A modified DeepLabV3+ based semantic segmentation of chest computed tomography images for COVID-19 lung infections

Hasan Polat

Department of Electrical and Energy, Bingol University, Bingöl, Turkey

Correspondence
Hasan Polat, Department of Electrical and Energy, Bingol University, Selahaddin-i Eyyübi Mah. Aydnılk Cad No: 1, 12000 Bingöl, Turkey.
Email: hasanpolat@bingol.edu.tr; hasanpolat1989@hotmail.com

Abstract
Coronavirus disease (COVID-19) affects the lives of billions of people worldwide and has destructive impacts on daily life routines, the global economy, and public health. Early diagnosis and quantification of COVID-19 infection have a vital role in improving treatment outcomes and interrupting transmission. For this purpose, advances in medical imaging techniques like computed tomography (CT) scans offer great potential as an alternative to RT-PCR assay. CT scans enable a better understanding of infection morphology and tracking of lesion boundaries. Since manual analysis of CT can be extremely tedious and time-consuming, robust automated image segmentation is necessary for clinical diagnosis and decision support. This paper proposes an efficient segmentation framework based on the modified DeepLabV3+ using lower atrous rates in the Atrous Spatial Pyramid Pooling (ASPP) module. The lower atrous rates make receptive small to capture intricate morphological details. The encoder part of the framework utilizes a pre-trained residual network based on dilated convolutions for optimum resolution of feature maps. In order to evaluate the robustness of the modified model, a comprehensive comparison with other state-of-the-art segmentation methods was also performed. The experiments were carried out using a fivefold cross-validation technique on a publicly available database containing 100 single-slice CT scans from >40 patients with COVID-19. The modified DeepLabV3+ achieved good segmentation performance using around 43.9 M parameters. The lower atrous rates in the ASPP module improved segmentation performance. After fivefold cross-validation, the framework achieved an overall Dice similarity coefficient score of 0.881. The results demonstrate that several minor modifications to the DeepLabV3+ pipeline can provide robust solutions for improving segmentation performance and hardware implementation.

KEYWORDS
computed tomography, COVID-19, deep learning, DeepLabV3+, ResNet, segmentation

1 | INTRODUCTION

The coronavirus disease (COVID-19) negatively affects the lives of billions of people around the world. The infection can cause many severe types of pneumonia that can result in death.1,2 COVID-19 has the ability to be transmitted from person to person through various ways, such as respiratory droplets and close contact.3 High
mobility in the human population yields an exponential increase in the spread of infection. The World Health Organization (WHO) declared the outbreak as a “Public Health Emergency of International Concern” on the January 30, 2020 and as a pandemic on the March 11, 2020. As of the beginning of 2022, approximately 373 million positive cases worldwide have been confirmed by WHO. Unfortunately, 5 658 702 of the positive cases resulted in death. It reveals the importance of the models proposed in the fight against COVID-19 due to the destructive effects of the outbreak on daily life routines, the global economy, and public health.

The primary assay for the diagnosis of COVID-19 is based on reverse transcription-polymerase chain reaction (RT-PCR). Although the RT-PCR test is considered the gold standard, previous studies have reported that the diagnostic test has low sensitivity and is time-consuming. The high rate of false-negative results in applied pathological tests reinforces the tendency for alternative methods for disease diagnosis and quantifying the severity of the infection. In this context, medical imaging techniques such as X-ray and computed tomography (CT) provide a robust solution for comprehensive analysis. Because of its relative ease of use and high sensitivity compared with RT-PCR, the chest CT scan is recommended by radiologists. The CT imaging technique is suitable not only for diagnosis but also for quantitative analysis of COVID-19 infections.

The pulmonary infections may vary according to the course of the disease. Ground glass opacities (GGOs), often observed in the early stage of the disease, are followed in the later stages by more characteristic findings such as consolidation, pleural effusion, crazy pavement, and reverse halo signs. In particular, GGO and consolidation are considered part of the diagnostic criteria and are widely used. The incidences of GGO and consolidation were reported as 86% and 29%, respectively. CT findings of COVID-19 lesions must be accurately identified, localized, and quantified by radiologists so that an early diagnosis can be made and effective treatments applied. However, manual analysis, which has to be meticulously performed by radiologists, is unfortunately tedious and time-consuming. In addition, the continuous increase in the number of confirmed and suspected cases during the epidemic causes intense pressure and excessive workload for radiologists. The development of automated diagnosis and segmentation models is critical to accelerate CT-based diagnosis and improve access to treatment.

In recent years, deep learning-based artificial intelligence (AI) technologies have made tremendous progress in many pattern recognition tasks such as object detection and speech recognition. Convolutional neural networks (CNNs), a subtype of deep learning, have attracted a lot of attention due to their breakthrough performance, especially in image and video processing. CNN-based models provide effective solutions for medical images such as early detection of various diseases, boundary tracking of irregularities, and quantitative assessment of lesions. It can be seen that CNN-based methods developed to assist healthcare professionals for COVID-19 commonly focus on the detection of pulmonary infections. These frameworks are designed to automatically extract features and classify image as a single pattern. However, deep learning-based semantic segmentation (SS) of infections will offer more effective solutions for the follow-up of serious complications that may occur in the lungs in parallel with the disease.

