



COVID-SegNet: encoder–decoder-based architecture for COVID-19 lesion segmentation in chest X-ray

Tarun Agrawal¹ · Prakash Choudhary²

Received: 9 May 2022 / Accepted: 10 April 2023 / Published online: 19 April 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The coronavirus disease 2019, initially named 2019-nCoV (COVID-19) has been declared a global pandemic by the World Health Organization in March 2020. Because of the growing number of COVID patients, the world's health infrastructure has collapsed, and computer-aided diagnosis has become a necessity. Most of the models proposed for the COVID-19 detection in chest X-rays do image-level analysis. These models do not identify the infected region in the images for an accurate and precise diagnosis. The lesion segmentation will help the medical experts to identify the infected region in the lungs. Therefore, in this paper, a UNet-based encoder–decoder architecture is proposed for the COVID-19 lesion segmentation in chest X-rays. To improve performance, the proposed model employs an attention mechanism and a convolution-based atrous spatial pyramid pooling module. The proposed model obtained 0.8325 and 0.7132 values of the dice similarity coefficient and jaccard index, respectively, and outperformed the state-of-the-art UNet model. An ablation study has been performed to highlight the contribution of the attention mechanism and small dilation rates in the atrous spatial pyramid pooling module.

Keywords Deep learning · ASPP module · Channel and spatial attention · Chest X-ray · COVID-19 lesion segmentation

1 Introduction

Coronavirus disease 2019, initially named as 2019-nCoV (COVID-19) is the deadliest disease that was first detected in China. It has the greatest impact on the elderly, and it can result in hospitalization, intubation, critical care, and even death [11]. Some common symptoms of COVID-19 are fever, dry cough, sore throat, shortness of breath and pneumonia. With comorbidities, such as heart, respiratory diseases, and diabetes, the mortality rate is significant [2]. COVID-19 detection tests include Real-Time Reverse Transcription-Polymerase Chain Reaction (RT-PCR) and Rapid Antibody Test (RAT). Blood samples are analyzed in RAT to see if antibodies are present. Although, it is not a direct

approach to diagnosing, but it shows the immune system's response. Antibodies are produced by the immune system to fight the infection. However, antibodies can take 9–20 days to appear, therefore, this approach of testing COVID-19 is ineffective. In RT-PCR swab of the patient's throat is obtained, and ribonucleic acid (RNA) is extracted. The patient is positive for COVID-19 if it has the same genetic sequence as SARS-CoV-2. It is the gold standard for COVID-19 detection, although it is expensive and yields a significant false-positive rate [16]. Additionally, it also does not provide useful information about the current status of the patient's lung.

CT-Scan (CTS) is advised for suspected COVID-19 patients at an earlier stage of the disease since it is a reliable imaging technique for early diagnosis [43]. It can capture observable imaging characteristics, such as lung fibrosis, consolidation, and ground-glass opacities (GGO) [4]. On the other hand, CTS has some limitations, such as being slow in acquisition and expensive. In comparison to CTS, chest X-ray (CXR) is easy to capture and also affordable. CXRs are the common first-line imaging tool used all around

✉ Prakash Choudhary
prakash.choudhary@curaj.ac.in

Tarun Agrawal
tarun20octagarwal@gmail.com; tarunagarwal@nith.ac.in

¹ Department of Computer Science and Engineering,
National Institute of Technology Hamirpur, Hamirpur,
Himachal Pradesh 177005, India

² Department of Computer Science and Engineering, Central
University of Rajasthan, Ajmer, Rajasthan, India

the world and also have less radiation exposure than CTSs.¹ Presently, CXRs and CTSs are widely used as an assisting imaging technique for COVID-19 prognosis, and they are said to have diagnostic potential in recent studies [11, 47].

Deep learning has emerged in the last decade, and it has proven to be superior in computer vision and pattern recognition [28]. Deep learning models are cutting-edge image classification algorithms that gained traction after successfully categorizing 1,000 classes in the ImageNet dataset [12]. CNN models are widely used in image classification and segmentation applications due to their superior expression ability and data-driven adaptive feature extraction technique [20, 41, 48]. Deep learning models are commonly used in medical image analysis [31]. Object segmentation, detection, and classification are a few popular tasks on which the deep learning community has worked extensively.

CTSs and CXRs have been significantly used for COVID-19 diagnosis in the last two years. Similarly, the majority of studies have utilized these two modalities for their deep learning models. The COVID-19 diagnostic studies can be categorized into image-level classification [2, 3, 35] and lesion segmentation [27]. Hassan et al. [19] have given a detailed review of studies done for image-level classification in CXR images. For CTSs, there are numerous research studies for both image-level classification [3] and lesion segmentation [6, 27]. While most of the studies with CXRs are focused on image-level classification [2, 35]. COVID-19 lesion identification in CXR is also important for the diagnosis. The segmented lesions can aid in determining the severity of pneumonia and ensuring that patients receive subsequent follow-up. With the help of a human-machine collaborative approach to obtain the exact affected locations, the first ground-truth masks for the COVID-19 infection in CXRs are introduced [11]. A similar approach is also used by Shan et al. [46] to segment the COVID-19 lesion in CTSs. These two studies have shown that interactions between radiologists and deep learning models can significantly reduce the annotation time [33].

Medical experts can use COVID-19 lesion segmentation to assess the current state of the illness and its severity to plan future treatment. The area covered by the lesion shows the location where pulmonary alveolar are abnormal [30]. Despite the great detection accuracy of COVID-19 in CXRs, the time required for interpretation by medical experts can increase the workload. However, deep learning models can be deployed for automatically segmenting the lesion in CXR. It has already been used successfully for lung segmentation in X-rays [7, 49] and CTSs [1, 25]. COVID-19 lesion segmentation in CXR is a challenging task because

the infection region in the X-ray shows features of different sizes and locations. Further, the COVID-19 does not have any definitive contour, which makes the segmentation of the infected region even more difficult.

UNet [41] is the most popular encoder–decoder architecture for medical image segmentation. The low- and high-level features of images are captured by the encoder, and those semantic features are concatenated in the decoder to generate the final segmented output. The COVID-19 lesion segmentation can localize the infection area, which can help the medical practitioners to analyze the severity of the COVID-19 infection in the lungs. However, manual segmentation requires a expert radiologist and is also a time-consuming process. In this paper, we have proposed the UNet-based novel encoder–decoder architecture for automatic COVID-19 lesion segmentation in CXRs. The lesion segmentation will help the radiologist to localize the infection in the lungs, and the obtained results show the proposed model can localize the lesion accurately in CXRs. The attention mechanism [42] is used to exploit the recalibration of features in both channel-wise and spatially. Further, the atrous spatial pyramid pooling (ASPP) [8] module is used to improve the model performance with the large view field. We have performed the ablation study to highlight the role of the attention mechanism and the ASPP module in improving the performance of the model. To the best of our knowledge, no study has proposed a similar architecture for the COVID-19 lesion segmentation in CXR images. The performance of the proposed model has been evaluated on both the dice similarity coefficient and jaccard index.

The rest of the paper is structured as: In Sect. 2, related work is discussed. Section 3 describes the dataset and the methodology. Section 4 discusses the evaluation metrics, implementation details and the obtained results along with ablation study. In Sect. 5, we have done the result analysis with UNet and related works followed by a conclusion in Sect. 6.

2 Related work

Researchers have used CXRs and CTSs with deep learning models for the diagnosis of COVID-19. Segmenting the infected area and image-level analysis in CXR and CTS are two ways to diagnose the disease. Most of the research work has done image-level analysis, such as Ozturk et al. [35], Agrawal and Choudhary [2], Ahuja et al. [3], and many others. There are very few works that have done the segmentation of the infected region for the accurate and precise diagnosis on the CTS [6, 27] and CXR images [11, 29]. Our primary focus in the proposed COVID-SegNet model is lesion segmentation. However, in addition, we have

¹ <https://www.health.harvard.edu/cancer/radiation-risk-from-medical-imaging>.

discussed a few state-of-the-art works using both image-level and lesion segmentation techniques.

