[13]. Fig. 2 describes the used kernel at the standard convolution layer, and for the depth-wise separable convolution layer.



(a) (b.1) (b.2)

Figure 2. The size of the used kernel in: (a) the standard convolution layer, and (b) the depth-wise separable convolution layer, where (b.1) the depth-wise convolution layer, and (b.2) the point-wise convolution layer.

Figure 2 (a) illustrates that if the standard convolution layer applies N numbers of kernels size $K \times K \times M$ on the input feature map with size $H \times W \times M$, it produces an output C of size $H \times W \times N$. Thus, the computational cost of the standard convolution C_1 is defined as follows [13]

$$C_1 = K \times K \times M \times N \times H \times W \quad (1)$$

where H, W are the height and width of the input feature map, respectively, M is the number of channel in the input feature map, N is number of kernels, K is the height and width of the applied filter at $S = 1, S$ is the stride (step size) which the kernel used to shift over the input feature map. In addition, if the depth-wise separable convolution is used, the depth-wise convolution layer with M number of kernels size $K \times K \times 1$ as shown in Fig. 2 (b.1) on the input feature map with size $H \times W \times M$, an output size $H \times W \times M$ is produces assuming also that $S = 1$ with applying the same convolution. Afterward, the PW layer is applied on the output of DW layer with N number of kernels size $1 \times 1 \times M$ as shown in Fig. 2 (b.2), which produces an output size $h \times w \times N$ assuming $S = 1$ and apply same convolution. The computational cost of the depth-wise separable convolution C_2 is defined as [14]

$$C_2 = (K \times K \times 1 \times M \times H \times W) + (1 \times 1 \times M \times N \times H \times W) \quad (2)$$

For determining the reduction percentage of computational cost R, divide C_2 by C_1 as [15]

$$R\% = \left(\frac{1}{N} + \frac{1}{K^2}\right) \times 100 \quad (3)$$

By assuming $N = 32$, and $K = 3$, the depth-wise separable convolution requires only 14.2% of the require computation in the case of standard convolution. Consequently, the consumption time is approximately seven times faster than the standard convolution.

(b) Inverted residuals and linear bottlenecks

For using real input images, the layer activations sets can be represented as a manifold of interest in low- dimensional subspaces to reduce the dimension space along the network for reducing the computational cost. The problem of reducing the channel numbers appears due to the use of non-linear transformation, such as Relu activation function, which transforms negative values to zero causing information loss. Nevertheless, using large lot number of channels will lead to saved information in other channels [6].

To solve the trade-off between the computational cost and the performance, a width multiplier parameter α is used, which has the range $]0,1]$. This α parameter represents the amount of reduction in the channels number; thus, the number of kernels is reduced by α factor. So, the computational cost after applies channel shrinking C^{α} is defined as follows [16].

Using mathematical approximations, the width parameter reduces the computational cost by a factor α^2 as follows:

$$C^{\alpha} = (K \times K \times 1 \times \alpha M \times H \times W) + (1 \times 1 \times \alpha M \times \alpha N \times H \times W) \quad (4)$$

$$C^{\alpha} = \alpha^2 C_2 \quad (5)$$

The low dimension channel is able to describe the necessary information, but to prevent the information loss due to the nonlinear transformation, a lot of channels is required. So, a linear bottleneck block is designed as shown in Fig. 3.

Figure 3 shows the start and the end of the block, which acts as bottlenecks. It consists of convolution layer with low number of channels, but in the middle of the block an expansion process is applied by PW layer to increase the number of channel and apply nonlinear transformation without information loss. Another PW layer is applied at the end of block to reduce number of channels again, where d is the reduced number of channels, t is the expansion factor.

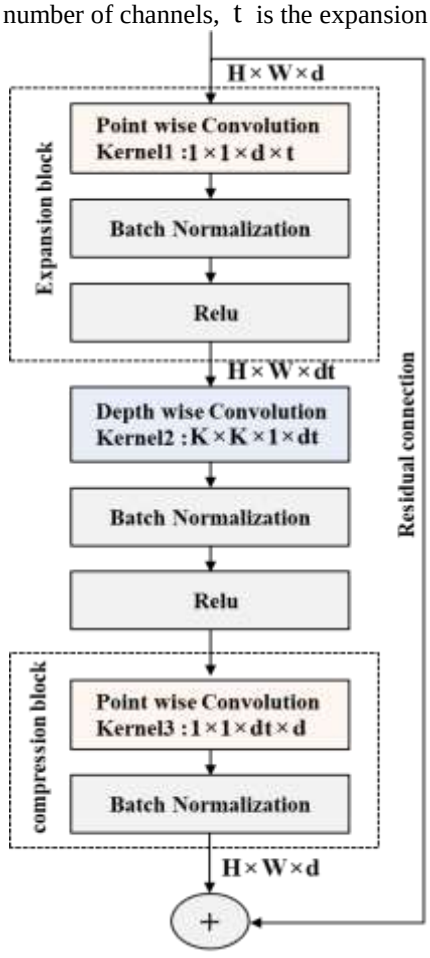


Figure 3. Inverted residuals and linear bottlenecks.