The SS is the method of associating each pixel within an image to one of the predetermined classes. In other words, the SS technique is pixel-wise image classification. It can be performed in a highly sophisticated way to predict the pixel labels corresponding to the region of interest. It is a crucial part of the image processing task and allows pixel-wise understanding of the whole image. This detailed pixel-level understanding plays a key role for numerous AI based systems that include scene understanding, human-machine interaction, medical image analysis, and autonomous driving. In the medical field, the SS can help to detect and assess the severity of the disease. The overall methodology of SS approach is to design a structure that extracts features through successive convolutions (encoder part) and uses that information to create a segmentation map as a segmented map (decoder part). Before deep learning, the effectiveness of segmentation models based on classical machine learning depended heavily on hand-crafted features. The fact that meaningful features are problem-specific, especially for medical images, makes the trial-and-error feature extraction approach unfavorable. Deep learning architectures are used by the scientific community as a backbone for segmentation tasks, as they have achieved breakthroughs in the automatic extraction of characteristic features in computer vision tasks. In deep learning-based segmentation pipelines, various CNN architectures form the general backbone in the encoder part. In this way, an attempt is made to exploit the powerful potential of CNN models in the automatic extraction of distinctive features. In this study, the DeepLabV3+ architecture, a redesigned CNN model, was used for SS of COVID-19 CT images. The pre-trained ResNet-50 architecture based on dilated convolutional operations was used as a feature extractor to extract contextual information from chest CT images. ResNet uses residual blocks to overcome the issue of vanishing and exploding gradients. Furthermore, these residual blocks are easier to optimize and can gain...
accuracy from considerably increased depth.22,27,28 The current advantages of the ResNet network, especially for applications with limited processing resources, were the driving force behind the selection of the backbone architecture. In the atrous spatial pyramid pooling (ASPP) module, lower atrous rates were applied compared to the original module. Thus, it is aimed at semantically more powerful segmentation of small and variable size COVID-19 lesions. In order to observe the effectiveness of the modified framework, the performance of the original DeepLabV3+ model in the segmentation of COVID-19 lesions was also evaluated. The main contributions of this paper can be summarized as follows:

1. The efficacy of the modified DeepLabV3+ algorithm, which incorporates lower ASPP rates in segmenting COVID-19 lesions, was introduced.
2. Although the convolution and pooling operations in the encoder part provide rich information, boundary information about the regions of interest can be lost. This paper demonstrated that the atrous convolution approach can minimize the loss of semantic information.
3. The proposed framework provides more effective and robust solutions in the segmentation of COVID-19 lesions from chest CT images than various state-of-the-art methods such as U-Net, and SegNet commonly used in the previous works.
4. The proposed modified DeepLabV3+ was compared with other popular segmentation models in terms of the number of learnable parameters used and segmentation performance. Thus, the feasibility of real-time segmentation applications can also be evaluated.

The rest of the paper is organized as follows: Section 2 presents the state-of-the-art segmentation and detection frameworks proposed in the literature to fight against COVID-19 outbreak. Details about the methods applied in line with the determined targets and the axial CT dataset used in the study are shared in Section 3. Experimental results obtained by the proposed model are extensively covered in Section 4. Based on the results obtained, the efficiency of the segmentation method, the factors affecting the performance and the benefits of the proposed model are shared in Section 5. In Section 6, some concluding remarks are given.

2 | RELATED WORK

Since the breakthrough of various CNN architectures for computer vision, the deep learning approach has become one of the most accurate and widespread methods for automated medical image analysis.9,29 Deep learning-based medical image analysis can be divided into two main areas: semantic segmentation and conventional classification. While classification involves assigning the image as a single pattern to a predefined class, the SS involves a pixel-level classification task. Previous works commonly focused on identifying COVID-19 at the lung level,30,31 but pixel-wise segmentation of pulmonary infections was relatively limited.

In order to fight against COVID-19, a comprehensive analysis of CT abnormalities of infections is required. The SS approach can enable sophisticated analysis for pulmonary infection in radiographic images. This approach can depict the regions of interests such as lung, lobes, bronchopulmonary segments, and infected regions or lesions, in the chest CT images.32 For this purpose, several deep learning-based robust segmentation frameworks have been proposed in the literature. The proposed models are mainly U-Net,33 SegNet,34 fully convolutional neural network (FCN)35 and optimized architectures based on these models. Amyar et al.19 proposed a novel multi-task deep learning framework to identify and segment COVID-19 infections. The segmentation pipeline was based on a U-Net but used a convolutional operation by stride factor of 2 instead of pooling. They obtained the best segmentation performance with 256 × 256 and 512 × 512 image resolutions. Zheng et al.36 proposed a variant of U-Net model named 3D CU-Net for automatic segmentation of COVID 19 lesions from 3D chest CT scans. They developed an attention mechanism to reach local cross-channel information interaction in an encoder. Thus, the encoder part enabled a multi-level feature representation. Thanks to the design of a pyramid fusion module with expanded convolutions at the end of the decoder part, segmentation performance was increased by almost 7% compared to conventional U-Net. Voulodimos et al.37 explored the efficiency of various state-of-the-art segmentation architectures such as U-Net and FCN. They pointed out that FCN achieved robust segmentation performance despite the class imbalance in the dataset and numerous annotation errors at the boundaries of symptom manifestation areas. Jin et al.38 proposed a multi-task model to support physicians in the early diagnosis and follow-up of COVID-19 lesions. The proposed pipeline was suitable for segmentation and classification tasks. For the segmentation task, the proposed 3D U-Net++ achieved a Dice similarity coefficient (DSC) of 0.754 on 732 cases. Yan et al.10 presented a feature variation (FV) block that adaptively adjusts the global properties of the features to enhance the capability of feature representation. They proposed a progressive ASPP module to fuse features at different scales. The tailored deep network pipeline achieved a DSC performance of 72.6%
in the segmentation of sophisticated infection areas. Müller et al. applied several preprocessing methods and used extensive data augmentation methods to overcome the disadvantage of limited datasets. They artificially increased the number of limited training images using several data augmentation methods based on spatial, color, and noise augmentation. Thus, the proposed 3D U-Net model achieved accurate and robust segmentation performance without overfitting risk. Wang et al. explored the efficiency of transfer learning approaches using 3D U-Net as a standard encoder-decoder method. They proposed a hybrid-encoder strategy based on a multi-lesion pretrained model. Experimental results demonstrated that the segmentation pipeline achieved an overall DSC of 0.704 with better generalization and lower overfitting risk. In addition, information about non-COVID-19 lesions proved to be significantly effective for a robust pre-trained segmentation network. Saood and Hatem evaluated the performance of two well-known structurally-different deep learning techniques in the segmentation of COVID-19 pulmonary lesions. They compared the performance of U-Net and SegNet architectures for two different segmentation strategies, binary and multi-class. The results showed that SegNet had superior binary segmentation capability, while U-Net performed better on the multi-class task. Karthik et al. proposed a contour-aware attention decoder CNN for the segmentation of pulmonary infections. For each decoding step, the network employs a cross-attention model to concatenate this auxiliary encoder feature-set with the incoming feature map at that step. The feature maps produced by the three convolutional branches were converted back to the high-resolution segmentation map. Their framework achieved a DSC performance of 85.43% for a combined CT dataset.