Ozturk et al. [35] developed the DarkCovidNet CNN model for the automatic detection of COVID-19 in CXR images. Instead of scratch CNN model, the authors adopted the DarkNet-19 [40] model design as a starting point. The model was trained and tested on 1127 chest images including 127 COVID-19 images. Agrawal and Choudhary [2] propose a deep learning approach to increase the accuracy of binary and three-class classification of COVID-19 in CXRs. Many research studies are also reported on CTSs in addition to CXRs. Ahuja et al. [3] proposed the COVID-19 detection model using the transfer learning approach in CT scan images. For the experimental evaluation, this work used well-known pre-trained architectures, such as ResNet-101, and SqueezeNet. Ragab et al. [38] proposed the use of capsule neural-based CapsNet model above traditional CNN because of its better generalizations and fewer parameters. It achieved an accuracy of 94% for the COVID-19 detection in CXR images. It has also outperformed the traditional CNN model. However, there are few research studies that have used unsupervised deep learning (UDL) for the COVID-19 detection in CXR images. Mansour et al. [32] used the UDL variational auto-encoder model for the COVID-19 diagnosis. Elzeki et al. [15] has proposed the CNN with less number of parameters in comparison to the pre-trained network on three datasets. There are different techniques available for COVID-19 detection, such as metaheuristic optimization [44] and web crawling, to update the datasets [14]. In the above-mentioned works, these models do the image-level classification for COVID-19 detection.

In the following, the deep learning models developed to segment the COVID-19 lesion segmentation in CTS and CXR images are discussed. Laradji et al. [27] introduced a framework that trains using a consistency-based loss function on a CTS dataset labeled with point-level supervision. The proposed model used the ImageNet pre-trained VGG16 architecture [48]. Meanwhile, Li et al. [29] has proposed the LViT model for COVID-19 infection segmentation using the Vision Transformer model [13]. It used the pixel-attention blocks to preserve the local features. The proposed model is trained on COVID-19 CT scans with ground truths lesions. The ground truth lesions were developed by physicians from several hospitals.

These research works have used the UNet [41] for the segmentation task. Laradji et al. [27] and Degerli et al. [11] have used the transfer learning technique in the encoder of the UNet. The pre-trained models used in the transfer learning techniques and UNet model have a large number of parameters, which makes them prone to overfitting on small medical datasets. While, in the proposed work, the COVID-SegNet model has been trained from scratch for COVID-19 lesion segmentation in CXR images. The

proposed COVID-SegNet model is lightweight in terms of parameters in comparison to other models.

3 Dataset and methodology

In this section, we have mentioned the dataset used and discussed the proposed architecture for COVID-19 lesion segmentation. Further also, we have discussed the attention mechanism and atrous convolution-based ASPP module before elaborating on the proposed architecture.

3.1 Dataset

In this study, we have used the QaTa-COV19 dataset [11] for COVID-19 infection segmentation. In this section, we have described the human–machine collaborative approach deployed by Degerli et al. [11] to obtain the ground-truth COVID-19 segmentation masks.

The development of various deep learning algorithms for image segmentation and classification has been aided by recent advancements in hardware. However, supervised deep learning methods require a substantial amount of annotated data. On the other hand, using the small dataset for training the deep learning model can lead to overfitting. A manual ground-truth mask by the radiologist for COVID-19 infection segmentation is labor-intensive work. As a result, Degerli et al. [11] suggested a human–machine collaborative technique for COVID-19 infection segmentation. The goal of the research was to reduce human labor and provide better segmentation masks. They suggested an iterative technique with two stages. In the first stage, MDs manually segment COVID-19 infected regions in the 500 CXRs. Then, the initial ground-truth masks were used to train multiple UNet-based models. Further, the MDs evaluated the segmented masks predicted by the network for the test set, as well as the original CXRs and the initial masks. The best segmentation masks are chosen from them.

On the collected lesion segmented masks from the first stage, five deep learning models based on UNet++ [52] and DLA [50] network are trained in the second stage. The remaining dataset is then utilized to estimate the segmentation region using these models. Then, the MDs choose the best masks from the prediction results, and if any erroneous mask was found, then it was manually segmented. A total of 2,951 CXRs (224×224) with COVID-19-infected regions were segmented using this human–machine collaboration technique. Figure 1 depicts the human–machine collaborative approach.

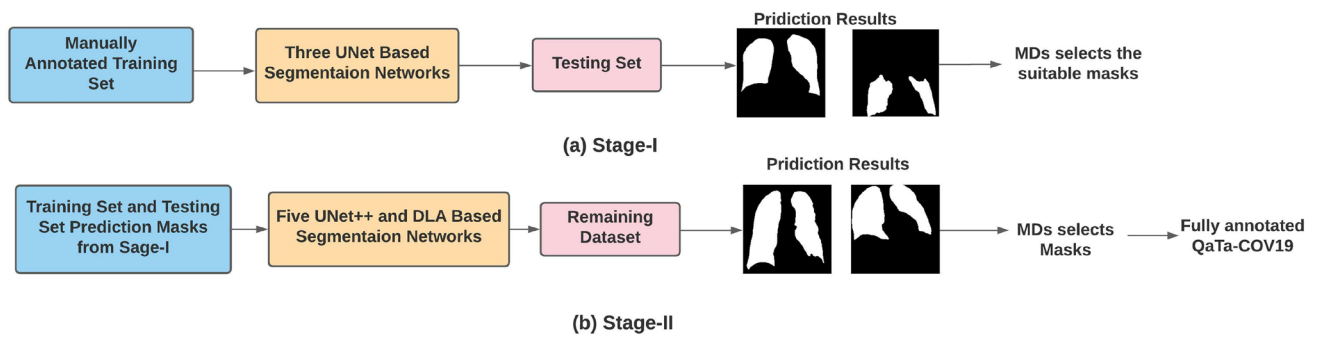
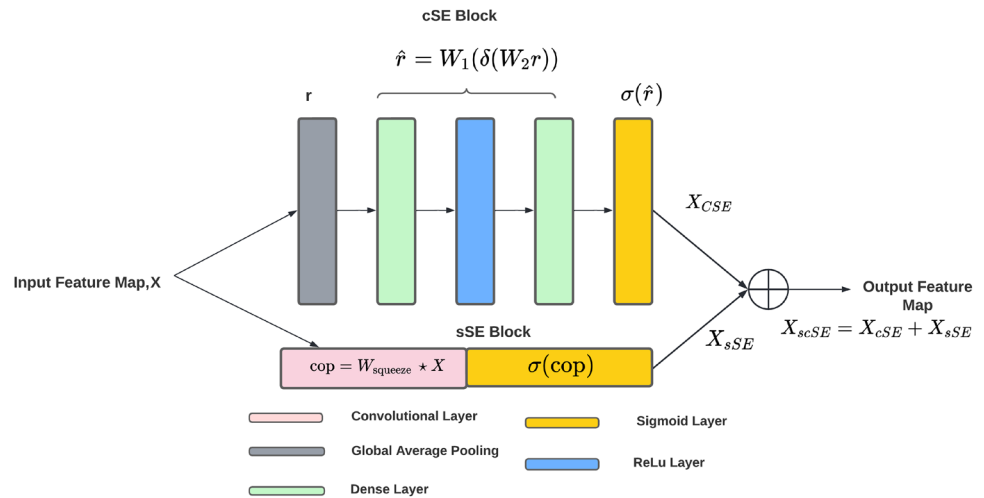


Fig. 1 Human-machine collaborative approach to generate the COVID-10-infected region ground-truth masks

Fig. 2 scSE attention mechanism



3.2 Attention mechanism

UNet is the encoder–decoder architecture, where the encoder is the convolutional neural network (CNN). The CNN’s are effective for a variety of visual applications. The image’s high-level features are captured by stacking convolutional layers with non-linear activation functions and down-sampling. CNN models have performed better than human experts for the disease classifications in CXRs [39]. The attention method can be used for any feature representation challenge, and it encourages the model to record rich contextual interactions to improve feature representations. So, squeeze and excitation network (SE) was introduced by Hu et al. [22] to calibrate channel-wise feature responses by explicitly modeling interdependencies between channels. In this study, to leverage the SE blocks performance in image classification, we introduce the spatial and channel squeeze and excitation block (scSE) [42] in the encoder blocks. The scSE comprise SE [22] and the newly introduced channel squeeze and spatial excitation (sSE) blocks. SE block has the global average pooling, and due to this every intermediary layer has the whole receptive field of the input image, which is an advantage of utilizing such a

block. Further, sSE blocks use pixel-wise spatial information for the fine-grained segmentation task. These sSE blocks do not change the receptive field but provide spatial attention to focus on certain regions. Figure 2 shows the scSE blocks. In the following, the scSE attention mechanism is explained.