As the input dimension of the block equal the output dimension of the block, a residual connection is applied over the block to solve the vanishing and exploding Gradients problem in the deeper network. It is called inverted residual as the residual connection is applied between two thin (bottlenecks) maps, unlike, in the traditional residual, where it is applied between the thick maps.

Since the more complex and deeper network are more accurate, when using depth-wise separable convolution layer instead of the standard convolution layer, the DW separates from the PW layer by batch normalization layer and Relu activation function. The batch normalization layer and Relu activation function are applied two times in the depth-wise separable convolution layer, which compensates the reduction of complexity due to the reduction of weights.

2. Atrous Spatial Pyramid Pooling (ASPP) module

The output of the encoder represents a small feature map which is suitable for classification problem, but the spatial information about small region usually lost. Although, in the case of segmentation problem, it requires increased feature maps resolution and receptive field of neurons. So, an Atrous convolution (Dilated convolution) is used to insert zero values between the weights of the filters for increasing the resolution of the image without the need to increase the number of parameters [5]. Fig. 4 describes the ASPP module.

Figure 4 shows that The ASPP is applied on the output of the encoder, which consists of four parallel dilated depth-wise separable convolution layers, each with different dilated rate. It helps in capturing multi-scale information and patterns depending on the different filter's views. The outputs of the four parallel branches are concatenated. In the proposed model, the dilated rates in the DeepLabV3+ combined with MobileNet-V2 are 6, 12, and 18.

3. Decoder

The DeepLabV3+ decoder is employed to restore the image size for producing the segmented image. Figure 5 describes the decoder architecture inside the DeepLabV3+ combined with MobileNet-V2.

Figure 5 shows that the output feature map of the ASPP module subject to convolution layer, then unsampled by a factor 4 using transpose convolution layer. Subsequently, this upsampled feature map concatenated with low level feature map from the earlier block which subjected to convolution layer to improve the efficiency of training process. The low-level feature maps are used due to its high level of spatial information. The output of concatenation subject to the depth-wise separable convolution and applies upsampled by a factor 4. Finally, a SoftMax and pixel classification layer are applied to generate the final segmented image [17].

4. Performance Metrics

The segmentation networks are evaluated different performance metrics to compare the output segmented images and their corresponding ground-truth masks. A confusion matrix is used which represents the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The positive class represents COVID-19 pixel, and the negative class represents background pixel.

The global accuracy denotes to the ratio between the number of correctly predicted pixels whether COVID-19 lesion class or non-COVID-19 (background) class to the total number of pixels along the whole number of images. It can be formulated as follows [18]:

$$\text{Global Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

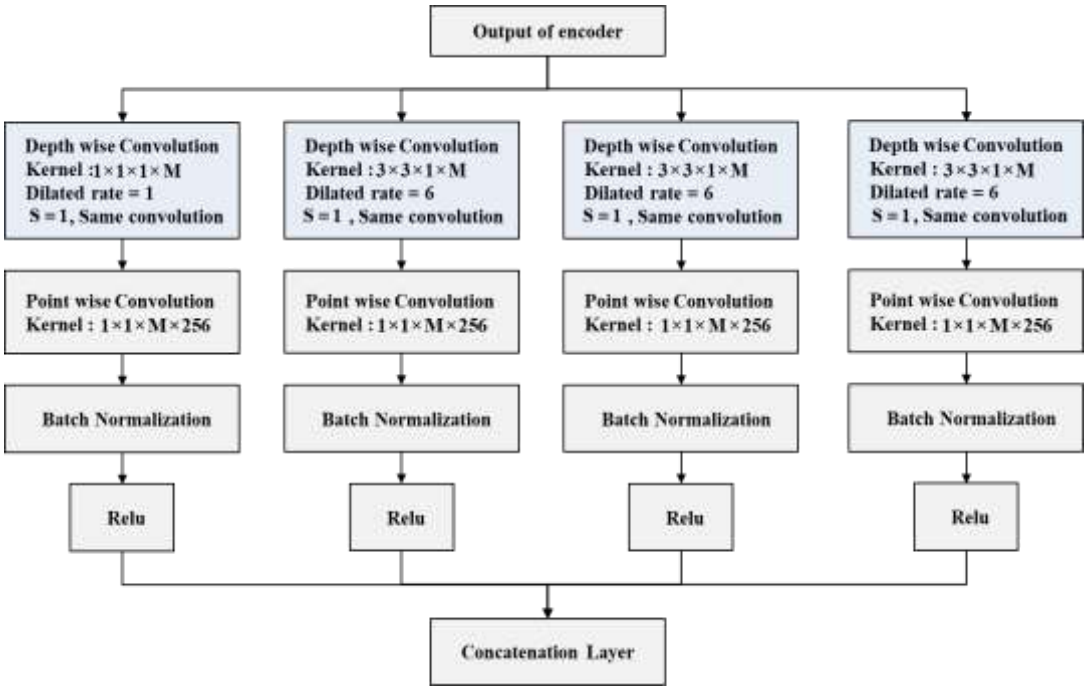


Figure 4. Atrous Spatial Pyramid Pooling (ASPP) module.

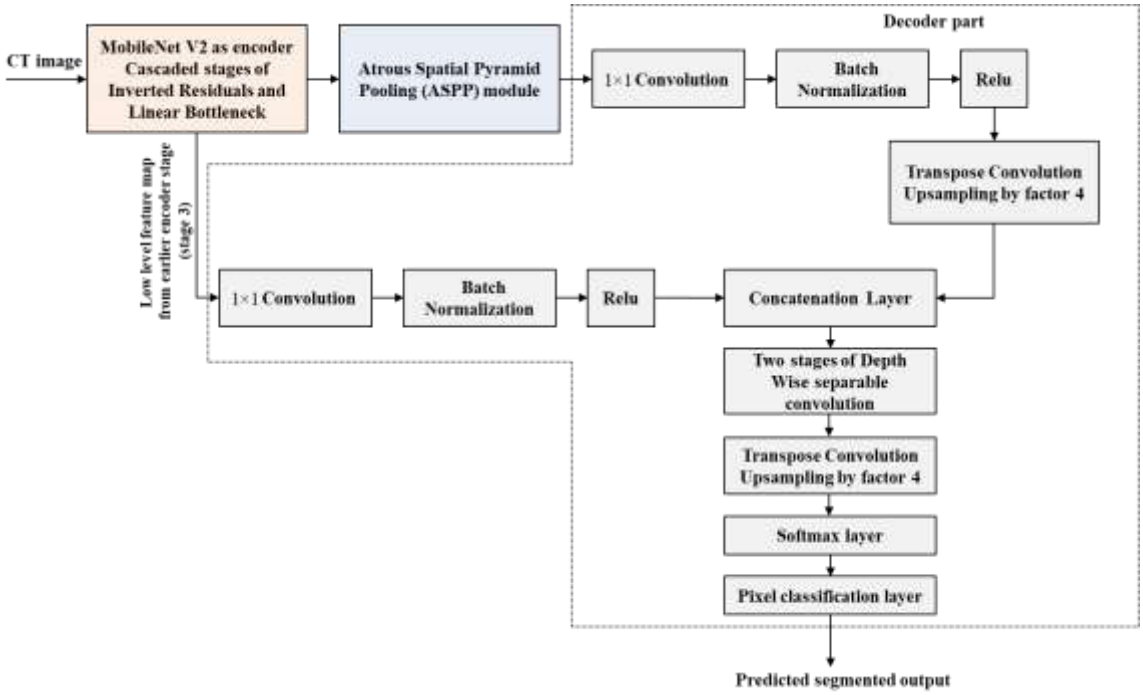


Figure 5. The construction of the DeepLabV3+ combined with MobileNet-V2.

The sensitivity represents the number of positive class pixels which predicted correctly, the specificity represents the number of negative class pixels which predicted correctly, the precision refers to the number of pixels predicted in the positive class and are really in positive class, which is given in [18]:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

The Jaccard and the dice indexes measure the similarity between output segmented images and their corresponding ground-truth masks, which is given in [19]:

$$\text{Jaccard} = \frac{TP}{TP + FP + FN} \quad (10)$$

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (11)$$

3. EXPERMINTEL RESULTS

In this section, an experimental evaluation of the DeepLabV3+ based segmentation network was conducted to segment COVID-19 lesions compare to the corresponding ground-truth. MATLAB R2021a was used for implementing and training the network on PC with the following specifications: AMD Ryzen 7 5800H processor with Radeon Graphics 3.20 GHz (16 GB RAM), and NVIDIA GeForce RTX 3060 GPU (6 GB RAM) using Microsoft Windows10 operating system. The pre-processed output CT images after augmentation using rotation of the image clockwise around its centre point by different rotation angles are shown in Fig.6.

