Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Effective multiscale deep learning model for COVID19 segmentation tasks: A further step towards helping radiologist

Abdul Qayyum [a,*], Alain Lalande [a,b], Fabrice Meriaudeau [a]

[a] *ImViA Laboratory, University of Bourgogne Franche-Comŕe, Dijon, France*
[b] *Medical Imaging Department, University Hospital of Dijon, Dijon, France*

ABSTRACT

Infection by the SARS-CoV-2 leading to COVID-19 disease is still rising and techniques to either diagnose or evaluate the disease are still thoroughly investigated. The use of CT as a complementary tool to other biological tests is still under scrutiny as the CT scans are prone to many false positives as other lung diseases display similar characteristics on CT scans. However, fully investigating CT images is of tremendous interest to better understand the disease progression and therefore thousands of scans need to be segmented by radiologists to study infected areas. Over the last year, many deep learning models for segmenting CT-lungs were developed. Unfortunately, the lack of large and shared annotated multicentric datasets led to models that were either under-tested (small dataset) or not properly compared (own metrics, none shared dataset), often leading to poor generalization performance. To address, these issues, we developed a model that uses a multiscale and multilevel feature extraction strategy for COVID19 segmentation and extensively validated it on several datasets to assess its generalization capability for other segmentation tasks on similar organs. The proposed model uses a novel encoder and decoder with a proposed kernel-based atrous spatial pyramid pooling module that is used at the bottom of the model to extract small features with a multistage skip connection concatenation approach. The results proved that our proposed model could be applied on a small-scale dataset and still produce generalizable performances on other segmentation tasks. The proposed model produced an efficient Dice score of 90% on a 100 cases dataset, 95% on the NSCLC dataset, 88.49% on the COVID19 dataset, and 97.33 on the StructSeg 2019 dataset as compared to existing state-of-the-art models. The proposed solution could be used for COVID19 segmentation in clinic applications. The source code is publicly available at https://github.com/RespectKnowledge/Mutiscale-based-Covid-_segmentation-usingDeep-Learning-models.

1. Introduction

COVID-19 has spread all over the world in the last few months and still the number of deaths increases day by day in many countries [1–3]. Computed tomography (CT) is an important technique that plays an important role in the fight against COVID-19 [4–6]. CT has shown good sensitivity in the premature diagnosis of COVID-19 infection. From CT images, quantitative information such as the percentage of high opacity, lung burden, and lung severity score could be used to monitor the disease development and might help clinicians to identify the progression of COVID-19 [7,8].

Common symptoms of COVID-19 patients are fever, cough, and shortness of breath [9] and may also include pneumonia [10,11]. Computed Tomography (CT) imaging plays a vital role in the assessment and detection of appearances associated with COVID-19 in the lungs [12,13]. The segmentation of the infection lesions based on CT scans could provide valuable information for more accurate diagnosis and follow-up assessment [14,11,15] of the disease progression [16]. These aspects are still under active debate [62] and thorough clinical investigations. Nonetheless, as manual segmentation of the lesions from 3D volumes is labor-intensive, time-consuming, and suffers from inter and intra-observer variabilities, automatic segmentation of the lesions is highly desirable in clinical practice [14]. The automatic segmentation of COVID-19 pneumonia lesions from CT scans could be challenging due to various reasons. The main and first reason is the appearance of infection lesions onto different complex forms such as Ground-Glass Opacity (GGO), reticulation, consolidation, and others [9]. The pneumonia lesions' sizes and positions are varying largely at different stages among different patients. Moreover, the lesions have

---

* Corresponding author.
  E-mail address: qayyum@enib.fr (A. Qayyum).

ambiguous boundaries and irregular shapes. The second reason is the lack of annotated data.

Deep learning-based methods have been widely used in medical imaging and quite heavily in fighting against COVID-19 [14] to propose data-driven solutions. For instance, AI could be used to build an imaging workflow that can avoid transmission from patients to health care benefactors [11]. Currently, Convolutional Neural Networks (CNNs) deep learning models have achieved state-of-the-art results for various medical image analysis tasks [3,11] and are showing promising results when applied to COVID-19 [7,12,13]. The automatic infection and diagnosis systems in COVID-19 studies depend on the initial segmentation from deep learning models [12,14,15,16]. Accurate and efficient solutions depend on large datasets and private datasets that may not easily be available publicly. One such effort of annotating thousands of multicentric CT scans should result in better data-driven models but has not been done yet due to lack of time from the clinicians busy fighting the disease.

However, a few small-size datasets, presented in the relevant subsection, exist and can be used as an initial step. As such, we propose a deep learning-based model for COVID-19 Lung Infection Segmentation to address the aforementioned issues. The proposed system can automatically segment small lesions that are scattered at different locations and positions within the lungs. The combination of an appropriate loss function along with a multiscale feature extractor provides an efficient solution. The following contributions have been addressed in this paper.

- The proposed deep learning model learns features at a multiscale level from encoder and decoder to minimize the segmented gap for efficient segmentation
- The proposed decoder and encoder modules used an efficient block that consists of expansion 1x1 Conv, depth-wise, and projection 1x1 Conv to tackle the feature maps using different depths and scales of feature maps. These modules enable to perfectly handle the challenging shape variations of COVID-19 infected areas.
- The proposed model has been validated on different COVID19 and NON-COVID19 segmentation datasets and experiment results prove that the proposed model produced excellent performances on each dataset.
- The different cross-validation techniques have been used to test our proposed model and different performance metrics are used to compare the results. Further statistical analyses have been applied to validate segmentation results.

As COVID-19 datasets are rather small, we are focusing on three areas such as few-shot learning, domain generalization, and knowledge transfer. The proposed model would be helpful to tackle limited COVID-19 CT scans and could generalize well on heterogeneous COVID-19 and non-COVID-19 CT scans.

2. Related work:

The existing classical U-Net or V-Net-based models could achieve promise segmentation performance on a well-labeled hundred cases dataset [14]. The various U-Net based models produced a Dice score range between 83.1% and 91.6% on a hundred cases training datasets. D.-P. Fan et al. [17] presented a deep learning model based on V-Net to segment the infection regions. They achieved dice coefficients between 85.1% and 91.0% with different labeled CT scans. Huang et al. [18] proposed U-Net [19] models based on 774 cases of lung infection for assessment and quantification of the disease area. The training of deep learning models with more datasets on segmentation could increase accuracy.

A few works have been proposed for the automatic segmentation of COVID-19 pneumonia lesions from medical images [14]. Li et al. [13] proposed the U-Net [19] for the segmentation of the lungs and lesions using COVID-19 pneumonia. Cao et al. [15] and Huang et al. [18] also presented work based on U-Net for segmentation of lungs and pulmonary opacities for assessment and quantification of lung diseases. The UNet++ [20] has been proposed for segmentation [21] and detection [22] of infection lesions based on CT scans. In [16], authors proposed VB-Net which is a combined form of V-Net [23] and the bottleneck structure [27] for the segmentation of different structures such as lung segments, infection regions using CT scans based on COVID-19 patients, lung lobes, and a human-in-the-loop approach used for efficient annotation. The authors proposed [20] U-Net++ [20] for the segmentation of lesions from COVID-19 patients based on CT images. They claimed that the proposed work could be used for the segmentation and assessment of COVID-19 and produced a comparable performance as the comparison with expert radiologists for the treatment of COVID19 patients. Other networks such as U-Net [19] and Res-UNet [24] have been used for developing AI-assisted COVID-19 diagnosis systems. Deep learning-based quantitative features could be used for the segmentation of infection regions in lung CT slices and could also be employed for lung infection quantification [25–27] of COVID-19, large-scale screening [28], and severity assessment [29].

Multi-scale features are useful features in many computer vision applications such as segmentation [30], saliency detection [31], and object detection [32]. Normally we can divide Multiscale feature learning into two categories.

In the first category, the combination of different level features using skip connections have been proposed such as skip-net [33], FPN [34], U-Net [19], and FED-Net [35]. These networks used an encoder to capture more context information gradually at downsamples, followed by a decoder that used upsampling layers to cater information in the segmentation. The convolutional blocks or attention gates are used through skip connection to extract low-level fine appearance features and fused them into coarse high-level features. The skip connections fuse multi-scale context and however, at the same time produce a big semantic gap between features at two ends of the connections. UNet++ [20] redesigned the skip-pathways to concatenate similar features at different levels to minimize the semantic gap of skip-net. The Pyramid structures were used to capture multi-scale features in computer vision tasks. The deep supervision method was used in image segmentation [36] and saliency detection [32] to obtain effective pyramid features. SegCaps [37] used a fusion of multi-scale features to preserve spatial information based on strides convolutional with max-pooling layers and achieved better segmentation performance as compared to UNet with significantly reduced parameter space.