Let the input feature map X is combination of channels x_i where, $X = [x_1, x_2, \dots, x_C]$ and $x_k \in R^{H \times W}$. Here C is the number of channel, H and W is the spatial and width, respectively. The global average pooling performs the spatial squeeze and generates the vector or tensor $r \in R^{1 \times 1 \times C}$ with its k^{th} element

$$r_k = \frac{1}{H \times W} \sum_i^H \sum_j^W x_k(i, j) \tag{1}$$

Then this vector r is transformed to $\hat{r} = W_1(\delta(W_2r))$ after applying the dense connected layers. W_1 and W_2 are weights of densely connected layers ($W_1 \in R^{C/p \times C}$, $W_2 \in R^{C \times C/p}$), while p is the reduction ratio, δ refers to the ReLU activation function. In this study, reduction ratio is set to 16. Further, to learn the channel-wise dependencies, the \hat{r} is passed through the Sigmoid activation layer $\sigma(\hat{r})$. The resultant tensor is shown in below equation:

$$X_{cSE} = [\sigma(\hat{r}_1)x_1, \sigma(\hat{r}_2)x_2, \dots, \sigma(r_{32})z_C] \tag{2}$$

In sSE, the feature map X is squeezed along the channel and exited spatially. Here, the input tensor can be considered as $X = [x^{1,1}, x^{1,2}, \dots, x^{i,j}, \dots, x^{H,W}]$ where $x^{i,j} \in R^{1 \times 1 \times C}$, i and j indicates the spatial location with $i \in 1, 2, \dots, H, j \in 1, 2, \dots, W$. Then the convolutional operation, $cop = W_{squeeze} \star X$, generating a vector $cop \in R^{H \times W}$ with weight $W_{squeeze} \in R^{1 \times 1 \times C \times 1}$. Then this scale is passed through the Sigmoid layer to excite X , spatially:

$$X_{sSE} = [\sigma(cop_{1,1})x^{1,1}, \dots, \sigma(cop_{i,j})x^{i,j}, \dots, \sigma(cop_{H,W})x^{H,W}] \tag{3}$$

Each value of $\sigma(cop_{i,j})$ represents the important spatial information of the corresponding feature map. In scSE activation, X_{cSE} and X_{sSE} are added to achieve the spatial and channel squeeze & excitation attention mechanism:

$$X_{scSE} = X_{cSE} + X_{sSE} \tag{4}$$

3.3 Atrous spatial pyramid pooling module

To perform accurate semantic segmentation, both local and global features are important. The encoder path in the UNet, on the other hand, is always confined to a smaller receptive field that captures the global information in the features [37]. The atrous convolutional layer can be used to deal with the small receptive field problem in the UNet encoder. Atrous convolution was introduced by Holschneider et al. [21] for efficient computing. Later, it was used in context of deep learning [17, 36, 45]. The atrous convolution-based deep learning models have been continuously investigated for semantic convolution. Within deep learning networks, it allows to deliberately control the resolution at which feature responses are generated. It also enables to effectively expand the field of view of filters to include more context without increasing the number of parameters or computation time [8]. The dilation rate is the additional parameter in the convolutional layer that introduces the space (zeros) consecutive kernel values. The dilation rate r introduces the $(r - 1)$ zeros, enlarging the filter of k size to K . The formula for enlarged filter size (EFS) is given in the below equation.

$$EFS_r = K = k + (k - 1) \times (r - 1) \tag{5}$$

The procedure performed does not affect the computational complexity or the number of parameters. For example, in a typical convolutional layer, a kernel of size 3 with a dilation rate of 2 is comparable to a kernel of size 5. It provides an effective technique for controlling the field of view and determining the appropriate trade-off between accurate localization (small field of view) and a large field of view.

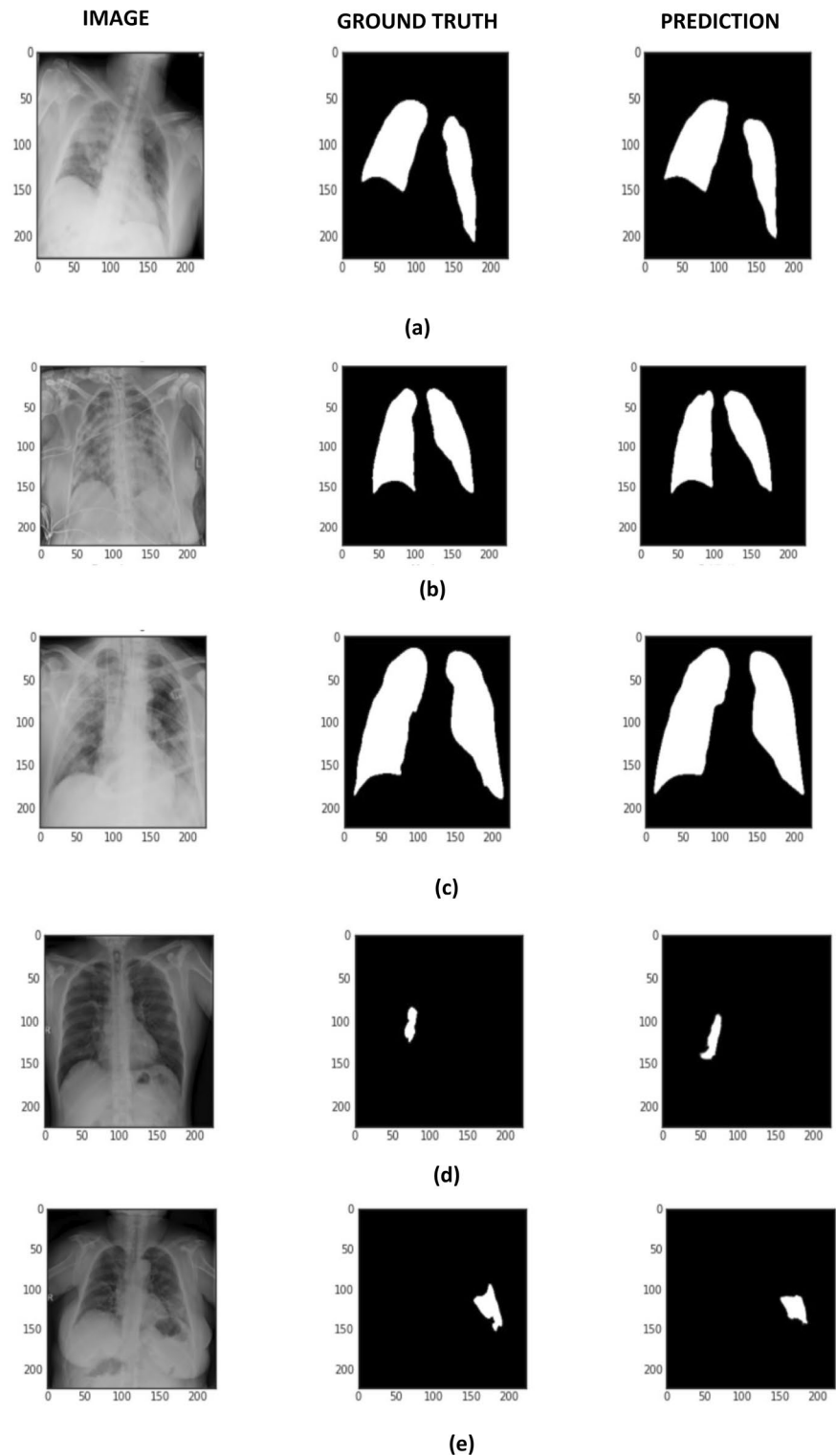
The same 5×5 receptive field can be obtained by stacking two 3×3 convolutional layers, but this increases the number of parameters in the models and slows the training process. [8] proposed the atrous spatial pyramid pooling (ASPP) module for semantic segmentation to address the lost spatial problem utilizing variable dilation rates. They utilized the four parallel convolutional layers (3×3 kernels) with a dilation rate of 6, 12, 18, and 24. Later, in their other works [9, 10], they proposed the different variations in the ASPP module.