It can be seen that the U-Net, SegNet, and FCN models are frequently used in studies of segmentation of COVID-19 irregularities from chest CT images. These architectures, formerly developed for understanding the road and indoor scene or for various clinical tasks, have been used on chest CT scans to allow rapid diagnosis and treatment. In this study, the efficiency of the DeeplabV3+ model, which outperforms U-Net, SegNet, and FCN architectures in various segmentation problems, was analyzed. In addition, minor modifications were applied to the DeeplabV3+ pipeline to increase segmentation performance.

3 | MATERIAL AND METHOD

3.1 | Dataset of COVID-19 axial CT images

This dataset is one of the first publicly available CT slice sets with annotated COVID-19 infection segmentation. The axial CT dataset is a collection of the Italian Society of Medical and Interventional Radiology (SIRM). It consists of 100 single-slice CT scans from >40 patients with COVID-19 and is presented as a collection of 512 × 512 pixel grayscale images. Ground-truth maps of COVID-19 irregularities annotated by an experienced radiologist are available in NIFTI file format. The radiologist annotated the infections at three levels: GGO (mask value = 1), consolidation (= 2), and pleural effusion (= 3). In addition, lung masks are also available for contribution to automated classification and segmentation tasks.

Segmentation of a medical image is a challenging task because the pixel distribution of classes is usually imbalanced. In this data set, the COVID-19 lesions also have an imbalanced distribution. It has been reported that the infections are generally in the form of GGO and consolidation, whereas pleural effusions are observed much less frequently. In addition, segmentation performance can be misled when each lesion type is considered a separate class, especially since pleural effusion findings are included in very few images. For reliable evaluation of segmentation performance, subdatasets reserved for training, validation, and testing should include all lesion samples. In order to overcome these drawbacks, all lesion types were labeled as COVID class. As a result, this paper carried out a multi-class segmentation task as lung (non-infected tissue), COVID (consisting of GGO, consolidation, and pleural effusion irregularities), and background. In addition, each CT image in the NIFTI file was rearranged in the 3D RGB color space. In this way, various state-of-the-art segmentation architectures can be applied to the dataset. Figure 1 shows an example of the raw CT scan, lung mask, ground-truth map for COVID-19 lesions, and the modified ground-truth map. For three segmentation classes, pixel count (total number of pixels in a class) and image pixel count (total number of pixels in images that had an instance of a class) showed in Table 1. Table 1 shows the comparison of the class sizes of the original dataset with the class sizes of the proposed segmentation.

3.2 | Technical details for modified DeeplabV3+

In this section, the core components of the proposed DeeplabV3+ model are briefly presented. The presentation consists of the backbone architecture as a feature extractor, the ASPP module, and the decoder. It also introduces the modification process of the network in technical detail. The implemented modified segmentation framework is illustrated in Figure 2.
3.2.1 The essential framework

In most models proposed for the SS, the main structure of the pipeline consists of encoder and decoder parts.\textsuperscript{25} The encoder part captures essential information from input images. The size of the feature maps is inevitably downsampled in the last layers of the module to capture the high-level details of the input image. In particular, the performance of CNN models has provided remarkable progress for feature extractors due to rich hierarchical features and an end-to-end trainable framework.\textsuperscript{45} Thus, the encoder part can provide rich information to the next module. In the decoder part, a segmentation map of the same size as the input image is created from the low-dimensional feature maps extracted by sequential deconvolution and upsampling operations.\textsuperscript{43}

FIGURE 1 Dataset samples and the modified ground-truth map. (A) The raw CT scan. (B) Masked lungs where dark gray is left lung, light gray is right lung. (C) Labeled classes where black is background, dark gray is consolidation, light gray is GGO, white is pleural-effusion. (D) Modified ground-truth map where black is background, white is lung (healthy tissue), and red is COVID-19 class.

TABLE 1 The comparison of the class sizes of the original dataset with the class sizes of the proposed segmentation

| Class          | Pixel count | Image pixel count | Class          | Pixel count | Image pixel count |
|----------------|-------------|-------------------|----------------|-------------|-------------------|
| “Background”   | 1.9197e+07  | 2.6214e+07        | “Background”   | 1.9197e+07  | 2.6214e+07        |
| “GGO”          | 1.1965e+06  | 2.5166e+07        | “COVID”        | 1.8199e+06  | 2.5953e+07        |
| “Consolidation”| 5.8921e+05  | 2.0447e+07        |                |             |                   |
| “Pleural effusion” | 34 265     | 6.5536e+06        |                |             |                   |
| “Lung”         | 5.1974e+06  | 2.6214e+07        | “Lung”         | 5.1974e+06  | 2.6214e+07        |

FIGURE 2 The framework of modified DeepLabV3+ for semantic segmentation of chest CT images. $d$ values denote the dilation factors.
3.2.2 Dilated residual network as encoder

The SS refers to assigning an image into several disjointed regions according to features such as grayscale, color, spatial texture, and geometric shapes. In this context, capturing meaningful features from the medical image is crucial and directly affects segmentation performance. Here, the backbone architecture used in the encoder part bears the computational load associated with the feature extraction. Several CNN architectures with various processing units and topological differences have a great potential for effective feature extraction. However, the increasing number of layers in state-of-the-art CNN architectures has a negative impact on hardware implementation. Furthermore, the goodness of the models does not increase in direct proportion to the number of layers. It is because deeper neural networks suffer from the problem of vanishing and exploding gradients. He et al. introduced skip connection or residual network (ResNet) architecture to overcome this problem. They have proven that deeper networks with the optimization capability of evolved residual blocks provide high performance. Despite its deeper architecture, lower computational cost and robust performance are the primary motivations for using the ResNet architecture as the backbone for the SS model in this study.