In the second category, a pyramid parsing module with either pyramid input analysis (PIA) or pyramid feature analysis (PFA) uses to capture multiscale features within the same convolutional level. PSP-Net [38] proposed spatial pyramid pooling to convolutional feature maps for pyramid feature analysis, Deeplab [39] and CE-Net [40] use parallel atrous convolution to extract multiscale features for semantic segmentation. These multiscale features used diverse effective receptive fields with a concatenation approach for progression feature representation ability of context information. PIA used features from input images of various sizes to create an image pyramid at multiscale. Kamnitsas et al. [41] and Farabet et al. [42] proposed the PIA method to extract effective features at multiple scales from input images for segmentation. The skip-net and pyramid parsing module can be helpful for image segmentation and could be used in combined form to enhance the segmentation performance in multi-organ segmentation. Some recent works like [43,44] integrate features from pyramid input images to the U-Net structure.

However, since those features are at different semantic abstraction levels, fusing those features from different scales may cause the problem of the semantic gap. Thus, those networks fail to mitigate the multi-scale context information, by only partially utilizing the pyramid shape of U-Net.

3. Material and methods

3.1. Datasets

In this study, there are two publicly available datasets based on COVID19 segmentation based on disease and lungs and we are comparing our methods on two other non-COVID-based datasets that have lung parts.

3.1.1. COVID 19 lesion dataset

The COVID-19 segmentation dataset [45] consists of the right lung, left lung, and COVID19 lesions. The dataset consists of 20 CT scans of patients and their respective annotated masks. The masks were created by junior annotators and were refined by senior radiologists having 5 years of experience. Finally, radiologists having 10 years' experience verified these annotations. On average, as CT scans have a good spatial resolution (250 slices), 400 min for delineating one CT scan volume were required.

3.1.2. Dataset based on COVID19 segmentation

COVID-19 CT images have been collected by the Italian Society of Medical and Interventional Radiology (SIRM). The dataset consists of 110 axial CT scans that are brought from 60 patients. Data were annotated by a trained radiologist with labels such as 1 = ground class opacification, 2 = consolidations, and 3 = pleural effusions. For our experiment, similarly to [46], a hundred CT images and masks have been used to perform the segmentation task.

3.1.3. StructSeg 2019 lung organ segmentation

StructSeg consists of various data modalities. 50 CT patients with lung cancer annotation can be used and these annotations are collected from one medical center. This dataset was first published in the MICCAI 2019 challenge. This dataset contains six numbers of classes such as left lung, right lung, heart, trachea, and spinal cord (https://structseg2019.grand-challenge.org). In this proposal, we only use the left and right lungs for validating the feasibility of the proposed model. The distribution of the dataset among training and testing samples is given below in Table 1.

3.1.4. NSCLC left and right lung segmentation

This dataset consists of 402 CT volumes and can be collected from the Cancer Imaging Archive on NSCLC Radiomics [47] platform. In the NSCLC dataset, annotations are given for the right lung and left lung for the 402 volumes as well as 78 cases are annotated for Pleural Effusion (PE).

3.1.5. Datasets preprocessing for training, testing, and validation

We have evaluated our proposed method on the different datasets and each dataset contains a different number of 2d images from the 3D volume. The input size of 2D slices is varied and the

detail of the input size of 2D images for each dataset is shown in Table 1. The total 2D images used for training, testing, and validation are shown in Table 1 for each dataset.

Three datasets are provided in 3D volumes in NIfTI (Neuroimaging Informatics Technology Initiative) compressed format. We have used NiBabel python-based library to convert the 3D volume into 2D images for each subject for 20 cases (covid), StructSeg2019, and NSCLC datasets. The 100 2D slices dataset was in 2D images jpeg format. The standard normalization method has been used as preprocessing steps for all datasets. The standard normalization technique that normalizes images using zero mean and unit standard deviation has been used to preprocess the input images. Each dataset has a different spatial resolution, the standard normalization method is working fine for our proposed method. Few slices contain empty lung pixels, we have trained our proposed model for all slices in each 3D volume for each patient in the training and validation dataset.

The image intensity values of all of the images were truncated to within the range of -384 to 384 HU to omit irrelevant information for the 20 cases (covid) cases dataset and for the StructSeg2019 dataset we have used the -512 to 512 HU range. For the other two datasets, we did not apply any window width and depth level. The input images into 256×256 for all datasets for training the proposed model and used linear interpolation method to get the equal input size for testing and validation of the proposed model.

3.2. Proposed methodology

3.2.1. Multiscale and multilevel feature extraction structure

In our proposed technique, the utilization of multiscale features at each pyramid level has been processed from the encoder to the decoder side tackles fully the problem of the semantic gap. To link the semantic gap caused by directly merging features from different scales, an equal convolutional depth block was introduced at each encoder side of the proposed solution. We assume that extracting and keeping multi-scale features through the model to gather hierarchical contextual information can expressively improve the segmentation performance.

The proposed method used fully fuses multi-scale context information and semantic similar features with one single network. The network performs spatial pyramid pooling on input and hierarchical abstract multi-scale features at each level imposed by a deep supervision mechanism. Unlike the classical U-net based methods, where the scale only reduces with the convolutional depth. The proposed approach has multi-scale features at each depth and therefore both global and local context information can be integrated to augment the extracted features. After going through one or more convolutional layers, the features are fused to have hierarchical structural information. The proposed multiscale approach extracted features and fused these features at each level went through the same number of convolutional layers. Same convolutional means, we have used equal convolutional depth block. It has been achieved using ResNet block as shown in maroon color in proposed Fig. 1. We have dealt with the problem of the semantic gap using an equal convolutional depth block. With the equal convolutional depth connections, all the fused features at each step are

Table 1
Datasets distribution for training, validation, and testing based on COVID19 and Non-COVID19.

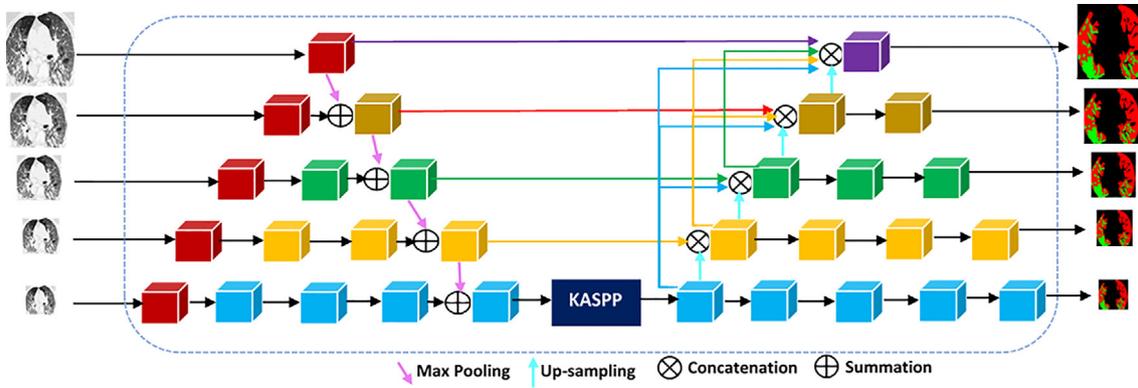| Datasets	Training cases	Training (2d slices)	Validation cases	Validation (2d slices)	Testing cases	Testing (2d slices)	Cross-Validation
20 cases (covid)	12	2452	4	560	4	508	5-fold
StructSeg2019	30	3807	10	465	10	503	5-fold
NSCLC	321	38,189	80	5139	81	5200	5-fold
100 cases (covid)	–	50	–	50	–	50	5-fold

Fig. 1. The proposed model is based on multiscale and multilevel feature extraction for COVID-19 and Non-COVID19 segmentation.

at the same semantic abstraction level to better exploit the pyramid shape of U-Net.

Moreover, the deep supervision method is applied to the outputs generated by the decoding path at a different scale. The spatial pyramid pooling has been performed to the ground truth segmentation to generate labels in all output scales during training for the proposed model. By applying weighted cross-entropy, the training loss is computed by using the corresponding output and ground truth segmentation at the same scale. Supervision layer can help relieve the problem of gradient vanishing in deep neural networks and learn deep-level features with hierarchical contexts. It also enforces the outputs in all scales to maintain structural information.