In our study, we have used the three parallel layers (3×3 filters) with 2, 3, and 4 dilation rates. The output of all three convolutional layers is concatenated and then passed by the convolutional layer with 1×1 kernel followed by batch normalization [24] and the ReLu activation function. We have used the ASPP module in every block of the encoder. Dilation rate is a critical hyper-parameter that must be properly selected. The behavior of the convolutional layer (3×3) can change if we employ a large dilation rate [37]. Specifically, for COVID-19 lesion segmentation in CXRs, a high dilation rate is not suitable. As can be seen in Figs. 3 and 4 the COVID-19 lesion area varies significantly. Therefore, a very high dilation rate can lead to the loss of fine-grained lesion contours in the lungs. For example, in the convolutional layer with kernel size 3 and dilation rate 32, there will be 31 zeros (spaces) between two non-zero numbers. As a result, critical information will likely be lost totally. Figure 5a illustrate the ASPP module, and Fig. 5b illustrate the convolutional weight when the dilation rate is 1 and 2.

3.4 Proposed architecture

The proposed UNet-based encoder–decoder architecture for COVID-19 lesion segmentation in CXRs is shown in Fig. 6. The encoder is essentially a CNN that downsamples to capture high-level characteristics. In this paper, max-pooling has been used for downsampling the resolution of the feature map. There are five blocks in encoder and decoder, respectively. In each encoder block, there is a convolutional block, a scSE block, and an ASPP module. Each convolutional block consists of a convolutional layer followed by batch normalization and the ReLU activation function. Every convolutional block is followed by the attention mechanism for the recalibration of intermediate feature maps. For the same, we have used the spatial and channel squeeze and excitation (scSE) block. These attention-based blocks can be exploited at any depth in the architecture. So, to gain the most from feature calibration, we placed scSE blocks in every encoder block. Further, we have introduced the ASPP module in the architecture to take advantage of varying receptive fields with different dilation rates. We have kept the dilation rate small, so local information is not lost. Each ASPP module

Fig. 3 Qualitative results of the proposed model for COVID-19 lesion segmentation on severe, moderate, and mild cases. **a–c** shows the prediction results severe and moderate cases, while **(d)** and **(e)** on the mild cases



has three parallel 3×3 convolutional layers with 2, 3 and 4 dilation rates. Then, their output is concatenated and passed through the 1×1 convolutional layer, followed by the batch normalization and activation function. The first block of the encoder has 16 kernels in every convolutional layer of the convolutional block and the ASPP block. Thereafter, in each

encoder block, the number of kernels in convolutional layers is increased by 2.

In the decoder part, there are also the five blocks corresponding to each encoder block in the architecture. In the decoder block, we have performed the upsampling to regain the spatial information lost during the downsampling

Fig. 4 These are the prediction results for moderate and mild cases on which the proposed model failed to produce accurate lesion segmentation

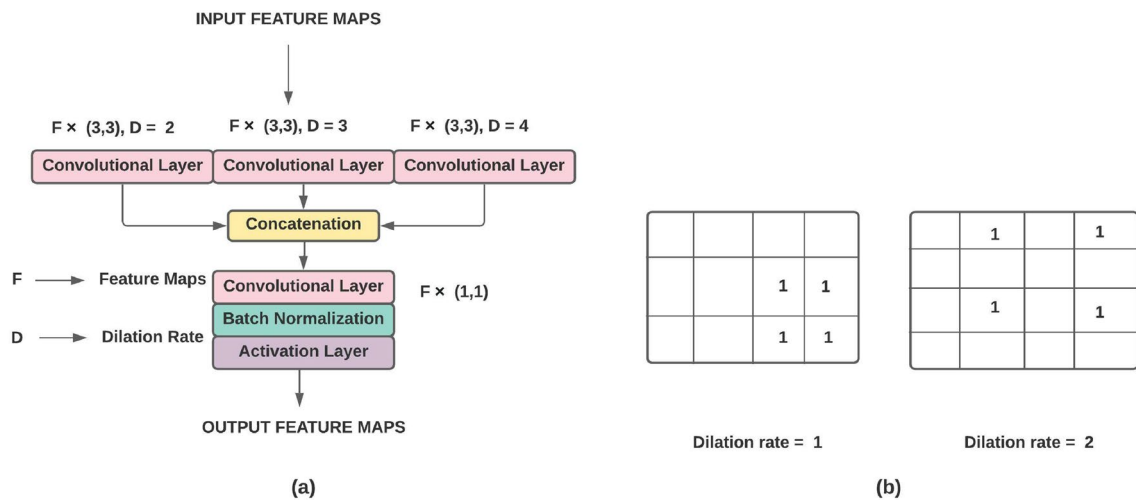
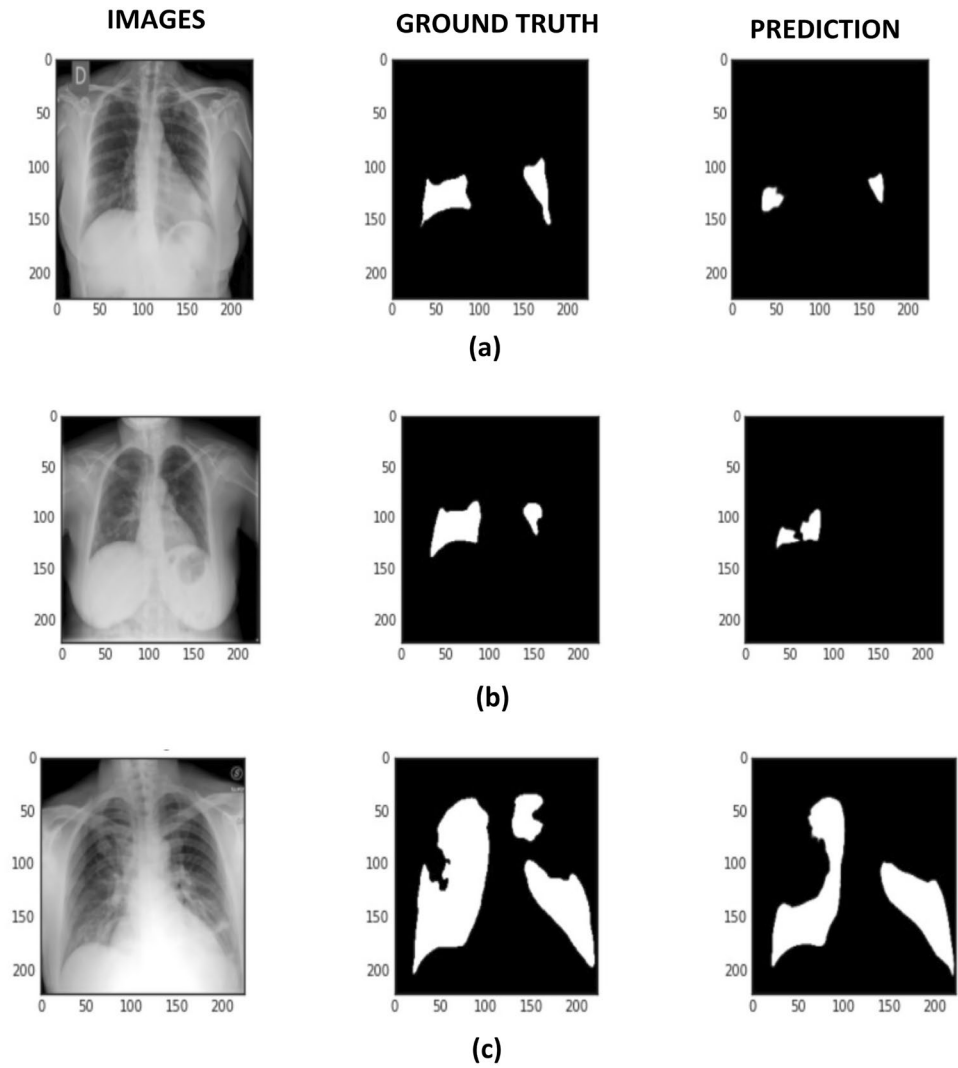