He et al. proposed several ResNet variants with different number of layers such as ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152. Although the proposed variants contain a different number of layers, their frameworks share a common form as shown in Figure 3. The overall pipeline of each ResNet structure consists of five convolution blocks with output spatial resolution. The number of layers in each convolution block with residual network differs depending on the architecture variant. This overall pipeline produces the final feature map resolution of $16 \times 16$ for an input image of $512 \times 512$. The rate of the final feature map size produced by the encoder to the input image size indicates the downsampling factor or output stride (OS). Thus, feature maps are produced by a downsampling factor of 32 when the original residual network is used.

Though small feature maps are suitable for image classification, spatial information of the small region of interests can be lost. Increasing the resolution of feature maps is accomplished by removing the stride or downsampling operation. However, this solution cause reduction of receptive field of neurons. The balance between the receptive field and the resolution of the feature map can be achieved with dilated convolution. Dilated convolution also called as Atrous convolution was originally developed for wavelet decomposition. The main idea of dilated convolution is to insert zeros between weights in the filters to boost image resolution. Thus, it allows effective feature extraction in backbone architecture. For dilated rates, while $d = 1$ means a standard convolution filter, $d > 1$ means upsampling a convolution filter. Dilated convolution provides a larger receptive field without increasing the number of parameters. Also, increasing the receptive field provides rich information to predict the image detail. Figure 4 illustrates the application of dilated convolution instead of standard convolution in the ResNet framework. This study attempted to decrease the output stride rate using dilated convolution instead of standard convolution. Thus, lower output stride rate OS = 16 was obtained. With a low OS rate, it is possible to capture more intense information about small lesions. Thus, larger feature maps created in the encoder part can make it possible to obtain effective segmentation maps.

3.2.3 Atrous spatial pyramid pooling

Spatial pyramid pooling (SPP) is a breakthrough tool to remove the fixed-size constraint of the deep networks. It is suitable for maintaining spatial information effectively. The ASPP technique was developed with the
motivation from the success of the SPP and dilated (atrous) convolution operations. The ASPP is utilized for resampling feature maps created from the encoder architecture at different atrous rates. It can capture multi-scale patterns and contextual information. The module consists of four parallel atrous filters with different atrous rates. These atrous convolutions are applied to encoder feature maps and their outputs are concatenated. The original DeepLabV3+ proposes to use atrous rates of 6, 12, and 18 in the ASPP module. However, the small and variable size of the COVID-19 lesion patterns compared to the whole image may impede the effectiveness of the original atrous rates. Thus, lower atrous rates may be more suitable for the robust segmentation of COVID-19 patterns in CT images. Modifying the atrous rates to 4, 8, and 12 have provided a better performance on lesion segmentation. It is aimed to obtain enough spatial details of the region of interest. In addition, the ASPP technique will not prevent the network from working effectively in real-time applications because it does not change the parameter size of the feature map. The details of the modified ASPP module are shown in Figure 5.

3.2.4 | Decoder module

The significant contribution of the DeepLabV3+ model over its predecessors is apparent in this module. DeepLabv3+ provides a simple yet useful decoder module to refine the segmentation results. In this way, the detailed region of interest boundaries can be recovered in a faster and stronger. In the proposed decoder, the features from the ASPP module are first upsampled bilinearly by a factor of 4. Subsequently, the upsampled features are concatenated with low-level features from the backbone architecture. Low-level features are subjected to $1 \times 1$ convolution before the concatenation for an effective training process. Combining low-level features having rich spatial information with high-level features improves segmentation accuracy. Then $3 \times 3$ convolution filter is applied followed by upsampling by a factor of 4 to create the final segmentation prediction as shown in Figure 2.

3.3 | Hyperparameter tuning

The deep networks used as backbone architecture commonly have millions of parameters to capture meaningful information from high-resolution input images. Effective segmentation can be possible by training the parameters in question on a huge amount of annotated images. However, most of the open access COVID-19 imaging data sets consist of fewer samples than other benchmarks image datasets. Thus, pre-trained networks on large image datasets are the robust alternative tool when there are insufficient samples to train a deep model from scratch.

In this study, the encoder part of the proposed framework employs a pre-trained residual network as a feature extractor. The backbone network produces rich information for pixel-wise segmentation by fine-tuning the pre-trained optimization parameters. A batch normalization and ReLU activation are applied after each convolutional layer, respectively. In particular, the ReLU is a useful activation function to provide a better contrast compared to the output of the convolution layer by removing the negative part of the input. For the optimization, the ADAM optimizer is used in the training network since it
has a fast convergence rate. The initial learning rate was set as 0.001, and then decreased by a value of 0.4 after every 4 epochs. Segmentation task was performed with batch size of 4 for 10 epochs. As the early stopping method was kept active to minimize the risk of overfitting, all training folds did not result in the entire 10 epochs and instead were early stopped.

### 3.4 Implementation and evaluation

In order to quantitatively evaluate the performance of the proposed model, the COVID-19 image dataset is divided into three subsets for training, validation, and testing, with proportions of 0.60, 0.20, and 0.20, respectively. Then, a fivefold cross-validation technique was carried out in partitioning the dataset to evaluate the generalization performance of the proposed model. The segmentation performance can only be reliably evaluated with suitable metrics. The SS process for medical tasks is usually between classes having imbalanced pixel numbers. Thus, the science community for medical image processing has usually focused on F-score-based metrics such as the Dice similarity coefficient (DSC) and intersection over union (IoU). The DSC and IoU metrics indicate the segmentation performance based on the overlap between the prediction of the proposed model and ground-truth maps. F-score-based measures range from 0 to 1, and the metric value is 1 indicates perfect overlap. DSC is the ratio of twice the pixels of the accurately associated region of interest to the sum of the total number of associated pixels, and the total number of ground-truth pixels. This expression clearly indicates \( \text{DSC} \geq \text{IoU} \). In addition to DSC and IoU, the sensitivity and specificity metrics are also robust metrics to improve the assessment of segmentation performance. Sensitivity represents the ratio of the number of correctly detected other class pixels to the total number of outside the region of interest pixels. The accuracy metric commonly used in classification tasks is not suitable for medical image segmentation because of the true-negative effect. The mathematical expressions of four widely popular evaluation metrics are defined as follows:

\[
\text{DSC} = \frac{2 \times TP}{2 \times TP + FP + FN} = \frac{2 |G \cap S|}{|G| + |S|},
\]

\[
\text{IoU} = \frac{TP}{TP + FP + FN} = \frac{|G \cap S|}{|G| \cup |S|},
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN},
\]

\[
\text{Specificity} = \frac{TN}{TN + FP}.
\]