A. Using the DeepLabV3+ combined with MobileNet-v2 for segmentation

In this section, the results of using the combined DeepLabV3+ with MobileNet-v2 are reported. This model was trained using the training dataset, where network was trained along eight epoch and eight minimum batches. Table 1 describes the parameters and their settings, which are used in the proposed model. Fig. 7 shows the model's convergence analysis, including the accuracy and loss function rates of the proposed model along the training and validation stages. The blue line represents the accuracy along different epochs during training stage, the red line represents the loss function value during the training stage, and the dash line represents accuracy and loss function during the validation stage.

Figure 7 demonstrates that the validation accuracy reaches 99.61%, and the cross-entropy loss function reaches about 0.01 as the network's weights are updated along the training phase to reduce the loss function.

Table 1. The used parameters and settings in the DeepLabV3+MobileNet-v2 model.

Parameter	
Learning rate	0.001
Optimizer	Adam
Loss Function	Cross-entropy
Number of epochs	8
Min Batch size	8
Dilated rates used in ASPP module	6,12,18

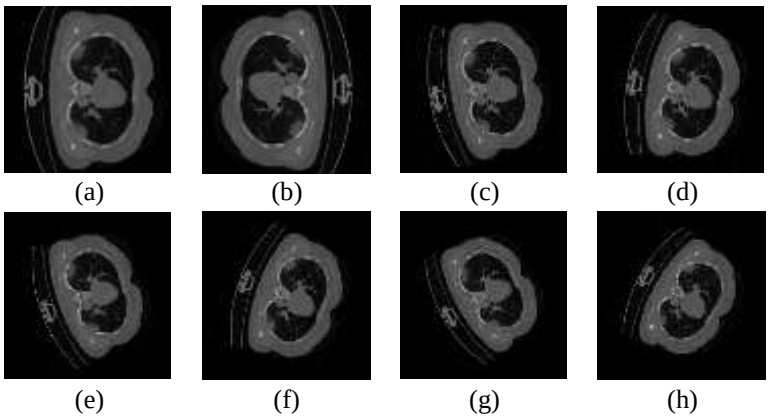


Figure 6. The pre-processed output CT images after augmentation, where: (a) original COVID-19 lung CT image, (b) CT image rotated by angle 180 degree, (c) CT image rotated by angle 10 degree, (d) CT image rotated by angle -10 degree, (e) CT image rotated by angle 20 degree, (f) CT image rotated by angle -20 degree, (g) CT image rotated by angle 30 degree, and (h) CT image rotated by angle -30 degree.

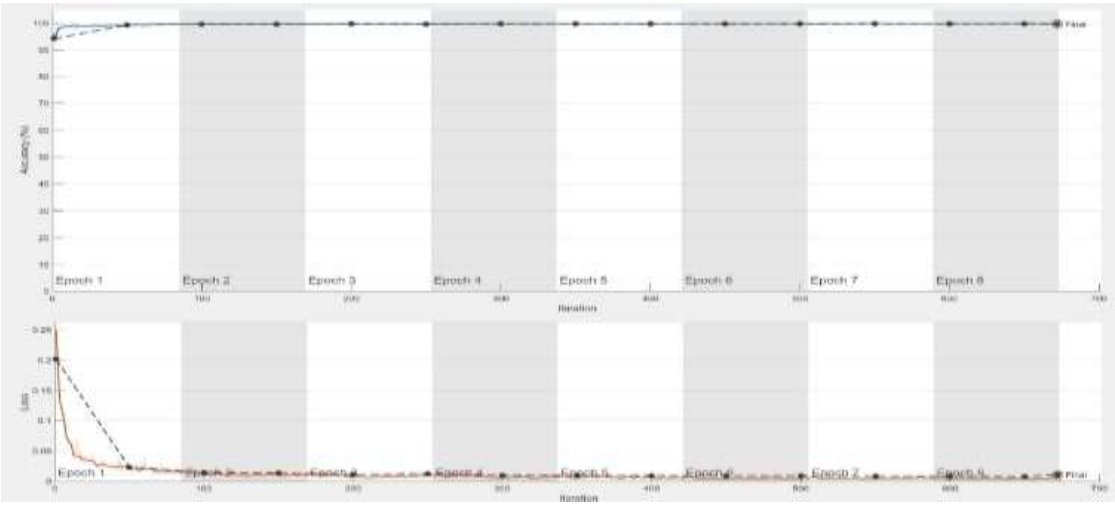


Figure 7. The accuracy and loss function rates of the DeepLabV3+ combined with MobileNet-v2 along different epochs during the training and validation phases.