Fig. 5 a ASPP module b Enlarged filter size when dilation rate is 1 and 2

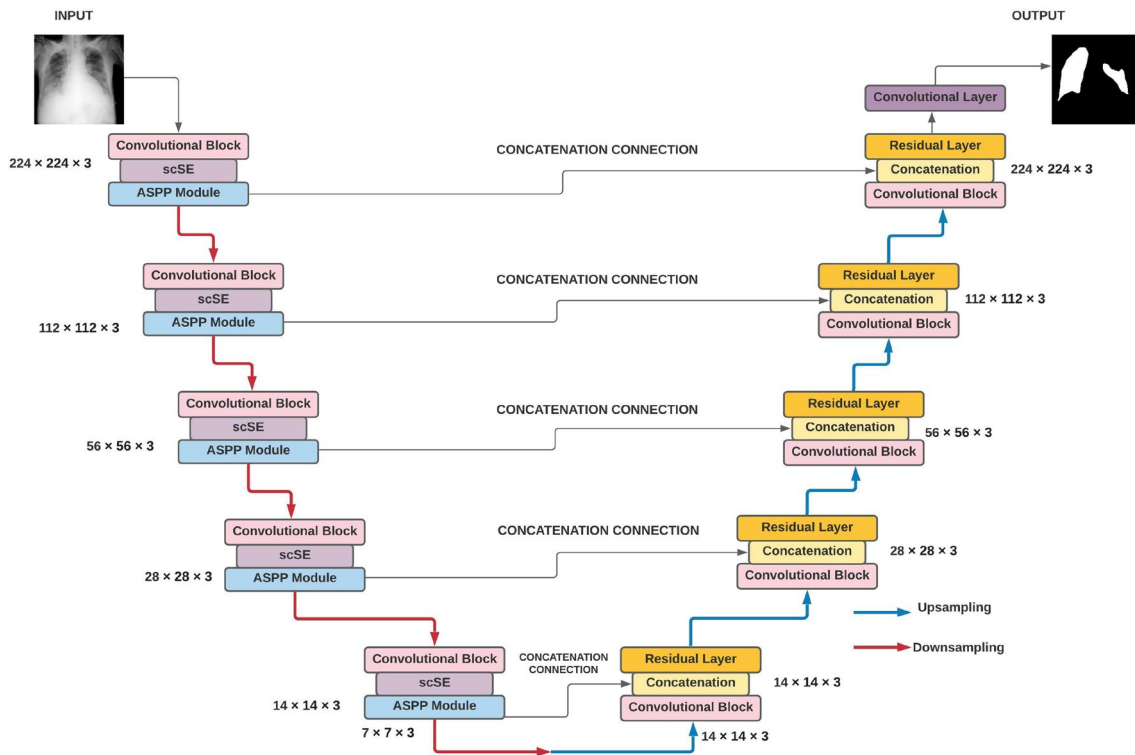


Fig. 6 Proposed encoder–decoder-based architecture

operation. The upsampled feature map is then fed to the convolutional block, and its output is concatenated with the feature maps of the corresponding encoder layer. This is an important step in localizing the high-resolution features. At last, these concatenated feature maps are passed through the residual layer. After each decoder block in the architecture (bottom to top), the number of feature maps in the convolutional layer is dropped by half. Finally, after the last decoder block, there is a 1×1 convolutional layer with Sigmoid activation function to produce the segmented COVID-19 lesion.

4 Performance evaluation

In this section, we have discussed the metrics used to evaluate the model for COVID-19 lesion segmentation, implementation details, and the obtained results.

4.1 Metrics

In this study, we have used the dice similarity coefficient (DSC) and jaccard index (JI) to evaluate the proposed model for COVID-19 lesion segmentation. We have also compared the predicted mask of the proposed model with the ground-truth segmented by an expert radiologist. In the following, we have presented the DSC and JI equations:

$$\text{DSC} = \frac{2 \times TP}{(TP + FP) + (TP + FN)} \quad (6)$$

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

where true positive values are denoted by TP, false-positive values are denoted by FP, and false-negative values are denoted by FN. If the model recognizes COVID-19 lesion pixels as exactly ground truth pixels, it will be TP. False-positive (FP) pixels are those that are incorrectly classified as COVID-19 lesion but are normal pixels, while false-negative (FN) pixels are those that belong to the COVID-19 lesion but are incorrectly identified as normal pixels.

4.2 Implementation details

Google Colaboratory, referred to as Colab, is used for experimentation. It provides 25 GB RAM and a P100 GPU for 24 h. All the experiments are implemented in Python using the Keras library. The Adam optimizer is employed with a 0.0005 learning rate, and the number of epochs is 200 with batch size 4. We have employed a dynamic approach to training the model. When the model stops improving, the ReduceLROnPlateau technique is used to adjust the learning rate. The minimum rate of learning is