The loss function computed at each scale is shown in the following equation (1).

$$Loss_{scale} = -\frac{1}{L}\sum_{l=1}^{L}\frac{1}{N_l}\sum_{i=1}^{N_l}\sum_{c=0}^{C}P_{i,l}^{c}logpr_{i,l}^{c} \qquad (1)$$

where $pr_{i,l}^{c}$ is predicted probability of voxel i in class c from scale l. The $P_{i,l}^{c}$ ground truth labels in scale l and N_l a total number of voxels in scale l. The $P_{i,l}^{c}$ are the parameters in scale l for class c. The C represents the total number of classes and L represents the number of scale levels. (in our case 5). The output feature maps are represented as $feat_{i,l}^{v}$.

$$pr_{i,l}^{c} = \frac{exp(feat_{i,l}^{v})}{\sum_{v}exp(feat_{i,l}^{v})} \qquad (2)$$

The feature maps at each encoder block is represented as.

$$feat_{l,v} = \begin{cases} conv(feat_{l,v-1}), v < l \\ Conv(feat_{l,v-1}) + MaxPool(feat_{l-1,v-1}), v = l \end{cases} \qquad (3)$$

The feature maps at each decoder block can be represented as.

$$feat_{l,v} = \begin{cases} U_1 = upsample(conv(feat_{l,v-1})) \\ U_{11} = concat(U_1, conv(feat_{l,v})), v+l = 2L \\ U_2 = upsample(U_{11}) \\ U_{22} = concat(U_1, U_2, conv(feat_{l,v})), v+l = 2L \\ U_3 = upsample(U_{22}) \\ U_{33} = concat(U_1, U_2, U_3, conv(feat_{l,v})), v+l = 2L \\ U_4 = upsample(U_{33}) \\ U_{44} = concat(U_1, U_2, U_3, U_4, conv(feat_{l,v})), v+l = 2L \\ conv(feat_{l,v-1}), v+l > 2L \end{cases}$$
$$\qquad (4)$$

The l is the level or scale at each encoder block. $feat_{l,v}$ input features v_{th} convolutional at scale l_{th} and L is the scale number of inputs. The features are concatenated at each scale in our proposed model. Similarly, features are presented at the decoder side with a concatenation layer. The U_1 is represented by the lower unsampled layer at the bottom of the decoder side and U_{44} is represented concatenation of features at the top level at decoder side is shown in Fig. 1.

Features information is obtained at different scales from deep supervision blocks that further fuse these features together at each scale to get an accurate segmentation map. We designed a module that gets adaptively the contextual information in different scales to obtain relative importance features information at each scale and also automatically fuses the score from each map. We have used the attention module on each pyramid input feature map to get the pyramid output features using a shared convolutional block. We also used Global average pooling (gavpool) and global max-pooling (gmaxpool) to squeeze into a single channel feature and to extract the global certainty score of the predictions at each scale.

The output at each scale can be obtained from the proposed model using *attentConv* block and then using these weights in Global average pooling (gavpool) and global max-pooling (gmaxpool). The values from different scales are then concatenated to feed into a softmax layer to get the corresponding weight for each scale. The weights W_l reflects the importance of feature at scale l.

$$a_{l_1}, b_{l_2}, c_{l_3}, d_{l_4}, e_{l_5} = ProposedModel()$$

$$W_1 = attentConv(a_{l_1})$$

$$W_2 = attentConv(b_{l_2})$$

$$W_3 = attentConv(c_{l_3})$$

$$W_4 = attentConv(d_{l_4})$$

$$W_5 = attentConv(e_{l_5}) \qquad (5)$$

The global score S_l at each scale, l can be computed by summing up the gavpool and gmaxpool.

$$S_l = \sum_{l=1}^{L}(\text{gavpool}(W_l) + \text{gmaxpool}(W_l)), l = 1, 2, 3, 4, 5 \qquad (6)$$

The weights W_l reflect the importance of feature at scale l. The scale weights are computed from equation (7).

$$W_l = \frac{\exp(S_l)}{\sum_l \exp(S_l)} \tag{7}$$

The softmax layer is applied to get the final segmentation map (SM).

$$SM = \frac{\exp(\sum_{l=1}^{L} W_l feat_l)}{\sum_c \exp(\sum_{l=1}^{L} W_l feat_l)} \tag{8}$$

We proposed a spatial pyramid pooling module with some proposed blocks to extract multiscale features at a different level of the encoder side and fused these features within a convolutional layer-based module to reduce the semantic gap and also used these features at multiple levels at the decoder side.

At the encoder side, we concatenate the features with decoder feature maps at every decoder layer and carry this information from the previous block to the next block as well. We have calculated the loss at each decoder block and adaptively accumulated all losses from all decoder blocks to compute the total loss during the training of our proposed model.

3.2.2. Basic units of proposed encoders

In the encoder block, three layers have been used with different activation functions. In the first layer. The expansion layer consisted of 1x1Conv, BN, and ReLU. The second layer contains a depth-wise layer with filter size 3×3, BN, ReLU activation, and the third layer is called the projection block consisting of 1x1Conv and BN layers. In the expansion layer, input data projects with a higher number of dimensions (channels) into a tensor with a lower number of dimensions. The 1×1 convolution is used to expand the number of channels in the data before it goes into the depth-wise convolution. Hence, this expansion layer always has more output channels than input channels. Exactly how much the data gets expanded is depends on the expansion factor. This is one of the hyperparameters that we used based on experimenting with different encoder blocks. The five (5) expansion factors produced better performance in our experiment. So, the input and the output of the proposed encoder block are low-dimensional feature maps, while the filtering step depth-wise that happens inside the block is done on a high-dimensional feature map. The residual connection works just like in ResNet and exists to help with the flow of gradients through the network. Each layer has batch normalization and the ReLU activation function. However, the output of the projection layer does not have an activation function applied to it. Since this layer produces low-dimensional data and non-linearity function may destroy the information. The expansion layer acts as a decompressor that first restores the data to its full form, then the depth-wise layer performs filtering on it that is important at this stage of the encoder, and finally, the projection layer compresses the data to make it small again. The expansions and projections are done using convolutional layers with learnable parameters, so the encoder can learn how to best decompress the data at each stage on the encoder side. The encoder block takes as an input a low-dimensional compressed representation which is first expanded to high dimension and filtered with a lightweight depth-wise convolution. Features are subsequently projected back to a low-dimensional representation with a linear 1×1 convolution. The layer structure is shown in Fig. 2 (b). The swish activation function is used after batch normalization in the expansion layer. Similarly, the activation is used after the depth-wise layer. The depth-wise separable convolutional layer is the building block of many deep neural learning models [48–50] and this layer is in our proposed model as compared to the standard convolutional layer. The depth-wise separable convolutional layer consists of two layers. In the first layer, it performs lightweight filtering by applying a single convolutional filter per input channel. The second layer is a 1×1 convolution, called a pointwise convolution, which

is responsible for building new features through computing linear combinations of the input channels. Depth-wise convolution produced an efficient solution as compared to standard convolutional layers. We have used a depth-wise convolutional layer and SE block on the decoder side to enhance the performance of the proposed model. The basic structure of the EDP (expansion, depth-wise, and projection) module proposed for encoder and decoder is shown in Fig. 2 (a).

3.2.3. Basic units of proposed decoders

On the decoder side, EDP (expansion, depth-wise, and projection layer block) have been proposed. 1×1Conv convolutional layer used to handle feature maps from expansion and projection layer block. The main objective of this decoder module is to used varying depth feature maps from lower to higher dimensions. We have used depth-wise convolutional layer and SE block before projection layer that increased little performance in our proposed model. In the encoder block, the expansion layer is based on a 1×1Conv, depth-wise layer with filter size 3×3, squeeze, and excitation block (SE block), and projection layer based on 1x1Conv has to be used. The expansion layer increases the number of feature maps from the input of the decoder block and then passed the high number of feature maps in depth wise convolutional layer and then the projection layer resumes the feature maps back to the input number of feature maps. The number of feature maps is expanded and then the depth-wise 2d layer is applied and then the feature maps are reduced back to the original in the projection layer. The complete layer structure for the decoder is shown in Fig. 2 (c).

In simple U-Net model-based encoder and decoder blocks may have limited effective feature learning capacity for complicated images task like multiclass segmentation. This limitation could be overcome by optimizing the network that has broadened parameter space to learn more representative features. The proposed encoder and decoder showed better and more effective feature learning capacity. The proposed encoder and decoder blocks produced a better performance as compared to normal standard convolutional layer-based encoder and decoder. The dynamics feature an expansion and reduction approach that has been used in a single encoder and decoder block with some activation layers such as swish activation that produced a better performance with a smaller number of parameters space. The expansion layer encodes the model's intermediate inputs and outputs while the inner layer encapsulates the model's ability to transform from lower-level concepts such as pixels to higher-level descriptors such as image categories. Finally, as with traditional residual connections, shortcuts enable faster training and better accuracy.