Evaluation metrics are based on the confusion matrix for binary classification, where \( TP, TN, FP, FN, G, \) and \( S \) represent true positive, true negative, false positive, false negative, ground truth and segmentation, respectively.

### 4 RESULTS

The proposed modified DeepLabV3+ approach on segmentation of axial CT images was performed in the Matlab (The MathWorks, Natick, MA, USA) environment. The programming platform was installed on a 2.7 GHz dual Intel i7 processor with 16 GB RAM, NVIDIA GeForce ROG-STRIX 256 bit, and 8GB GPU hardware.

In order to better analyze the contribution of the proposed framework to the segmentation performance, the modified model was compared to the original
DeepLabV3+ architecture. The segmentation scenario is the same for both models and provides a more reliable evaluation. The comparison is based on a multi-class segmentation task as lung (non-infected tissue), COVID (consisting of GGO, consolidation, and pleural effusion irregularities), and background. Table 2 shows the segmentation results for the original residual network-based DeepLabV3+ architecture with the proposed modified model.

After training, the segmentation results proved that the lungs and COVID-19 infected region could be distinguished with satisfactory performance. The original DeepLabV3+ based on residual network presented a DSC of 0.738 and an IoU of 0.585 for COVID-19 infection segmentation. Table 2 demonstrates that the modified model provided higher accuracy and efficiency. The lower atrous rates in the ASPP module significantly improved segmentation performance. Modified model based on lower atrous rates in the ASPP module achieved a DSC and IoU of 0.901 and 0.820 for lungs, as well as 0.754 and 0.606 for COVID-19 infection. The segmentation results for the COVID class were improved in terms of DSC and IoU by around 2%. Both networks showed near-perfect segmentation performance for the background class due to the characteristic color, spatial texture, and geometric shape. The proposed model and the original DeepLabV3+ framework achieved the same DSC performance of 0.987 for background segmentation. Furthermore, the proposed model obtained sensitivity and specificity of 0.828 and 0.972 on the overall testing set for COVID-19 lesions, respectively. In particular, the sensitivity metric is a remarkable indicator in evaluating the robustness of segmentation models. Due to the small number of images in the dataset, segmentation performance was evaluated for all models using a fivefold cross-validation technique. Training, validation, and test data are reassigned for each fold, paving the way for more reliable and consistent evaluation.

After fivefold cross-validation, segmentation results for lung and COVID-19 lesions reveal the highest variance for sensitivity and IoU metrics. For the modified framework, the sensitivity and IoU for the fivefold experiment of the lung segmentation presented standard deviations of 0.021 and 0.026, respectively. For COVID-19 lesion segmentation, the standard deviation values were obtained as 0.024 and 0.021 for sensitivity and IoU, respectively. In particular, the fact that COVID-19 lesions have low pixel counts and exhibit spatial, textural, and morphological differences causes high variance. More details on fivefold experiments are listed in Table 3.

In order to explore the robustness of the proposed method to noise in CT images, artificially lower quality images were generated. In this context, a matrix of random values drawn from a Gaussian distribution was injected into each high-quality CT image. Table 4 shows the segmentation results for noisy CT images of the original residual network-based DeepLabV3+ architectures with the proposed modified model. It has been observed that the overall segmentation performances for both architectures decrease for noisy CT images. The original DeepLabV3+ based on the residual network offered a DSC of 0.702 and an IoU of 0.542 for COVID-19 infection segmentation. The modified model based on lower atrous rates in the ASPP module achieved a DSC and IoU of 0.879 and 0.786 for lungs, as well as 0.714 and 0.556 for COVID-19 infection. The results showed that the proposed model has superior and still satisfactory performance in CT images with low-quality resolution.

Table 2: Comparison of the proposed DeepLabV3+ and the original DeepLabv3+ model in terms of DSC, IoU, the sensitivity (Sen.), and the specificity (Spec.) metrics

| Network | Class   | DSC  | IoU  | Sen.  | Spec.  |
|---------|---------|------|------|-------|--------|
| Original DeepLabV3+ with ResNet-50 | “Background” | 0.987 | 0.975 | 0.982 | 0.983 |
| | “COVID” | 0.738 | 0.585 | 0.834 | 0.967 |
| | “Lung” | 0.889 | 0.800 | 0.865 | 0.979 |
| Modified DeepLabV3+ with ResNet-50 | “Background” | 0.987 | 0.975 | 0.982 | 0.984 |
| | “COVID” | 0.754 | 0.606 | 0.828 | 0.972 |
| | “Lung” | 0.901 | 0.820 | 0.890 | 0.979 |
4.1 Comparison to the state-of-the-art models

In order to better observe the effectiveness of the proposed modified DeepLabV3+ model in segmenting the COVID-19 lesion, the previously proposed state-of-the-art models commonly used in the literature were applied to the same dataset. Specifically, the comparison for the proposed method was performed with U-Net and SegNet. Note that all methods used the same segmentation procedure. Furthermore, the ResNet-18 based original DeepLabV3+ framework was also run to evaluate the segmentation efficiency of the backbone architecture.

The proposed DeepLabV3+ framework with lower atrous rates was compared to other state-of-the-art models in terms of various performance metrics and the number of parameters. The trade-off between segmentation performance and the number of parameters summarized in Table 5 is an essential indicator for evaluating the robust end-to-end framework in the segmentation of COVID-19 lesions.