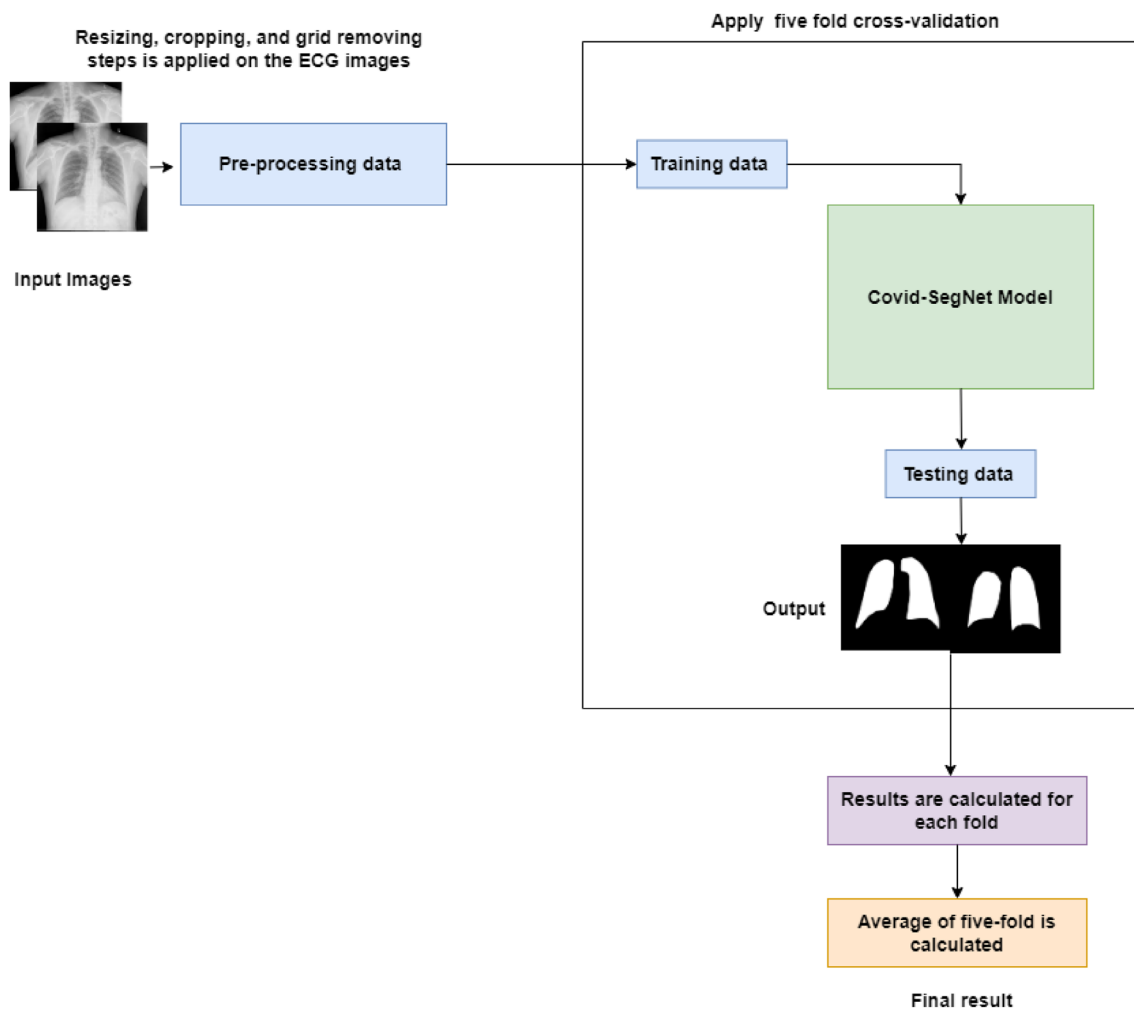


Fig. 7 Framework of the proposed work for COVID-19 lesion segmentation

fixed at $1e-6$ with a 0.5 factor, and patience equals 4. The validation loss is monitored at every epoch, and if it does not improve for a continuous 25 epoch, then the training of the model is stopped.

Cross-entropy is the popular loss function, but in the case of a class-imbalance problem, the dice coefficient loss performs better [18, 25]. So, in this study, we have used the dice coefficient loss function. Its calculation formula is $1 - DSC$. A cross-validation approach can be used to check the general effectiveness of the model [23]. In the K-fold cross-validation, the dataset is randomly divided into mutually exclusive K folds of equal or nearly equal size [23, 26]. The 5-fold cross-validation method is employed in this study. The dataset is split into five sets, with four sets used for training and the fifth set used to test the model. This procedure is repeated five times, with the findings for each test set being recorded. Finally, the proposed architecture is evaluated using the average of all five results. The framework of the proposed study is shown in Fig. 7.

4.3 Results

In the following, we have presented and discussed the obtained results. In addition, we have performed the ablation study to demonstrate the influence of the attention mechanism and ASPP module. We have also shown the effect of dilation rate on the results. For comparison, we have used the using DSC and JI values.

4.3.1 Proposed architecture

In the following, we have discussed the results obtained by the proposed architecture for COVID-19 lesion segmentation. We have also compared the predicted masks with the ground-truth masks produced by the human–machine collaborative approach.

In this study, we have used the ASPP module and the attention mechanism to gain from an expanded field of view and improved feature representations, respectively. Table 1

Table 1 Obtained results for the COVID-19 lesion segmentation by proposed model

Dilation Rate	Fold	DSC	JI
(2,3,4)	1	0.8202	0.6952
–	2	0.8292	0.7083
–	3	0.8407	0.7253
–	4	0.8361	0.7184
–	5	0.8364	0.7189
–	Average	0.8325	0.7132

The bold values represent the results obtained by the proposed model

shows the DSC and JI values obtained by the proposed model in each fold. The model has obtained the highest DSC and JI values of 0.8407 and 0.7253 in fold 3. While, the lowest DSC and JI values of 0.8202 and 0.6952 are obtained in fold 1. The average DSC value of 0.8325 and JI value of 0.7132 is obtained by COVID-SegNet for all five-folds.

Figures 3 and 4 shows the qualitative results of our proposed model for the COVID-19 lesion segmentation in CXRs. Figure 3a, b, and c shows the COVID-19 lesion segmentation ground-truth and prediction results for the severe and moderate cases. In Fig. 3, it can be clearly seen that the proposed model is able to accurately segment the COVID-19 lesion area for severe and moderate cases. While Fig. 3d and e show the prediction results on the mild COVID-19 cases. In both, Fig. 3c and d, the model again has predicted correctly the COVID-19 lesion location. It has slightly predicted a larger lesion area in Fig. 3d and has slightly missed the lower region in Fig. 3e. Overall, the location of the COVID-19 has been predicted accurately.

Figure 4 shows the qualitative results on which the proposed model failed to segment the lesion correctly. However, in Fig. 4a it predicted the location correctly but does not segment the whole lesion. While in Fig. 4b, it predicted the location correctly in one lung and completely missed the lesion in the second lung. In Fig. 4c also as in Fig. 4b, it correctly predicts the lesion area in the lower lungs, while missing the lesion in the upper lungs. In other words, the proposed model can segment the lesion in these unsuccessful cases, but with false-negative regions. The reason for the failure cases can be the presence of rib-cages and clavicles bone in CXR images. The lesion might be beneath these bones, which the model may fail to predict, and this can result in an increase in false negatives, which results in a decrease in the efficiency of the model.

4.3.2 Ablation study

Table 2 shows the results obtained by different ablation studies. These studies have been conducted to investigate the

Table 2 Results of ablation studies

Ablation studies	DSC	JI
Without dilation	0.8290	0.7080
Dilation rate (3,5,7)	0.8310	0.7110
Without scSE	0.8315	0.7119
scSE replaced by SE	0.8308	0.7107
ASPP module replaced by convolutional block	0.8300	0.7094
Proposed model	0.8325	0.7132

The bold values represent the results obtained by the proposed model

individual contributions of the different modules in improving the results of the proposed model. In the following, studies and their results are discussed.

In these ablation studies, we have not changed the decoder. The purpose of this experiment is to show that the atrous convolution can capture the multi-scale context with varying dilation rates in increasing the efficiency of the model. Therefore, in the first study, we remove the dilation rate from all the convolutional layers in the ASPP block to evaluate the effect. Upon evaluation, the proposed model performance dropped to a DSC value of 0.8290 and a JI value of 0.7080. This drop in results shows that the multi-scale view can help increase the performance of the model. The multi-scale view helps capture better feature representation to improve efficiency.

High dilation rates are used in different DeepLab versions [8–10] for semantic segmentation in benchmark datasets, but the use of high dilation rates is not recommended in medical images [37]. In the second study, we have used (3,5,7) the dilation rate to evaluate the effect of increased dilation on the results. The proposed model reported a DSC value of 0.8310 and a JI value of 0.7110. These findings provide further confirmation that the use of a high dilation rate is not appropriate when dealing with segmentation issues in medical images. Because of the high rate of dilation, there is an increase in distance between kernel elements, which may result in the loss of fine-grained information. In Qamar et al. [37] also, the best segmentation results for skin lesions were obtained with the small dilation rates. These two ablation studies show the importance of atrous convolution and particularly small dilation rates in increasing the performance of the model for segmentation in CXRs.

In the third and fourth study, we have evaluated the impact of the attention mechanism on the performance of the model. The attention mechanism helps in extracting the useful feature from the image. Therefore, in the third study, we have shown that the spatial and channel-wise attention increase the performance of the model. Therefore, after removing the scSE attention mechanism, and upon evaluation, the DSC value of 0.8315 and the JI value of 0.7119 are

obtained. In the fourth study, we evaluated the model with the SE model [22] to check how the model performs with the channel excitation mechanism, while the scSE attention mechanism employs both the channel and spatial mechanisms. So, we replaced the scSE block with the SE module. It obtained a DSC value of 0.8308 and a JI value of 0.7107. These results show that spatial and channel-wise attention, when used together, increase the performance of the model.

In the fifth study, we replaced the ASPP block with the convolutional block to evaluate its contribution to the proposed model. In doing so, the performance of the model dropped to a DSC value of 0.8300 and a JI value of 0.7094. These obtained results show the significant role played by the ASPP block in improving the performance of the proposed model. So, the obtained results in these ablation studies suggest that the attention mechanism, the ASPP module, has individually contributed to improving the performance of the proposed model for COVID-19 lesion segmentation.

5 Results analysis and discussion

This section presents a comparative results, computational analysis of UNet and other related works with the proposed model.