Table 2. Performance metrics of the DeepLabV3+ MobileNet-v2.

Network	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score	weighted Dice	Sensitivity	Specificity	Precision	Dice	Jac
										Mean ± STD	Mean ± STD
Combined DeepLabV3+ with Mobilenetv2	99.6%	97.5%	89.7%	99.3%	92.4%	97.5%	95.2%	99.7%	99.7%	83.7% ± 4.9%	72% ±6.8%

B. Visual results of DeepLabV3+MobileNet v2 segmentation-based deep learning network

The DeepLabV3+ combined with MobileNet-V2 trained for COVID-19 lesions segmentation. Figure 8 demonstrates the output of the proposed network showing the segmented COVID-19 lung CT original /augmented images, and their corresponding ground-truth.

Figure 8 illustrates that the dice coefficients showing the similarity between the predicted segmented images in Fig. 8 (c1-c3) and their corresponding ground-truth masks in Fig. 8(d1-d3) are 94%, 90%, and 92%, respectively. In addition, Table 1 demonstrates the performance evaluation metrics of the DeepLabV3+ combined with MobileNet-V2.

Table 2 shows the performance metrics, including the global accuracy, mean accuracy, mean IOU, weighted IOU, mean BF score, weighted dice, sensitivity, specificity, precision, dice, and Jac. The implemented model achieves a reliable performance of 95.2% sensitivity, which means that 95.2% of COVID-19 lesion pixels are correctly segmented. Also, 99.7% specificity, which means that 99.7% of background pixels are correctly segmented. Moreover, high values of weighted IOU and weighted dice of 99.3%, and 97.5%, respectively, were realized.

C. Comparative study with other well-known deep learning networks

A comparative study was conducted between the DeepLabV3+ combined with MobileNet-V2 against other well-known deep learning networks, namely DeepLabV3+ with Resnet18, DeepLabV3+ with Resnet50, AlexNet, and U-Net. Figure 9 demonstrates visually the segmented output using the previous segmentation models, and their ground-truth masks.

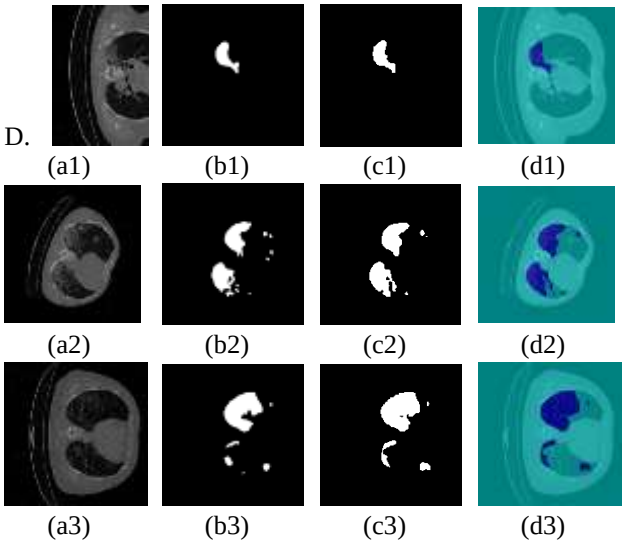


Figure 8. The outputs of the proposed segmentation network on original/augmented CT images, where: (a) COVID-19 lung CT image, (b) ground-truth corresponding to each CT image, (c) the segmented COVID-19 lesion output using the proposed system, and (d) the segmented lesion laid on the original CT image.