### 3.2.4. Proposed KASPP module

Inspired by Deeplab [42] and CE-Net [43], the proposed atrous spatial pyramid pooling module was introduced at the bottom of the proposed model. This proposed ASPP module captured features that are fused with features extracted from the spatial pyramid pooling module at the bottom of the proposed model is shown in Fig. 1. The 3×3 kernel is shared with atrous convolutional layers with different dilation rates. In this work, we have extended KSAC based ASPP module. The proposed KSAC based ASPP (later noted KASPP) module captures features from low levels as well as features from different down-sampled layers to obtain texture and position information from encoder side feature maps. Kernels with small atrous rates in convolutional layer branches would be able to learn detailed information and effectively handle small semantic classes well. A kernel with big atrous rates can extract features with large receptive fields and may miss detailed information, the generalization about kernel in atrous convolutional layer branches is limited and the number of parameters increases linearly using the parallel branches with separate kernels. In Fig. 3,
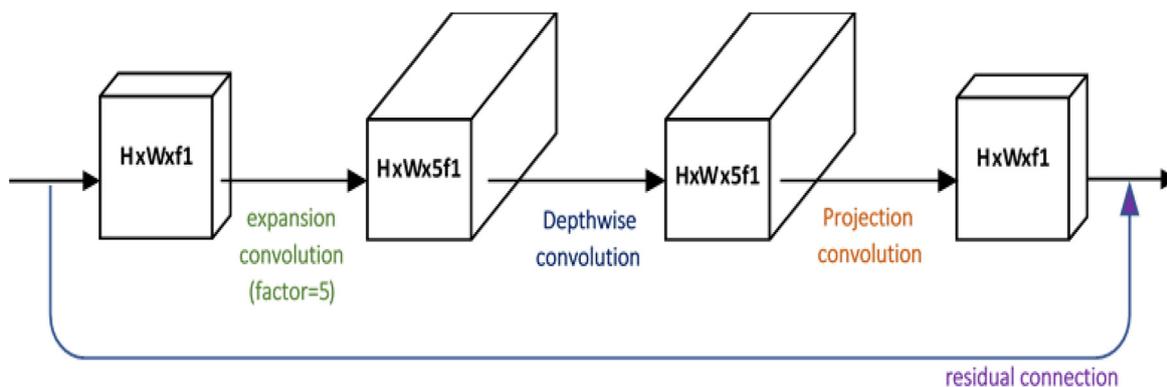
(a) proposed EDP (expansion, depth-wise, and projection) block
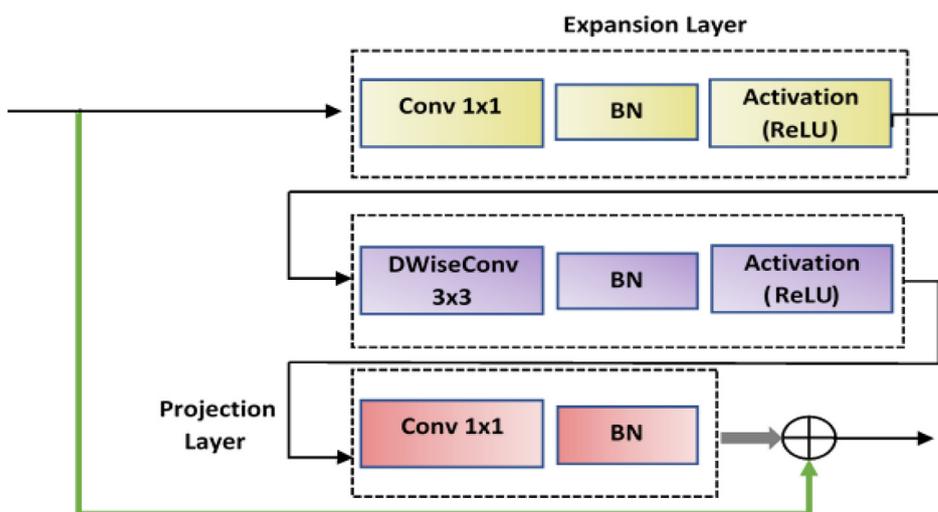
(b) proposed encoder block
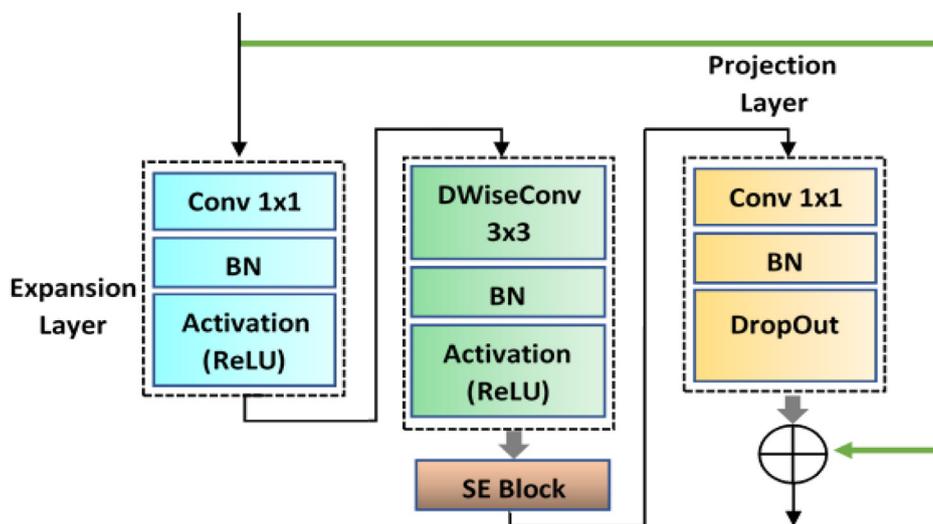
(c) proposed decoder block

Fig. 2. Proposed encoder and decoder module used in the proposed segmentation model.

as one can see, a 3×3 kernel is shared by multiple parallel branches at different atrous rates and captured features with different receptive fields. Without sharing kernel, in a simple atrous spatial pyra-

mid pooling layer, the number of parameters increases using parallel convolutional layer branches with different atrous rates. The kernel sharing approach improved the performances in terms
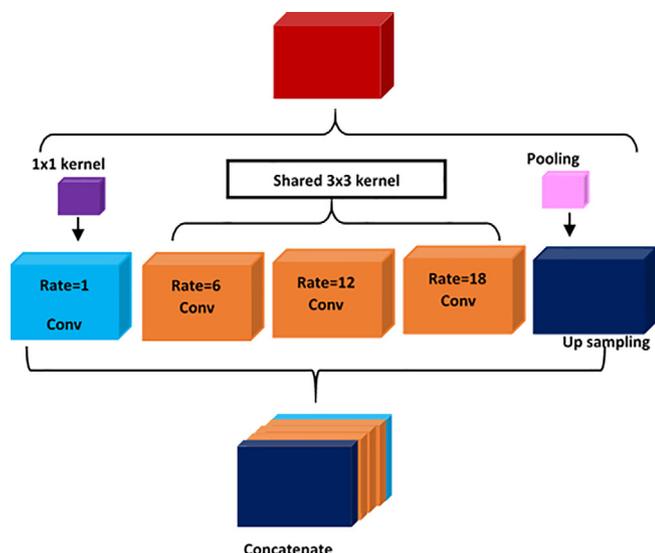
Fig. 3. The proposed kernel atrous spatial pyramid pooling layer (KASSP).

objects. The feature maps produced by the shared KASPP layer are comprehensive, expressive, and discriminative as compared to features produced by simple ASPP. This strategy increases the efficiency of features map reuse and fuses feature information at the downsampling level. In our proposed model, the features are extracted from three levels that enable the model to get benefit from the multi-scale transformation of high-level semantic information and low-level information of position and texture. In our approach, fourth downsampling layers information are passed to the proposed KASPP module to guarantee improved cross-level feature connection and complementary between cross-level information. The proposed KASPP can be used at any level of the encoder side. We have tried this module at each encoder block, however, it produced better performance when we placed it at the bottom layer of the proposed model. The KASPP module handles challenging shape variations of COVID-19 infection areas.

3.3. Evaluation metrics

Dice similarity coefficients (DSC) and Normalized surface Dice (NSD) [51] were used to measure the performance between the proposed and existing deep learning models. DSC is used to estimate the similarity between predicted and ground truth segmentation maps. NSD is used to evaluate the closeness of the boundary between segmentation and ground truth surfaces. Higher DSC and NSD mean better segmentation performance and 100% mean perfect segmentation. Moreover, Hausdorff distance and normal-

of less computational time (using a smaller number of parameters) and also increases the segmentation performance. The shared kernel approach enhanced the generalization capability by enhancing the learning for detailed features for small objects locally and globally increasing semantic information of rich features for large

Table 2
Average Dice similarity coefficients (DSC) of proposed and existing deep learning models.

Datasets	Methods	DSC Left Lung	DSC Right Lung	DSC Covid19 Infection	DSC Average
NSCLC dataset	DeepLabV3	0.927	0.944	–	0.935
	UNet	0.917	0.938	–	0.928
	Proposed model	0.945	0.959	–	0.952
COVID19 dataset	DeepLabV3	0.944	0.936	0.6965	0.859
	UNet	0.892	0.905	0.6707	0.823
	Proposed model	0.953	0.981	0.71994	0.884
StructSeg2019 dataset	DeepLabV3	0.900	0.947	–	0.924
	UNet	0.898	0.947	–	0.923
	Proposed model	0.981	0.965	–	0.973

Table 3
HD of proposed and existing models for different datasets.