5.1 The proposed architecture vs state-of-the-art architectures

UNet [41] model is an encoder–decoder architecture. UNet is divided mainly into three parts: encoder, bridge module and decoder. There are four encoder–decoder blocks and a bridge module. Each block in the encoder consists of a two convolutional layer stack: ReLu activation and a max-pooling layer for contraction. The bride module has two convolutional layers with a ReLu activation function. In the decoder, the feature map is upsampled and then concatenated with the feature map of the corresponding encoder layer. These feature maps then work as an input to the stack of convolutional layers. At last, the output is obtained with a 1×1 convolutional layer and the Sigmoid activation function.

For the UNet training, all the hyperparameters are kept the same as in the proposed model except the learning rate. It performed badly at a 0.0005 learning rate, so we have used a 0.00001 learning rate. Both UNet and proposed models are evaluated using 5-fold cross-validation. In Table 3, we can see that the proposed model outperformed the original UNet on both DSC and JI evaluation metrics.

Apart from the UNet model, the proposed COVID-SegNet model has also been compared against state-of-the-art Attention UNet [34], R2UNet [5], and UNet++ [51]. The attention UNet model obtained the DSC value of 0.8124 and 0.6875 JI value. While R2UNet obtained the 0.8317 DSC

Table 3 Quantitative comparison between proposed model and state-of-the-art architectures

Models	DSC	JI
UNet	0.7950	0.6598
Attention UNet	0.8124	0.6875
R2UNet	0.8317	0.7087
UNet++	0.8297	0.7094
Proposed Model	0.8325	0.7132

The bold values represent the results obtained by the proposed model

value and 0.7087 JI value. Moreover, the UNet++ obtained a 0.8297 DSC value and a 0.7094 JI value. The COVID-SegNet model outperformed all the state-of-the-art models for the COVID-19 lesion segmentation.

5.2 Comparative analysis with related works

There are numerous studies available for the COVID-19 detection in CXRs and CTs. Ozturk et al. [35], Agrawal and Choudhary [2], and Ahuja et al. [3] performed the image-level classification for the detection of COVID-19.

In the image-level prediction, the model does not localize the lesion area. The lesion segmentation can help the radiologist identify the severity of the patient. Therefore, Laradji et al. [27] proposed the deep learning model for COVID-19 lesion segmentation in CTs. It obtained a DSC value of 0.730 and a JI value of 0.570 on 60 CTs images. In other work, Amyar et al. [6] proposed the UNet model for COVID-19 lesion segmentation in CTs which achieved the 0.880 DSC value on 100 CTs images.

Most of the lesion segmentation studies use CTs images as modality. QaTa-COV19 dataset [11] is the only CXR dataset available for the COVID-19 lesion segmentation along with ground-truth masks. This dataset uses a unique human–machine collaborative approach for preparing the ground-truth masks of the COVID-19 lesion. Apart from annotating the dataset, Degerli et al. [11] have also evaluated different models using a five-fold cross-validation scheme. They have trained the three segmentation networks (UNet, UNet++, and DLA) with different pre-trained encoders (CheXNet, DenseNet-121, Inception-v3, and ResNet-50). Despite the lack of a powerful transfer learning approach and fewer parameters, our proposed model produced significantly comparable outcomes. In some cases, the proposed model outperformed the pre-trained CheXNet and Inception-v3 encoder based on UNet networks on DSC values. The proposed model is lightweight in terms of parameters. It has just 6.5 million parameters. While the CheXNet and inception-v3 model has total 12.15 million and 29.94 million parameters. The COVID-SegNet model is five times light weighted than the Inception-v3 model and two times

Table 4 State-of-the-art works reported for COVID-19 lesion segmentation and detection in CXRs and CTScs

Model	Modality	COVID-19 Images	Application	Technique	DSC	JI	IT (ms)
VGG-16 [27]	CTS	60	Lesion segmentation	VGG16 as encoder, transfer learning for training	0.730	0.570	–
MTL [6]	CTS	100	Lesion segmentation	Tailored-design CNN, training from scratch	0.880	–	–
Inception-v3 [11]	X-ray	2951	Lesion segmentation	Inception-v3 as encoder, transfer learning for training	0.8147	–	2.53
CheXNet [11]	X-ray	2951	Lesion segmentation	CheXNet as encoder, transfer learning for training	0.8290	–	2.56
LViT [29]	X-ray	2951	Lesion Segmentation	Vision Transformer	0.8366	–	–
Covid-SegNet	X-ray	2951	Lesion segmentation	Tailored-design CNN, ASPP, attention mechanism, training from scratch	0.8325	0.7132	6.7

The bold values represent the results obtained by the proposed model

light weighted than the CheXNet model. The training time of CheXNet and the Inception-v3 model is not specified in the articles; however, the inference time (IT) is reported. The training and inference time of the models on the given dataset can vary as they depend on the hardware's computational power. The proposed model (IT) is 6.7 milliseconds, compared to 2.53 milliseconds and 2.56 milliseconds for Inception-v3 and CheXNet, respectively, for a single image. Meanwhile, Laradji et al. [27] proposed model took 1–3 s to label a pixel of infected region. The COVID-SegNet tool five hours approximately for the training. LViT [29] model proposed the infection segmentation using the vision transformer models. It obtained a 0.8366 DSC value. The DSC value obtained by the LViT model is slightly higher than the value obtained by the COVID-SegNet model. However, the reason for the higher value is that the LViT model has been evaluated with a single train-test split and not with the cross-validation scheme as we did for COVID-SegNet. Further, for one-to-one comparison with the LViT model, we have also trained and tested the COVID-SegNet model with a single train-test split. And, the proposed COVID-SegNet model obtained 0.8472 DSC value. Again, the performance evaluation with cross-validation is recommended because it gives generalized results for the dataset. The training and inference time of the LViT model is not given in the paper. Table 4, we have given the brief overview of the works performed for COVID-19 using the deep models and the proposed model.

6 Conclusion

The rapid and accurate diagnosis of highly contagious COVID-19 is critical in halting the virus spread. Automatic diagnosis can ease the burden on the medical community. Therefore, CXRs are used in this paper because they are less expensive, more accessible, and faster than other regularly used procedures like RT-PCR and CTS. In this paper, the UNet-based encoder–decoder architecture is proposed for

the COVID-19 lesion segmentation. The COVID-19 lesion segmentation in CXRs will increase the diagnosis precision. The attention-based mechanism and the ASPP blocks are utilized to increase the performance of the model. The COVID-SegNet obtained 0.8325 DSC and 0.7132 JI, which outperformed the other models on CXR images. Furthermore, an ablation study shows that the atrous convolution, attention mechanism, and the ASPP module helped enhance the performance of the model. The limitation of the proposed model is that it failed in some mild COVID-19 cases. However, it correctly predicts the location of the lesion by segmenting a portion of it. The DSC and JI values obtained are very low. Therefore, in future work, we will work to improve the DSC and JI values and the lesion segmentation in mild cases. The availability of annotated data is another issue while working in medical image analysis. Because the number of labeled datasets is limited and the number of unlabeled datasets is large, self-supervised learning is gaining popularity. Genetic algorithms can also be used to optimize the UNet models. Therefore, in future, self-supervised learning and genetic algorithms can be used for COVID-19 lesion segmentation.

Author contributions TA implement and wrote the paper. PC prepared all figures. TA and PC reviewed the manuscript.

Data availability The data used in the experiments is publicly available at <https://www.kaggle.com/datasets/ayseudegerli/qatacov19-dataset>.

Declarations

Conflict of interest Authors do not have any competing interests to disclose.