The visual comparison between the predicted images and ground-truth masks is shown in Fig. 9 proving that the combined DeepLabV3+ with MobileNet-V2 model outputs are the most compatible segmented images to their corresponding ground-truth masks. Fig. 9 (c) is the segmented lesion using the DeepLabV3+ with Resnet18 showing smooth outputs, but large number of COVID-19 pixels were not classified correctly. Fig. 9 (d) shows that the DeepLabV3+ with Resnet 50 has the ability to classify larger number of COVID-19 pixels during the segmentation of the lesion as this network is deeper than DeepLabV3+ with Resnet18. However, the segmented lesions' borders does not achieve the perfect match with the ground-truth borders and does not segment the small lesions. Fig. 9 (e) illustrates that the segmented lesion using modified Alex has the worst output between models as it classifies falsely most pixel as COVID-19 which is unacceptable. Fig. 9 (f) shows that the segmented lesion using U-Net appears as discrete points and fails to segment the COVID-19 lesion. Fig. 9 (g) is the segmented lesion using the combined DeepLabV3+ with MobileNet-V2 laid on the CT image, the segmented lesions are smooth and closely matches their ground-truth masks, and DeepLabV3+ with MobileNet-V2 can segment the small lesions. Moreover, Table 2 illustrates a comparative study in terms of the performance evaluation metrics of the different combinations with the combined DeepLabV3+ with MobileNet-V2.

The comparison in Table 3 by measuring the performance metrics illustrates that using the combined DeepLabV3+ with MobileNet-V2 model improved the dice index to 84% with a small standard deviation value 4.9%, and the Jac index is also improved to 72% with a standard deviation 6.8%. Also, 97.5% mean accuracy and 95.2% sensitivity were achieved. These results are the highest compared to the other models except Alex model which achieved 96.9% sensitivity. Although Alex model is unacceptable due to the low value of specificity 93.4%. So, the combined DeepLabV3+ with MobileNet-V2 solves the trade-off between the sensitivity 95.2%, and the specificity 99.7%.

4. DISCUSSION

A comparison between the combined DeepLabV3+ with MobileNet-V2 model and other existing studies used the same dataset is carried out to assess the performance of the semantic segmentation model. A semantic segmentation network based on U-Net was presented by Khalifa *et al*. [18] using a network consisted of 3 encoder and decoder stage and start with 64 number filters. A two sequential segmentation networks designed by Xie *et al*. [19] to segment the lung in CT images using the first network and segment the COVID-19 lesion from the segmented lung by the second network.

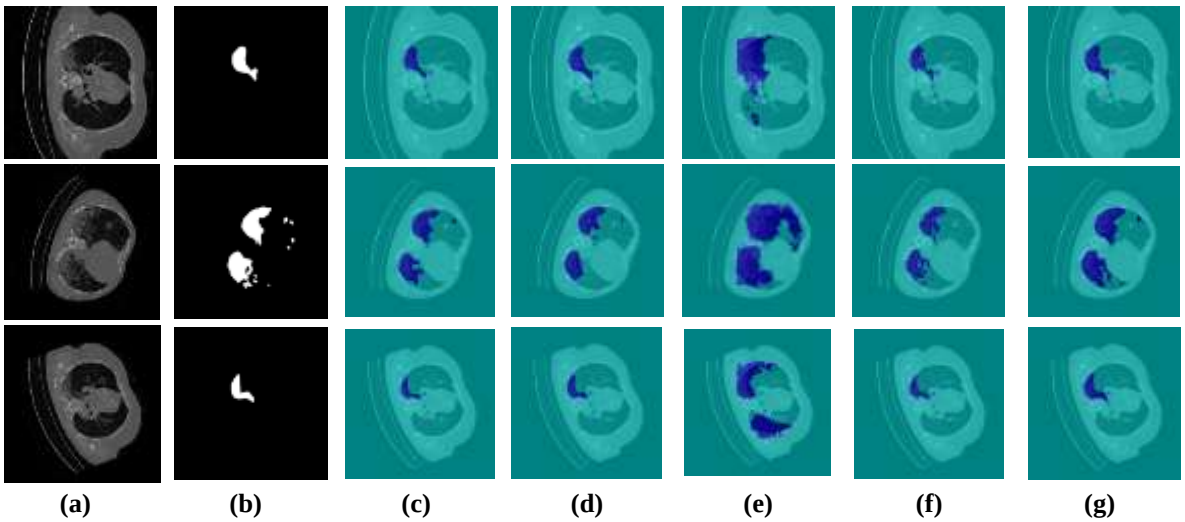


Figure 9. The segmented output using of the different models, where: (a) COVID-19 lung CT images, (b) ground-truth masks, (c) segmented lesion of DeepLabV3+ with Resnet18 laid on the CT image, (d) segmented lesion of DeepLabV3+ with Resnet50 laid on the CT image, (e) segmented lesion of modified Alex laid on the CT image, (f) segmented lesion of U-Net laid on the CT image, and (g) segmentation of lesions using the DeepLabV3+ with MobileNet-V2 laid on the CT image.