Datasets	Methods	HD Left Lung	HD Right Lung	HD Covid19 Infection	Average
NSCLC dataset	DeepLabV3	13.78	15.33	–	14.56
	UNet	14.56	19.17	–	16.87
	Proposed model	12.91	14.72	–	13.81
COVID19 dataset	DeepLabV3	25.82	25.94	32.23	28.00
	UNet	28.05	27.02	33.15	29.41
	Proposed model	15.40	17.30	25.32	19.34
StructSeg2019 dataset	DeepLabV3	21.48	18.90	–	21.68
	UNet	23.53	19.60	–	23.61
	Proposed model	17.23	18.11	–	18.87

Table 4
Normalized average volume difference (NAVD) for proposed and existing deep learning models using different COVID19 and NON-COVID19 datasets.

| Datasets	Methods	NAVD Left Lung	NAVD Right Lung	NAVD Covid19 Infection	NAVD Average
NSCLC dataset	DeepLabV3	0.124	0.097	–	0.111
	UNet	0.127	0.092	–	0.110
	Proposed model	0.110	0.093	–	0.101
COVID19 dataset	DeepLabV3	0.294	0.095	0.594	0.361
	UNet	0.387	0.165	0.484	0.339
	Proposed model	0.273	0.098	0.424	0.278
StructSeg2019 dataset	DeepLabV3	0.050	0.037	–	0.043
	UNet	0.065	0.030	–	0.047
	Proposed model	0.046	0.032	–	0.039
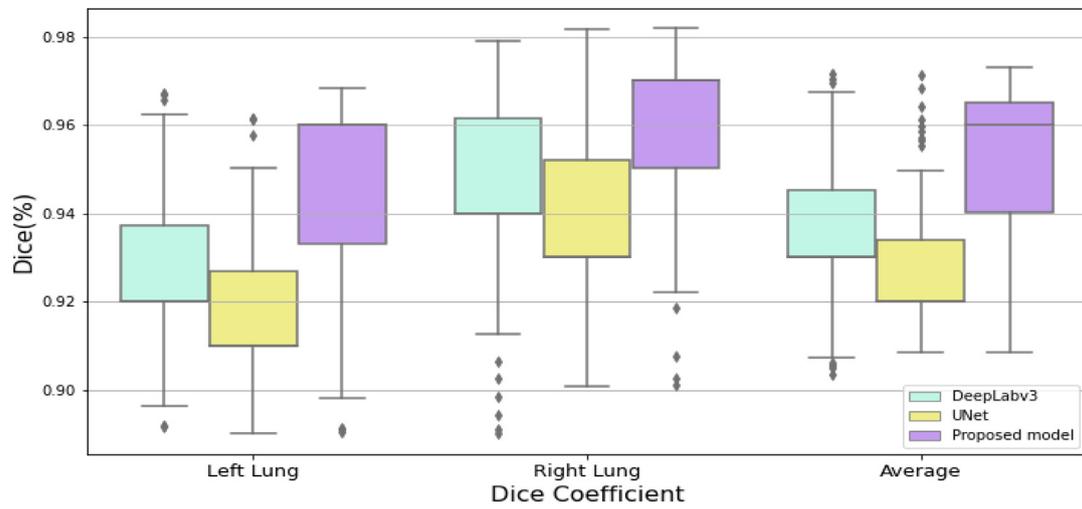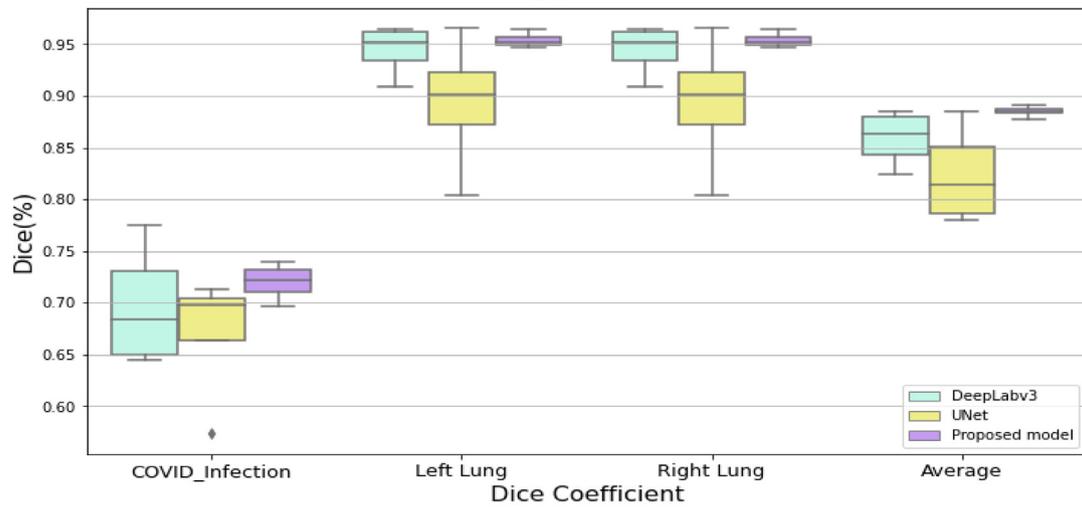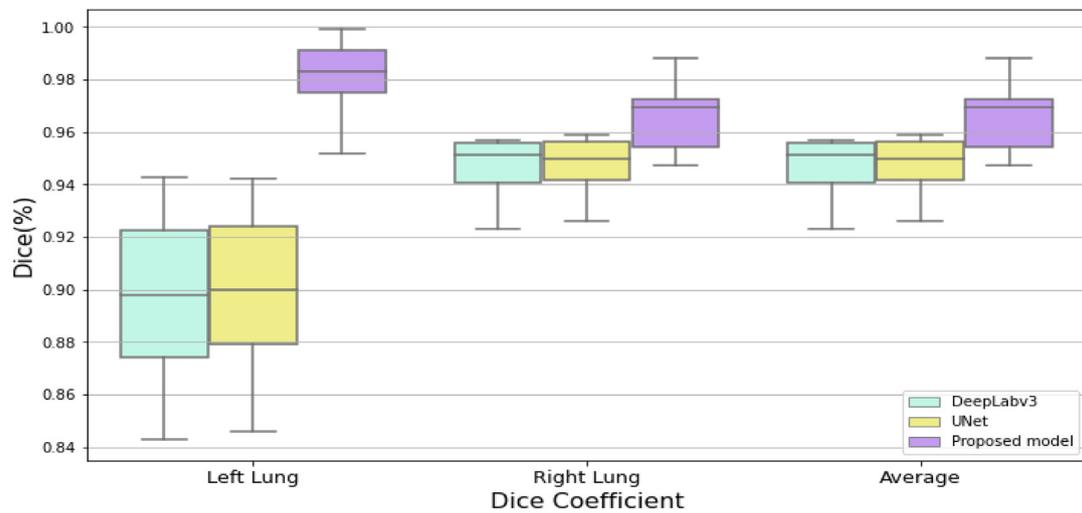