References

1. Agnes, S.A., Anitha, J., Peter, J.D.: Automatic lung segmentation in low-dose chest ct scans using convolutional deep and wide

- network (cdwn). *Neural Comput. Appl.* **32**(20), 15845–15855 (2020)
2. Agrawal, T., Choudhary, P.: Focuscovid: automated covid-19 detection using deep learning with chest x-ray images. *Evol. Syst.* 1–15 (2021)
 3. Ahuja, S., Panigrahi, B.K., Dey, N., Rajinikanth, V., Gandhi, T.K.: Deep transfer learning-based automated detection of covid-19 from lung ct scan slices. *Appl. Intell.* **51**(1), 571–585 (2021)
 4. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., Xia, L.: Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology* **296**(2), E32–E40 (2020)
 5. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. Preprint at [arXiv:1802.06955](https://arxiv.org/abs/1802.06955) (2018)
 6. 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.* **126**, 104037 (2020)
 7. Candemir, S., Antani, S.: A review on lung boundary detection in chest x-rays. *Int. J. Comput. Assist. Radiol. Surg.* **14**(4), 563–576 (2019)
 8. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2017)
 9. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. Preprint at [arXiv:1706.05587](https://arxiv.org/abs/1706.05587) (2017)
 10. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818 (2018)
 11. Degerli, A., Ahishali, M., Yamac, M., Kiranyaz, S., Chowdhury, M.E., Hameed, K., Hamid, T., Mazhar, R., Gabbouj, M.: Covid-19 infection map generation and detection from chest x-ray images. *Health Inf. Sci. Syst.* **9**(1), 1–16 (2021)
 12. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. IEEE (2009)
 13. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. Preprint at [arXiv:2010.11929](https://arxiv.org/abs/2010.11929) (2020)
 14. ElAraby, M.E., Elzeki, O.M., Shams, M.Y., Mahmoud, A., Salem, H.: A novel gray-scale spatial exploitation learning net for covid-19 by crawling internet resources. *Biomed. Sig. Process. Control* **73**, 103441 (2022)
 15. Elzeki, O.M., Shams, M., Sarhan, S., Abd Elfattah, M., Hassanien, A.E.: Covid-19: a new deep learning computer-aided model for classification. *PeerJ Comput. Sci.* **7**, e358 (2021)
 16. Gao, K., Su, J., Jiang, Z., Zeng, L.L., Feng, Z., Shen, H., Rong, P., Xu, X., Qin, J., Yang, Y., et al.: Dual-branch combination network (den): towards accurate diagnosis and lesion segmentation of covid-19 using ct images. *Med. Image Anal.* **67**, 101836 (2021)
 17. Giusti, A., Cireşan, D.C., Masci, J., Gambardella, L.M., Schmidhuber, J.: Fast image scanning with deep max-pooling convolutional neural networks. In: *2013 IEEE International Conference on Image Processing*, pp. 4034–4038. IEEE (2013)
 18. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **58**, 101563 (2019)
 19. Hassan, E., Shams, M.Y., Hikal, N.A., Elmougy, S.: Covid-19 diagnosis-based deep learning approaches for covidx dataset: A preliminary survey. *Artificial Intelligence for Disease Diagnosis and Prognosis in Smart Healthcare* p. 107 (2023)
 20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
 21. Holschneider, M., Kronland-Martinet, R., Morlet, J., Tchamitchian, P.: A real-time algorithm for signal analysis with the help of the wavelet transform. In: *Wavelets*, pp. 286–297. Springer (1990)
 22. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141 (2018)
 23. Ibtehaz, N., Rahman, M.S.: Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Netw.* **121**, 74–87 (2020)
 24. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *International conference on machine learning*, pp. 448–456. PMLR (2015)
 25. Khanna, A., Londhe, N.D., Gupta, S., Semwal, A.: A deep residual u-net convolutional neural network for automated lung segmentation in computed tomography images. *Biocybern. Biomed. Eng.* **40**(3), 1314–1327 (2020)
 26. KOHAVI, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International Joint Conference on Artificial Intelligence*, pp. 1137–1143 (1995)
 27. Laradji, I., Rodriguez, P., Manas, O., Lensink, K., Law, M., Kurzman, L., Parker, W., Vazquez, D., Nowrouzezahrai, D.: A weakly supervised consistency-based learning method for covid-19 segmentation in ct images. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2453–2462 (2021)
 28. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
 29. Li, Z., Li, Y., Li, Q., Zhang, Y., Wang, P., Guo, D., Lu, L., Jin, D., Hong, Q.: Lvit: language meets vision transformer in medical image segmentation. Preprint at [arXiv:2206.14718](https://arxiv.org/abs/2206.14718) (2022)
 30. Liang, T., Liu, Z., Wu, C.C., Jin, C., Zhao, H., Wang, Y., Wang, Z., Li, F., Zhou, J., Cai, S., et al.: Evolution of ct findings in patients with mild covid-19 pneumonia. *Eur. Radiol.* **30**(9), 4865–4873 (2020)
 31. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
 32. Mansour, R.F., Escorcía-Gutiérrez, J., Gamarra, M., Gupta, D., Castillo, O., Kumar, S.: Unsupervised deep learning based variational autoencoder model for covid-19 diagnosis and classification. *Pattern Recogn. Lett.* **151**, 267–274 (2021)
 33. Mohammadi, A., Wang, Y., Enshaei, N., Afshar, P., Naderkhani, F., Oikonomou, A., Rafiee, J., de Oliveira, H.R., Yanushkevich, S., Plataniotis, K.N.: Diagnosis/prognosis of covid-19 chest images via machine learning and hypersignal processing: challenges, opportunities, and applications. *IEEE Signal Process. Mag.* **38**(5), 37–66 (2021)
 34. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. Preprint at [arXiv:1804.03999](https://arxiv.org/abs/1804.03999) (2018)
 35. Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Acharya, U.R.: Automated detection of covid-19 cases using deep neural networks with x-ray images. *Comput. Biol. Med.* **121**, 103792 (2020)

36. Papandreou, G., Kokkinos, I., Savalle, P.A.: Modeling local and global deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 390–399 (2015)
37. Qamar, S., Ahmad, P., Shen, L.: Dense encoder-decoder-based architecture for skin lesion segmentation. *Cogn. Comput.* **13**(2), 583–594 (2021)
38. Ragab, M., Alshehri, S., Alhakamy, N.A., Mansour, R.F., Koundal, D.: Multiclass classification of chest x-ray images for the prediction of covid-19 using capsule network. *Comput. Intell. Neurosci.* **2022** (2022)
39. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., et al.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. Preprint at [arXiv:1711.05225](https://arxiv.org/abs/1711.05225) (2017)
40. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)
41. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer (2015)
42. Roy, A.G., Navab, N., Wachinger, C.: Recalibrating fully convolutional networks with spatial and channel “squeeze and excitation” blocks. *IEEE Trans. Med. Imaging* **38**(2), 540–549 (2018)
43. Salehi, S., Abedi, A., Balakrishnan, S., Gholamrezanezhad, A.: Coronavirus disease 2019 (covid-19): a systematic review of imaging findings in 919 patients. *Am. J. Roentgenol.* **215**(1), 87–93 (2020)
44. Samee, N.A., El-Kenawy, E.S.M., Atteia, G., Jamjoom, M.M., Ibrahim, A., Abdelhamid, A.A., El-Attar, N.E., Gaber, T., Slowik, A., Shams, M.Y.: Metaheuristic optimization through deep learning classification of covid-19 in chest x-ray images. *Computers, Materials and Continua*, pp. 4193–4210 (2022)
45. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. Preprint at [arXiv:1312.6229](https://arxiv.org/abs/1312.6229) (2013)
46. Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., Xue, Z., Shen, D., Shi, Y.: Abnormal lung quantification in chest ct images of covid-19 patients with deep learning and its application to severity prediction. *Med. Phys.* **48**(4), 1633–1645 (2021)
47. Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., Shen, D.: Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for covid-19. *IEEE Rev. Biomed. Eng.* **14**, 4–15 (2020)
48. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Preprint at [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
49. Tang, Y.B., Tang, Y.X., Xiao, J., Summers, R.M.: Xlsor: A robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation. In: International Conference on Medical Imaging with Deep Learning, pp. 457–467. PMLR (2019)
50. Yu, F., Wang, D., Shelhamer, E., Darrell, T.: Deep layer aggregation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2403–2412 (2018)
51. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11. Springer International Publishing, Cham (2018)
52. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11. Springer (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.