Table 3. Performance metrics of the combined DeepLabV3+ with MobileNet-v2, Resnet18, Resnet50, Alex, and U-Net.

Network	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score	weighted Dice	Sensitivity	Specificity	Precision	Dice	Jac
										Mean ± STD	Mean ± STD
DeepLabV3+ with Resnet18	99.5%	86.9%	85.2%	99%	91.1%	88.3%	73.8%	**99.9%**	**99.9%**	67% ± 9.3%	0.51 ± 10%
DeepLabV3+ with Resnet50	99.6%	89.1%	87.8%	99.2%	91.3%	90.1%	78.3%	**99.9%**	**99.9%**	7% ± 10%	0.54 ± 12%
Modified Alex	93.4%	95.2%	56.4%	92.1%	58.1%	95.1%	**96.9%**	93.4%	93.6%	39% ± 14.3%	0.25 ± 11%
U-Net	99%	82%	75%	98%	80%	98.4%	63%	99.6%	99.4%	58%±16%	0.43±15%
Combined DeepLabV3+ with Mobilenetv2	**99.6%**	**97.5%**	**89.7%**	**99.3%**	**92.4%**	**97.5%**	95.2%	99.7%	99.7%	**84%** ±4.9%	**72%** ±6.8%

Table 4. Comparison between the DeepLabV3+ with MobileNet-V2 model and existing studies used the same dataset.

Method	Accuracy %	Dice %	IOU %	Precision %	Sensitivity %	Specificity %
Khalifa *et al*. [18]	99.00	97.00	98.86	99.00	95.00	Not given
Xie *et al*. [19]	99.06	87.06	77.09	Not given	90.84	99.58
Combined DeepLabV3+ with Mobilenetv2	**99.60**	**97.50**	**99.30**	**99.70**	**95.20**	**99.70**

To assess the performance of the semantic segmentation model in the comparison previous studies different performance parameters, such as the global accuracy, precision, sensitivity, specificity, dice coefficient, and Jaccard coefficient, were measured in Table 4. Table 3 indicates that the performance of the model by Khalifa *et al.* [18] achieved 99% accuracy, 97% dice, 98.86% IOU, 99% precision, and 95% sensitivity due to using the direct skip connection which affected relatively the performance of the network as the skip connection transmit low features maps from earlier layer. Also, in the model designed model by Xie *et al.* [19] the performance affected by the complexity of model architecture 99.06% accuracy, 87.06% dice, 77.09% IOU, 90.84% sensitivity, and 99.58% specificity.

By comparing the results of the DeepLabV3+ with MobileNet-V2 model and existing studies used the same dataset, the comparison indicates that the proposed model achieves the best segmentation performance as indicated in all the performance metrices. It achieves 99.6% accuracy, 97.5% dice, 99.3% IOU, 99.7% precision, 95.2% sensitivity, and 99.7% specificity. The DeepLabV3+ with MobileNet-V2 model provides the ability to predict 95.2% of covid-19 pixels correctly. Also, 99.7% of background pixels correctly which achieves a high dice coefficient 97.5%.

The combined DeepLabV3+ with MobileNet-V2 model has high performance due to its structure as it uses MobileNet-V2 as encoder for extract features which based on inverted residuals and linear bottlenecks that used the depth-wise separable convolution layers to decrease numbers of learnable parameter and make network faster, and also has a residual connection along the block which solve the problem of overfitting. A ASPP module which consist of parallel branches of dilated convolution layers with dilated rate 6, 12, and 18 to capture multi-scale information and increase feature maps resolution.

5. CONCLUSION

Due to the rapid spreading of COVID-19 virus in all over the world. A need for a method that diagnosis infected patients earlier and take precautions measures as earlier as possible to reduce the spreading. Chest CT scans represent 3D representation of lung and also covid-19 lesions appear in it. As CT scans has a high sensitivity, availability, and low cost compared to RT-PCR kits. So, researchers become interested to design an automatic semantic segmentation network by the help of chest CT images to segment COVID-19 lesion.

The combined DeepLabV3+ with MobileNet-V2 model is a semantic segmentation model that combines between MobileNet-V2 (classification network), and DeepLabV3+ (segmentation network). The MobileNet-V2 was used as an encoder for extracting features and the inverted residuals and linear bottlenecks of MobileNet-V2 reduce number of learnable parameter due to the use of depth-wise separable convolution layers instead of standard convolution. and solve the problem of vanishing and exploding Gradients as it has a residual connection between the thin bottlenecks of each block.