Table 4 shows that the original DeepLabV3+ with dilated ResNet-50 with ASPP rates of 6, 12, and 18 achieves 0.871 of DSC for overall segmentation. The experimental results show that reducing the atrous rate to 4, 8, and 12 in the ASPP module provides a more effective segmentation and achieves an overall DSC of 0.881. Moreover, the modified framework presents better performance than other state-of-the-art models. U-Net and SegNet models achieve DSC scores of 0.525, and 0.816, respectively. The original DeepLabv3+ model, which applies the backbone architecture as a dilated ResNet-18, achieves a DSC performance of 0.858. As can be seen, the highest segmentation accuracy is achieved with the...
proposed model. However, the ResNet-18 based architecture has competitive segmentation performance with almost half the parameter count of the ResNet-50 based architecture. That is why DeepLabV3+ has options that enable lightweight architectures for real-time applications. Figure 7 compares all segmentation networks in terms of DSC performance, training time, and the number of parameters graphically.

In order to contribute to all these quantitative evaluations, annotated ground-truth maps and the segmentation results of each framework are visualized in Figure 8. While the first two columns of Figure 8 show raw CT images and ground-truth maps, respectively, columns 3–7 show the segmentation prediction for each network. As seen in Figure 8, the segmentation performance of the proposed model is more successful. From the general observation, the modified DeepLabV3+ defines the boundaries of the regions of interest at a reasonable level. Although correctly predicting the small but crucial region of interest is the main challenge, minor boundary variations in the segmentation map do not affect overall visual performance. Thus, it is clear that the produced prediction as the false-negative pixel is relatively on a low level, but it significantly affects the performance metrics score. The visual results show that building a robust segmentation model is achievable by applying various modifications in the network pipeline.

5 | DISCUSSION

In medical imaging, robust SS enables a better understanding of infection morphology and rapid tracking of lesion boundaries. Segmentation of the COVID-19 lesion signs is a difficult task because requires rich semantic features. The spatial location, size, and morphologic differences of COVID-19 lesions such as GGO, consolidation, pleural effusion, crazy paving, reverse halo make them tough to detect on high-resolution images. Furthermore, the lesion regions in question are so small compared to the rest of the image that even a few incorrect pixels assigned can have an enormous impact on segmentation performance. Due to a high imbalance in the distribution between classes, it is crucial to focus on the most suitable metrics to evaluate the performance improvements of the proposed models. Therefore, high specificity and accuracy performances can be misleading. In this paper, DSC and IoU metrics are prioritized in the performance evaluation of each segmentation network. Visual segmentation findings reveal that DSC and IoU metrics are more reliable indicators, in contrast to high specificity.

Many segmentation architectures are suitable to make various modifications for improving segmentation performance and hardware feasibility. Especially the DeepLabV3+ pipeline offers many options in this regard.
For hardware implementation, a lightweight model can be built with a suitable backbone architecture in the encoder. Also, the backbone architectures utilized in the encoder directly affect segmentation accuracy. Additional changes to parameters such as number of convolutional filter, atrous rate, pooling size, and upsampling factor in modules other than the encoder can improve segmentation performance. In this study, various dilated ResNet were applied to the original DeepLabV3+ model as the encoder. The original DeepLabV3+ with ASPP rates 6, 12, and 18 with dilated ResNet-50 achieved 0.738 DSC for segmentation of COVID-19 lesions. The modified framework based on decreasing ASPP rates to 4, 8, and 12 proves more effective in segmenting COVID-19 infections because it achieved a DSC of 0.754. The improvement of DSC performance without increasing the number of parameters in the network is a significant advancement.

Due to the rapid spread of the coronavirus, many scientists urgently tried to develop automatic systems to support physicians in the fight against the epidemic. However, for robust training of the proposed deep learning-based models, the diagnosed and annotated lung images were limited. Previous works have used data augmentation approaches or combined different medical imaging datasets to overcome this drawback. Although the data augmentation approach artificially increases the number of training images, the validation performance achieved by the model, in particular, departs from the real scenario. Also, efficient normalization methods based on resizing images to use different datasets in common can lead to loss or distortions in semantic information. The modified DeepLabV3+ based on a pre-trained encoder has enabled satisfactory performance without interfering with the number and size of images in the original dataset.

Previous works have proposed various effective frameworks that can successfully segment COVID-19 pulmonary lesions from CT images to assist physicians in rapid diagnosis and treatment. Most of these segmentation works used popular architectures such as U-Net and SegNet, or various modifications on these architectures. For further evaluation, the proposed modified framework was compared with other previous COVID-19 segmentation work. Table 6 outlines information and additional details of the comparison process.
Table 6 shows that the modified DeepLabV3+ performs similarly or better in segmenting COVID-19 pulmonary infections than previous works using appropriate backbone architecture and ASPP rates. For a proposed network, apart from successful segmentation results, the number of parameters, computing time cost, and network complexity are also important issues to be considered. These issues make a segmentation model infeasible for real-world applications. Various backbone and pipeline parameter options can make the DeepLabV3+ framework more effective than other architectures. The success achieved in this work with a minor modification reinforces this idea.

5.1 Limitations

Although deep learning-based frameworks provide robust segmentation results, most of these models are not suitable for clinical use. The number of parameters in the modified DeepLabV3+ still needs to be drastically reduced. This is one of the main issues that need to be discussed when evaluating the effectiveness of the modified architecture. Another problem that needs to be discussed is that the proposed model was trained only with images related to COVID-19. Therefore, the robustness of the proposed model is unclear in distinguishing between COVID-19 lesions and other types of pneumonia or completely different medical conditions such as cancer. The comprehensive evaluation of the segmentation model is only possible if high-quality annotated data sets on various medical conditions are available.