Fig. 4. The dice coefficients for proposed and existing deep learning models using test cases for all datasets, (a) NSCLC dataset, (b) COVID19 dataset, (c) StructSeg2019 dataset.
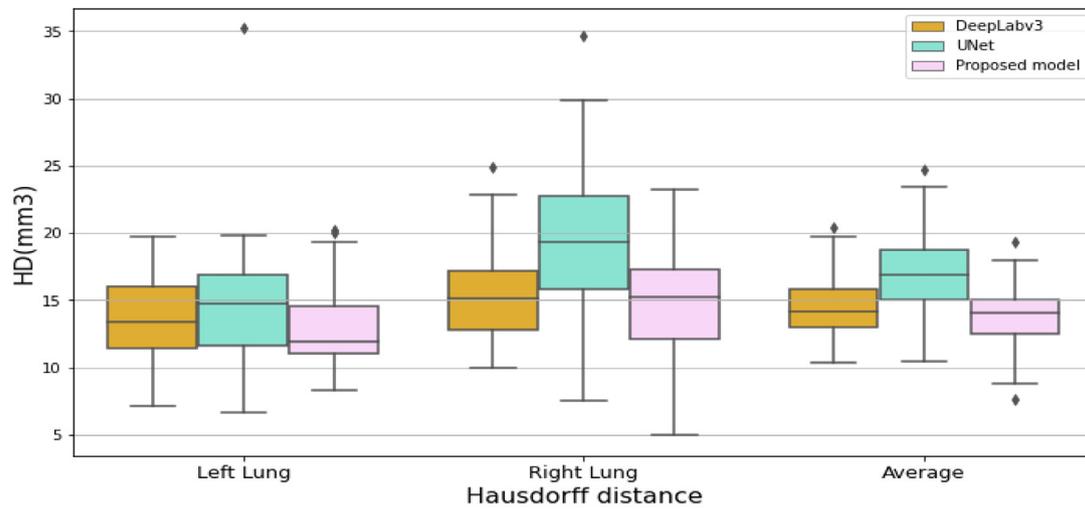
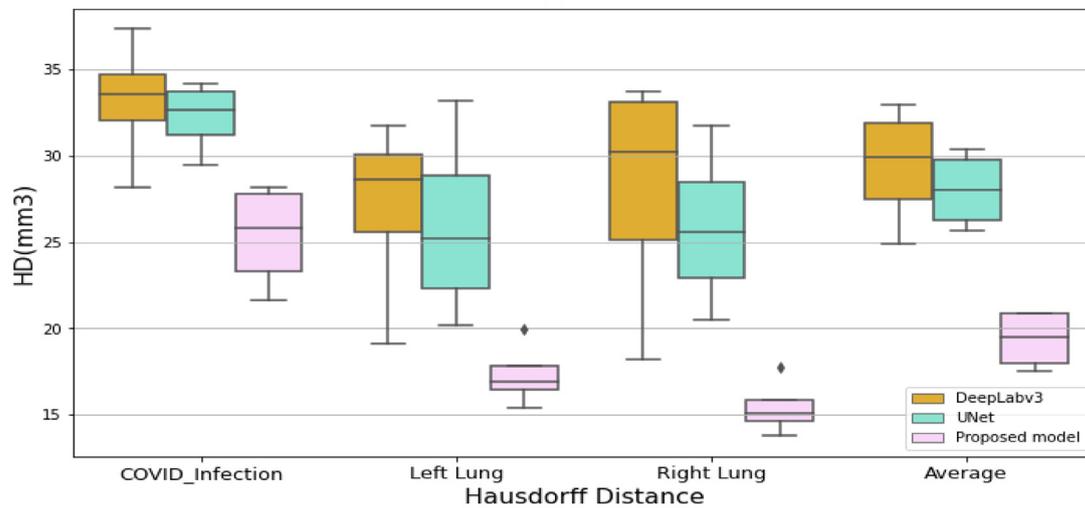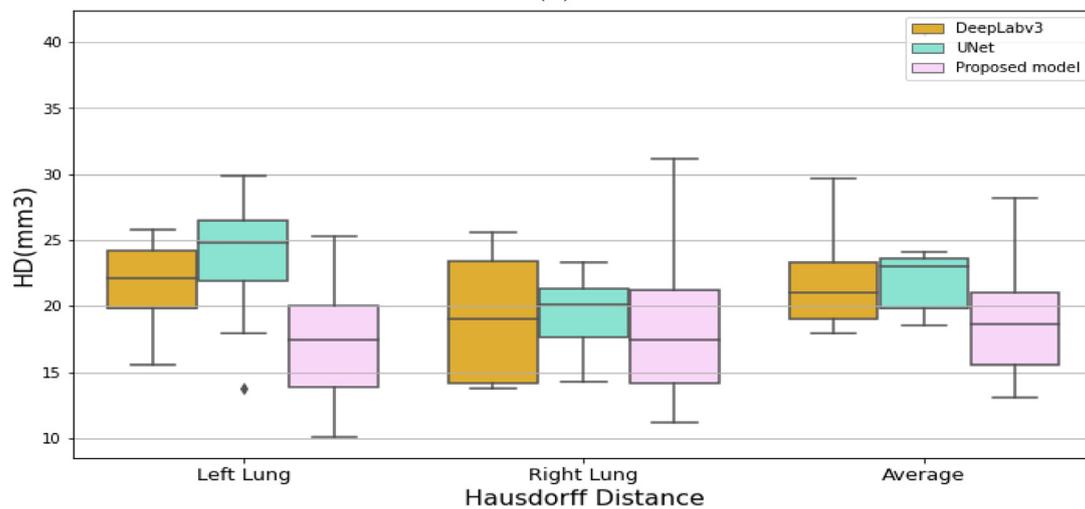**Fig. 5.** The HD for proposed and existing deep learning models using test cases for all datasets, (a) NSCLC dataset, (b) COVID19 dataset, (c) StructSeg2019 dataset.
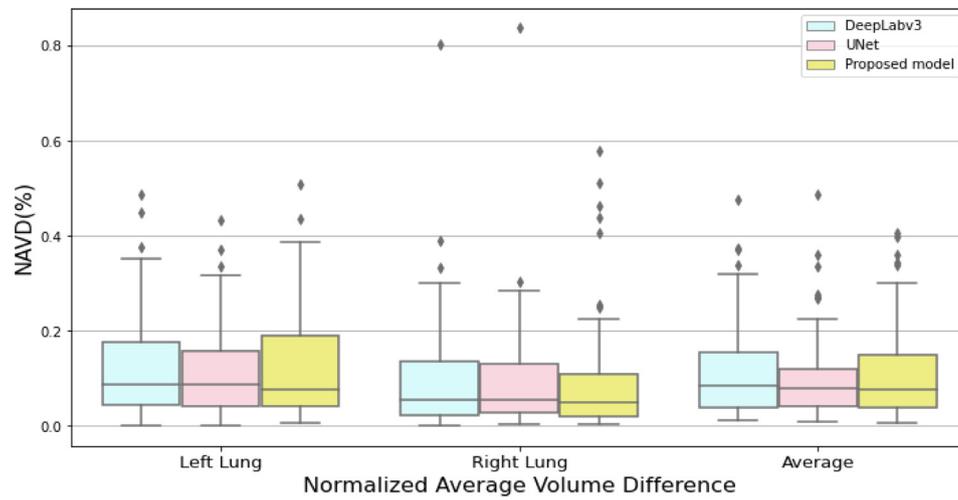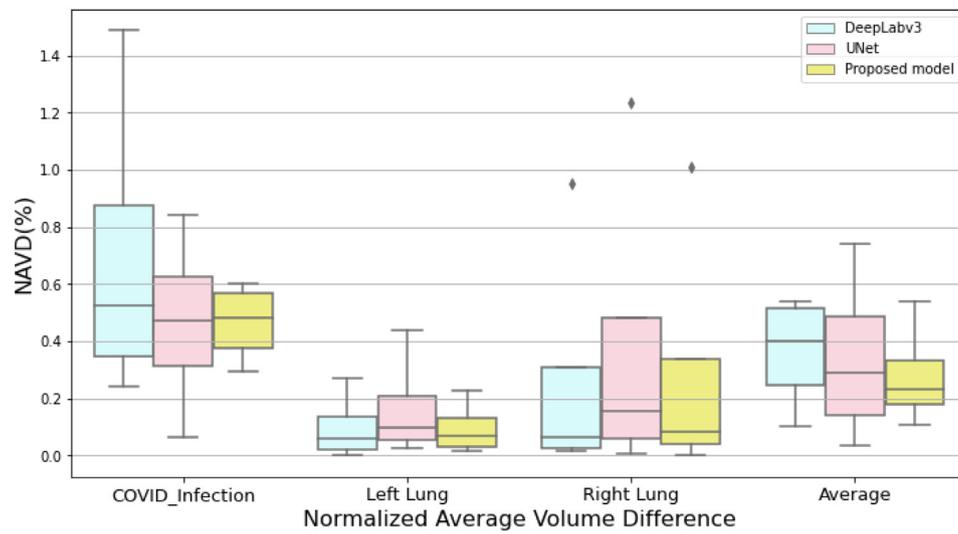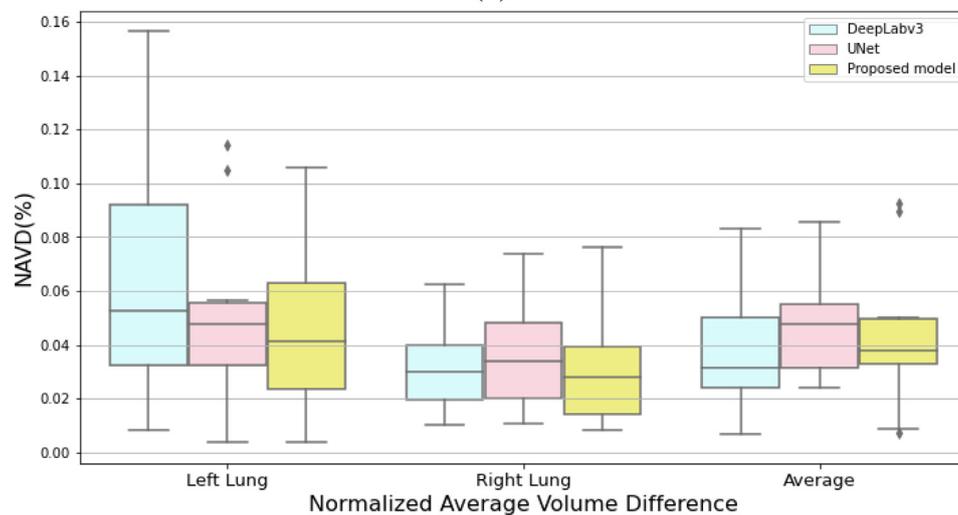
Fig. 6. The NAVD for proposed and existing deep learning models using test cases for all datasets, (a) NSCLC dataset, (b) COVID19 dataset, (c) StructSeg2019 dataset.
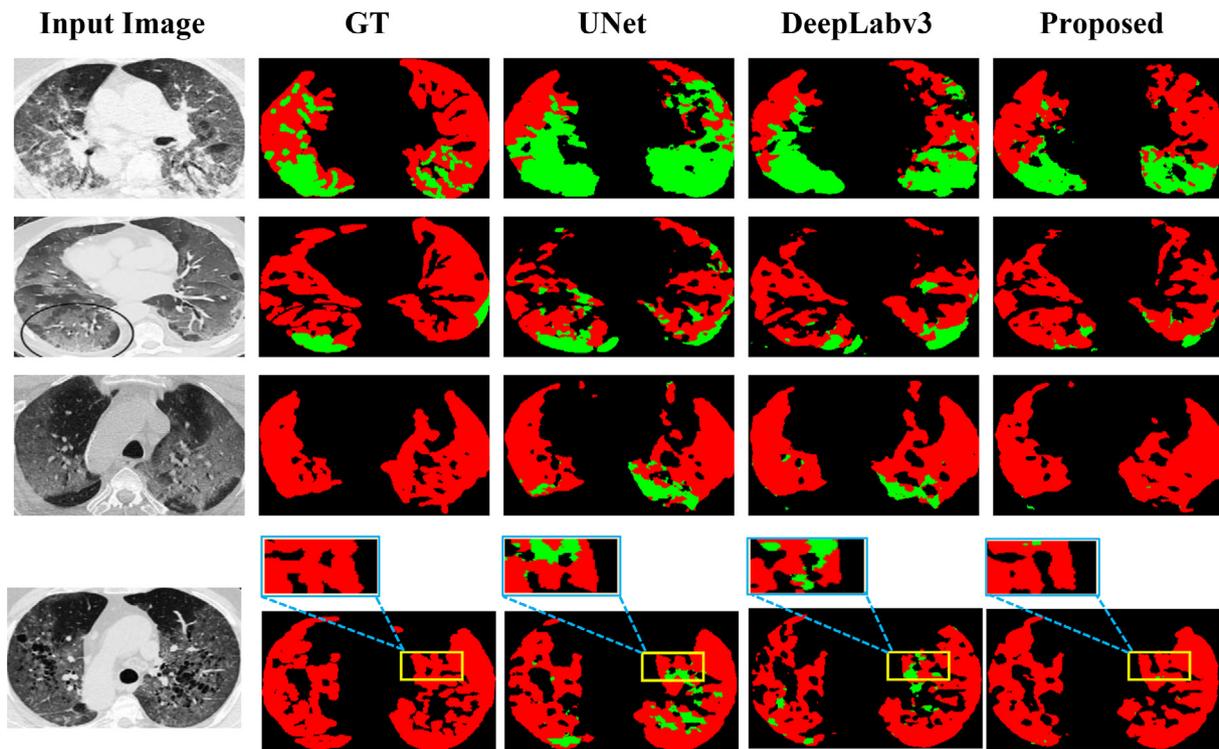
Fig. 7. The visualization of the segmentation map is based on the proposed and existing deep learning model for COVID19 100 case dataset.
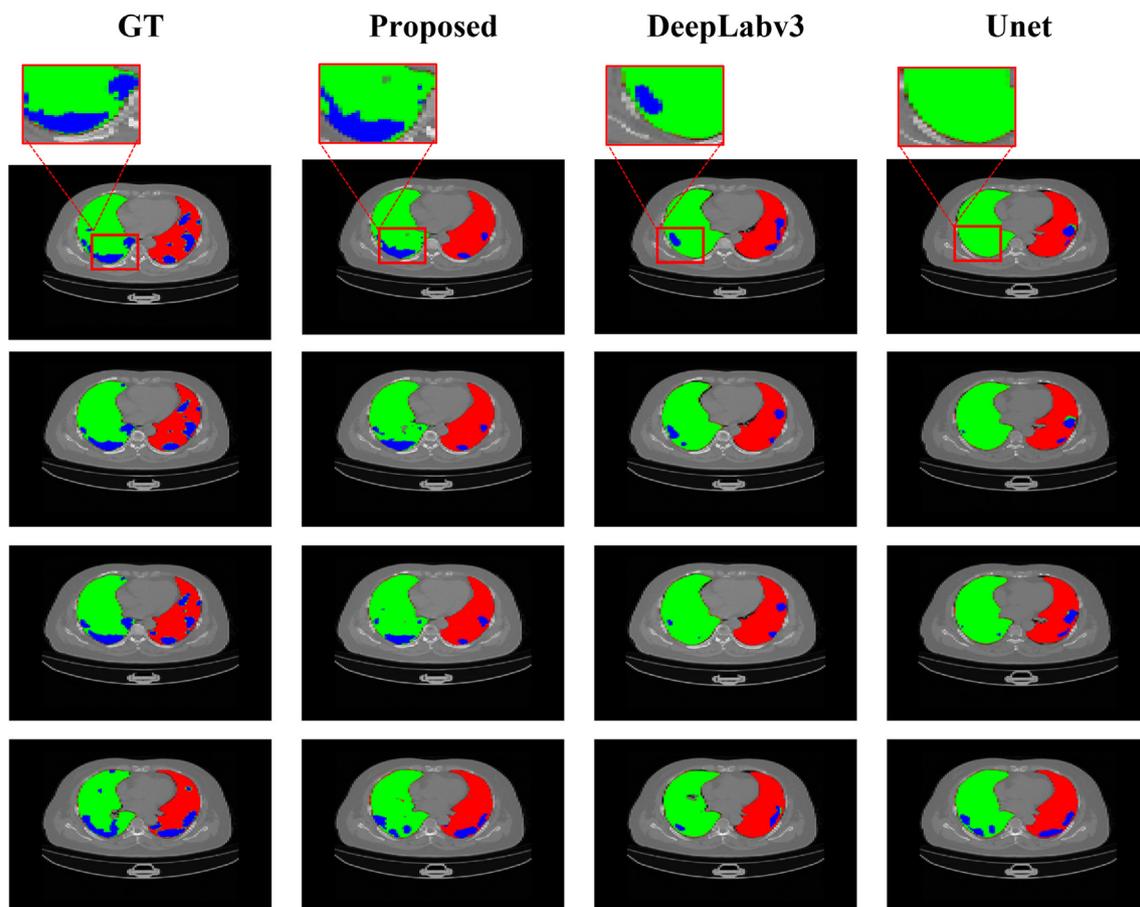
Fig. 8. The ground truth and predicted segmentation mask using the COVID19 dataset. The first row shows slice 1 and the second row shows slice2 and so on. Green color and red color represent the left and right lungs and the blue color denotes infection disease pixels on the lungs in the COVID19 dataset.