Then DeepLabV3+ was used the extracted features from encoder part to produce the predicted segmented image. It applies ASPP module to capture a multi scale information, the makes up-sampling process to reach to the segmented output image with the help of earlier feature maps that has a highly sensitive spatial information.

By solving the trade-off between all performance metrics, the experimental results proved that The Proposed DeepLabV3+ MobileNet-V2 model achieves the best performance along all metrics. It achieves 99.6% accuracy, 97.5% dice, 99.3% IOU, 99.7% precision, 95.2% sensitivity, and 99.7% specificity.

Funding: This research has not received any type of funding.

Conflicts of Interest: The author declares that there is no conflict of interest.

REFERENCES

[1] Oulefki, A., Agaian, S., Trongtirakul, T. and Laouar, A.K., 2021. Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images. Pattern recognition, 114, p.107747.

[2] Ashour, A.S., Eissa, M.M., Wahba, M.A., Elsawy, R.A., Elgnainy, H.F., Tolba, M.S. and Mohamed, W.S., 2021. Ensemble-based bag of features for automated classification of normal and COVID-19 CXR images. Biomedical Signal Processing and Control, 68, p.102656.

[3] Zhao, W., Jiang, W. and Qiu, X., 2021. Deep learning for COVID-19 detection based on CT images. Scientific Reports, 11(1), pp.1-12.

[4] Polat, H., 2022. Multi-task semantic segmentation of CT images for COVID-19 infections using DeepLabV3+ based on dilated residual network. Physical and Engineering Sciences in Medicine, pp.1-13.

[5] Polat, H., 2022. A modified DeepLabV3+ based semantic segmentation of chest computed tomography images for COVID- 19 lung infections. International Journal of Imaging Systems and Technology, 32(5), pp.1481-1495.

[6] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520.

[7] Zheng, R., Zheng, Y. and Dong-Ye, C., 2021. Improved 3D U-Net for COVID-19 chest CT image segmentation. Scientific Programming, 2021.

[8] Jha, D., Riegler, M.A., Johansen, D., Halvorsen, P. and Johansen, H.D., 2020, July. Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)* (pp. 558-564).

[9] Zhang, J., Ding, X., Hu, D. and Jiang, Y., 2022. Semantic segmentation of COVID-19 lesions with a multiscale dilated convolutional network. *Scientific Reports*, *12*(1), pp.1-12.

[10] Jabber, B., Lingampalli, J., Basha, C.Z. and Krishna, A., 2020, December. Detection of covid-19 patients using chest x-ray images with convolution neural network and mobile net. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 1032-1035).

[11] Covid19-CT-Scans; https://zenodo.org/record/3757476#.Yp8mr6gzZ3gn.d.

[12] Liu, X., Song, L., Liu, S. and Zhang, Y., 2021. A review of deep-learning-based medical image segmentation methods. Sustainability, 13(3), p.1224.

[13] Srivastava, H. and Sarawadekar, K., 2020, October. A depthwise separable convolution architecture for CNN accelerator. In 2020 IEEE Applied Signal Processing Conference (ASPCON) (pp. 1-5). IEEE.

[14] Girish, G.N., Saikumar, B., Roychowdhury, S., Kothari, A.R. and Rajan, J., 2019, July. Depthwise separable convolutional neural network model for intra-retinal cyst segmentation. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 2027-2031).

[15] Sun, W., Zhang, X. and He, X., 2020. Lightweight image classifier using dilated and depthwise separable convolutions. Journal of Cloud Computing, 9(1), pp.1-12.

[16] Kamal, K.C., Yin, Z., Wu, M. and Wu, Z., 2019. Depthwise separable convolution architectures for plant disease classification. Computers and Electronics in Agriculture, 165, p.104948.

[17] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).

[18] Khalifa, N.E.M., Manogaran, G., Taha, M.H.N. and Loey, M., 2022. A deep learning semantic segmentation architecture for COVID- 19 lesions discovery in limited chest CT datasets. Expert Systems, 39(6), p.e12742.

[19] Xie, F., Huang, Z., Shi, Z., Wang, T., Song, G., Wang, B. and Liu, Z., 2021. DUDA-Net: a double U-shaped dilated attention network for automatic infection area segmentation in COVID-19 lung CT images. International Journal of Computer Assisted Radiology and Surgery, 16(9), pp.1425-1434.