6 Conclusion

Nowadays, the development of autonomous systems for understanding medical images in the fight against COVID-19 is one of the essential research topics. In particular, the focus is on deep learning-based models for rapid detection and quantitative assessment of COVID-19 abnormalities from lung CT images. This paper has proposed various modifications such as backbone architecture and ASPP rates within the framework of DeepLabV3+ segmentation. In addition, a comprehensive comparison with other state-of-the-art segmentation methods was conducted to observe the effectiveness of the proposed model. The modified DeepLabV3+ achieves good segmentation performance using around 43.9 M parameters. After fivefold cross-validation, the framework achieves an overall DSC of 0.881 for whole CT segmentation. The results prove that the DeepLabV3+ framework has great potential for improving segmentation accuracy and hardware applicability.

As future research, it is planned to build the most accurate and lightweight model by integrating various optimization algorithms into the DeepLabV3+ segmentation pipeline. Furthermore, the contributions of different backbone architectures in the encoder part to the segmentation performance will be investigated.
CONFLICT OF INTEREST
The author declares no conflicts of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in COVID-19—Medical segmentation at http://medicalsegmentation.com/covid19/, reference number 44.

ORCID
Hasan Polat https://orcid.org/0000-0001-5535-4832

REFERENCES
1. Tiwari S, Jain A. Convolutional capsule network for COVID-19 detection using radiography images. Int J Imaging Syst Technol. 2021;31(2):525-539. doi:10.1002/ima.22566
2. Selvaraj D, Venkatesan A, Mahesh VGV, Joseph Raj AN. An integrated feature frame work for automated segmentation of COVID-19 infection from lung CT images. Int J Imaging Syst Technol. 2021;31(1):28-46. doi:10.1002/ima.22525
3. Loey M, Manogaran G, Taha MHN, Khalifa NEM. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Meas J Int Meas Confed. 2021;167:108288. doi:10.1016/j.measurement.2020.108288
4. Coronavirus Disease (COVID-19). Accessed January 29, 2022. https://www.who.int/emergencies/diseases/novel-coronavirus-2019
5. Ucar F, Korkmaz D. COVIDiagnosis-net: deep BayesSqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med Hypotheses. 2020;140-(April):109761. doi:10.1016/j.mehy.2020.109761
6. Ai T, Yang Z, Hou H, 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(2):E32-E40. doi:10.1148/radiol.2020200642
7. Fang Y, Pang P. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology. 2020;296:15-17.
8. Alalwan N, Abozeid A, ELHabshy AAA, Alzahrani A. Efficient 3D deep learning model for medical image semantic segmentation. Alex Eng J. 2021;60(1):1231-1239. doi:10.1016/j.ajegy.2020.10.046
9. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42-(December):60-88. doi:10.1016/j.media.2017.07.005
10. Yan Q, Wang B, Gong D et al. COVID-19 chest CT image segmentation—a deep convolutional neural network solution; arXiv preprint; 20004:1-109897. http://arxiv.org/abs/2004.10987
11. Bernheim A. Chest CT findings in COVID-19. Radiology. 2020;19(3):1-19.
12. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, et al. Clinical, laboratory and imaging features of COVID-19: a systematic review and meta-analysis. Travel Med Infect Dis. 2020;34(March):101623. doi:10.1016/j.tmaid.2020.101623
13. Li M. Chest CT features and their role in COVID-19. Radiol Infect Dis. 2020;7(2):51-54. doi:10.1016/j.rird.2020.04.001
14. Khalifa NEM, Manogaran G, Taha MHN, Loey M. A deep learning semantic segmentation architecture for COVID-19 lesions discovery in limited chest CT datasets. Expert Syst. 2021;(March):1-11. doi:10.1111/exsy.12742
15. Li Y, Zhang X, Li H, Zhou Q, Cao X, Xiao Z. Object detection and tracking under complex environment using deep learning-based LPF. IET Comput Vis. 2019;13(2):157-164. doi:10.1049/iet-cvi.2018.5129
16. Parmar Y, Natarajan S, Sobha G. Deeprange: deep-learning-based object detection and ranging in autonomous driving. IET Intell Transp Syst. 2019;13(8):1256-1264. doi:10.1049/iet-its.2018.5144
17. Ogawa A, Hori T. Error detection and accuracy estimation in automatic speech recognition using deep bidirectional recurrent neural networks. Speech Commun. 2017;89:70-83. doi:10.1016/j.specom.2017.02.009
18. Ballesteros DM, Rodriguez-Ortega Y, Renza D, Arce G. Deep4SNet: deep learning for fake speech classification. Expert Syst Appl. 2021;184(November):115465. doi:10.1016/j.eswa.2021.115465
19. Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation. Comput Biol Med. 2020;126(October):104037. doi:10.1016/j.compbiomed.2020.104037
20. Benameur N, Mahmoudi R, Zaid S, Arous Y, Hmida B, Bedoui MH. SARS-CoV-2 diagnosis using medical imaging techniques and artificial intelligence: a review. Clin Imaging. 2021;76(July):6-14. doi:10.1016/j.clinimag.2021.01.019
21. Li B, Shi Y, Qi Z, Chen Z. A survey on semantic segmentation. IEEE International Conference on Data Mining Workshops (ICDMW). Vol 2018- November; IEEE; 2019:1233-1240. doi:10.1109/ICDMW.2018.00176
22. Baheti B, Gajre S, Talbar S. Semantic scene understanding in unstructured environment with deep convolutional neural network. IEEE Region 10 Annual International Conference, Proceedings/TENCON. Vol 2019- October; IEEE; 2019:790-795. doi:10.1109/TENCON.2019.871339
23. Ozturk O, Saritürk B, Seker DZ. Comparison of fully convolutional networks (FCN) and U-net for road segmentation from high resolution imageries. Int J Environ Geoinf. 2020;7(3):272-279. doi:10.30897/ijegeo.737993
24. Minaei S, Kafreh R, Sonka M, Yazdani S, Jamalipour Soufi G. Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. Med Image Anal. 2020;65:101794. doi:10.1016/j.media.2020.101794
25. Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. BMC Med Imaging. 2021;21(1):1-11. doi:10.1186/s12880-020-00529-5
26. Müller D, Soto-Rey I, Kramer F. Robust chest CT image segmentation of COVID-19 lung infection based on limited data. Inform Med Unlocked. 2021;25(ly):100681. doi:10.1016/j.imu.2021.100681
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol 2016-December; IEEE; 2016:770-778. doi:10.1109/CVPR.2016.90
28. Yao X, Wang X, Wang SH, Zhang YD. A comprehensive survey on convolutional neural network in medical image analysis. Multimed Tools Appl. 2020. doi:10.1007/s11042-020-09634-7
29. Wang G. A perspective on deep imaging. IEEE Access. 2016;4:8914-8924. doi:10.1109/ACCESS.2016.2624938
30. Gavathri JL, Abraham B, Sujarani MS, Nair MS. A computer-aided diagnosis system for the classification of COVID-19 and non-COVID-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. Comput Biol Med. 2022;141(June):105134. doi:10.1016/j.compbiomed.2021.105134
31. Baghdadi NA, Malki A, Abdelaliem SF, Magdy Balaha H, Badawy M, Elhosseini M. An automated diagnosis and classification of COVID-19 from chest CT images using a transfer learning-based convolutional neural network. Comput Biol Med. 2022;144(January):105383. doi:10.1016/j.compbiomed.2022.105383
32. Budak Ü, Çebuk M, Çömert Z, Şengür A. Efficient COVID-19 segmentation from CT slices exploiting semantic segmentation with integrated attention mechanism. J Digit Imaging. 2021;34(2):263-272. doi:10.1007/s10278-021-00434-5
33. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Lect Notes Comput Sci. 2015;9351:234-241. doi:10.1007/978-3-319-24574-4_28
34. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 2017;39(12):2481-2495.
35. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell. 2017;39(4):640-651. doi:10.1109/TPAMI.2016.2572683
36. Zheng R, Zheng Y, Dong-Ye C. Improved 3D U-Net for COVID-19 chest CT image segmentation. Sci Program. 2021;2021:1-9. doi:10.1155/2021/9999368
37. Voulodimos A, Protopapadakis E, Katsamanis I, Doulamis A, Doulamis N. Deep learning models for COVID-19 infected area segmentation in CT images. ACM International Conference Proceeding Series; Association for Computing Machinery; 2021:404-411. doi:10.1145/3453892.3461322
38. Wang B, Jin S, Yan Q, et al. AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system. Appl Soft Comput. 2021;98:1-22. doi:10.1016/j.ascom.2020.106897
39. Wang Y, Zhang Y, Liu Y, et al. Does non-COVID-19 lung lesion help? Investigating transferability in COVID-19 CT image segmentation. Comput Methods Programs Biomed. 2021;202:106004. doi:10.1016/j.cmpb.2021.106004
40. Karthik R, Menaka R, Hariharan M, Won D. Contour-enhanced attention CNN for CT-based COVID-19 segmentation. Pattern Recognit. 2022;125:108538. doi:10.1016/j.patcog.2022.108538
41. Menteşe E, Hançer E. Nucleus segmentation with deep learning approaches on histopathology images. Eur J Sci Technol. 2020;(November):95-102. doi:10.31590/ejosat.819409
42. Khan Z, Yahya N, Alsaiah K, Ali SSA, Meriaudeau F. Evaluation of deep neural networks for semantic segmentation of prostate in T2W MRI. Sensors (Switzerland). 2020;20(11):1-17. doi:10.3390/s20113183
43. Baheti B, Innani S, Gajre S, Talbar S. Semantic scene segmentation in unstructured environment with modified DeepLabV3+. Pattern Recognit Lett. 2020;138:223-229. doi:10.1016/j.patrec.2020.07.029
44. COVID-19—Medical Segmentation. Accessed October 10, 2021. http://medicalsegmentation.com/covid19/
45. Wang P, Chen P, Yuan Y et al. Understanding convolution for semantic segmentation. IEEE Winter Conference on Applications of Computer Vision (WACV). Vol 2018-January; IEEE: 2018:1451-1460. doi:10.1109/WACV.2018.001163
46. Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. Sustain. 2021;13(3):1-29. doi:10.3390/su13031224
47. Yao XJ, Zhu ZQ, Wang SH, Zhang YD. Cgbbnet: an explainable deep learning framework for Covid-19 detection. Diagnostics. 2021;11(9):1712. doi:10.3390/diagnostics11091712
48. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. Lect Notes Comput Sci. 2018;11211:833-851. doi:10.1007/978-3-030-01234-2_49
49. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. Lect Notes Comput Sci. 2014;8691(Pt 3):346-361. doi:10.1007/978-3-319-10578-9_23
50. Yuan X, Shi J, Gu L. A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst Appl. 2021;169(June):114417. doi:10.1016/j.eswa.2020.114417
51. Ma J, Wang Y, An X, et al. Toward data-efficient learning: a benchmark for COVID-19 CT lung and infection segmentation. Med Phys. 2021;48(3):1197-1210. doi:10.1002/mp.14676
52. Kingma DP, Ba JL. Adam: a method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015; arXiv preprint 2014:1-15:arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980
53. Asgari Taghanaki S, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G. Deep semantic segmentation of natural and medical images: a review. Artif Intell Rev. 2021;54(1):137-178. doi:10.1007/s10462-020-09854-1
54. Jun M, Cheng G, Yixin W, et al. COVID-19 CT lung and infection segmentation dataset. 2020. doi:10.5281/ZENODO.3757476
55. Pei HY, Yang D, Liu GR, Lu T. MPS-net: multi-point super-resolution for medical images: a review. Sensors (Switzerland). 2021;21(9):105383. doi:10.3390/s2109105383
56. Research and Practical Clinical Center of Diagnostics and Telemedicine Technologies. Department of Health Care of Moscow; 2020. http://www.mosmed.ai

**How to cite this article:** Polat H. A modified DeepLabV3+ based semantic segmentation of chest computed tomography images for COVID-19 lung infections. *Int J Imaging Syst Technol.* 2022; 32(5):1481-1495. doi: 10.1002/ima.22772