ized volume differences are also used to measure the quality of the predicted segmentation map.

Absolute volume difference (AVD) between predicted volumes measured from manual (V_M) and predicted (V_A) segmentations and their normalized value (NAVD) were reported as volume-based metrics.

$$NAVD = \frac{|V_A - V_M|}{V_M} \qquad (9)$$

3.4. Implementation of model

The proposed model was implemented using the Python-based PyTorch framework. The proposed model was trained using Adam optimizer for optimization at a learning rate of 0.0001. Batch size 16 was used for training the proposed 2D model, and the number of epochs was set to 200 for all datasets. The training was performed using the NVIDIA 100 GT T machine equipped with two GPUs, each having a 12 GB GPU memory. The training required 2 h.

4. Simulation results

80 percent of the dataset was used for training and 20 percent for testing and validation for all datasets. The results obtained from the proposed models are compared with standard UNet and DeeplabV3 models using COVID19 datasets and validated on two datasets that belong to NON-COVID19 datasets. The Dice similarity coefficients (DSC) have been computed for the proposed and exist-

ing deep learning models such as UNet and DeeplabV3. The proposed model produces better dice coefficients for all datasets as compared to the existing deep learning model as shown in Table 2. The Hausdorff distance (HD) has been computed for all datasets using our proposed and existing deep learning models as shown in Table 3. Similarly, the Normalized average volume difference has been computed based on the proposed and the existing model as shown in Table 4.

DSC box plots are shown in Fig. 4. Higher DSCs show that our proposed model produces a better performance as compared to existing deep learning models.

HD has been computed for test cases using the proposed and existing deep learning models, results are shown in Fig. 5. The lower the HD distance, the better the predicted segmentation map. Our proposed model produced lower HD as compared to the existing deep learning models for all datasets.

The normalized average volume difference has been computed based on the proposed and existing deep learning models and reported in Fig. 6. Our proposed model produced optimal performance in terms of NAVD for all datasets.

4.1. Segmentation visualization results

Below are some visualizations of some samples to see the predicted segmentation maps using the proposed and existing state-of-the-art deep learning models for the different datasets.

In all cases, as demonstrated by the quantitative results presented before, the proposed model produced fewer false-positive pixels in the segmentation maps. Fig. 7 presents the results
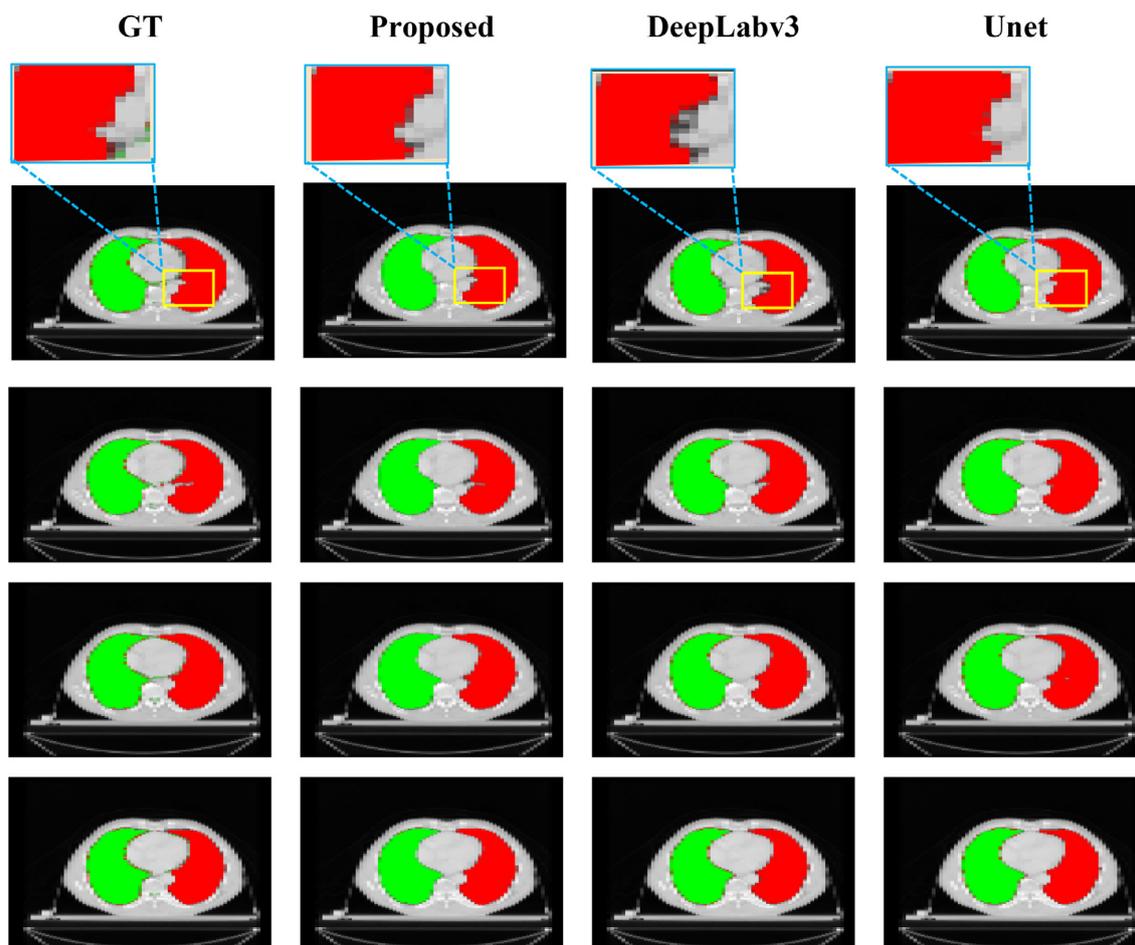

Fig. 9. The ground truth and predicted segmentation mask using the structseg2019 dataset. The first row shows slice 1 and the second row shows slice2 and so on.

obtained on the COVID19 100 case dataset. The visualization capability of the proposed model can be shown in the zoom part of the last row in Fig. 7. The proposed model produced very less wrong pixel values for covid infection as compared to DeepLabV3 and Unet model. The green color showed the infection pixel value and in this particular slice, there is no covid infection pixel that exists that is shown in the ground-truth segmentation mask. The

DeepLabV3 and UNet predicted wrong covid infection pixels instead of predicted lung pixel values.

Fig. 8 shows some results from the COVID19 20 dataset. The first, second, and fourth rows represent a different number of slices that belong to a single case.

The first row in Fig. 8 represented the visualization of ground-truth, proposed model, DeepLabV3, and 2D Unet in zoom area.
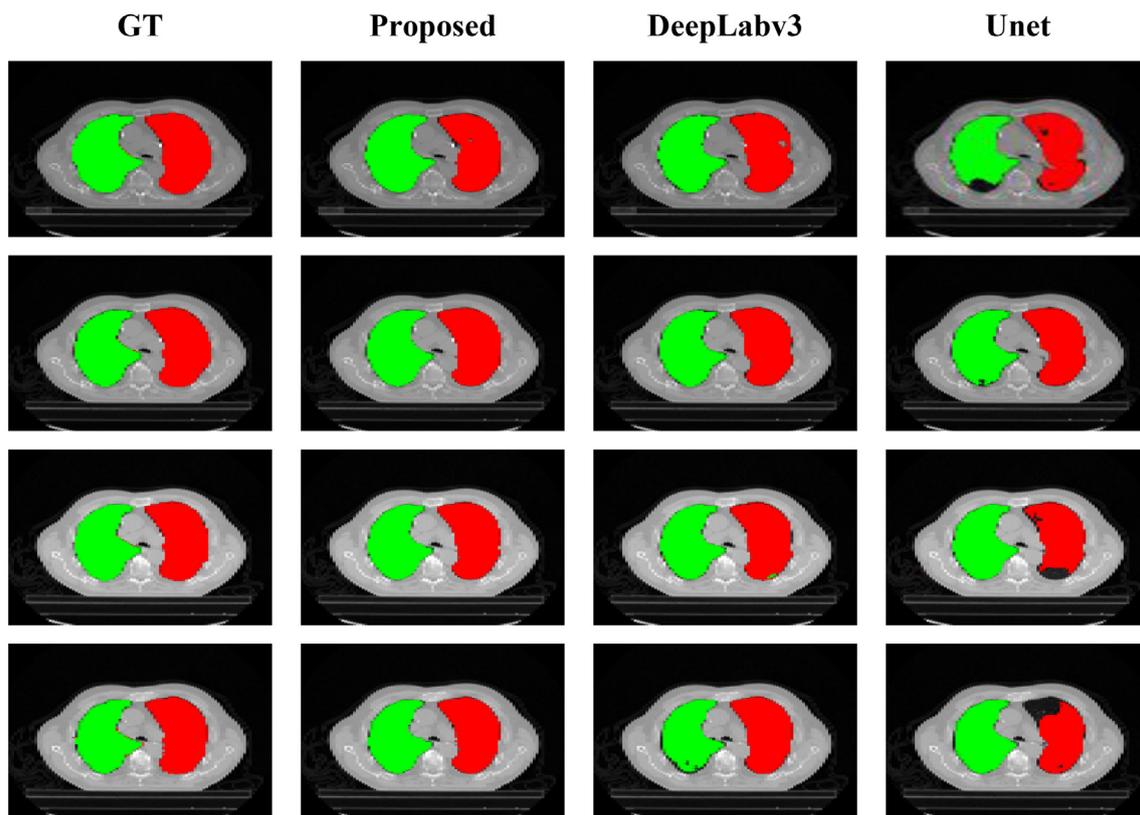

Fig. 10. The ground truth and predicted segmentation mask using the NSCLC dataset. The first row shows slice 1 and the second row shows slice2 and so on.

Table 5
The comparison of the proposed model with state-of-the-art models on different publicly available datasets (COVID19 and NON-COVID19).

Methods	Dataset	DSC Left Lung	DSC Right Lung	Infection (COVID-19-CT-Seg)
2D UNet Jun Ma [30]	20 Cases	*95.1 ± 7.9*	95.6 ± 7.4	60.9 ± 24.5
–	Task3-StructSeg	*96.3 ± 7.6*	96.7 ± 7.0	–
–	Task3-NSCLC	*92.5 ± 17.3*	93.3 ± 15.9	–
3D UNet Jun Ma [30]	20 Cases	85.8 ± 10.5	87.9 ± 9.3	67.3 ± 22.3
–	Task3-StructSeg	97.3 ± 2.1	**97.7 ± 2.1**	–
–	Task3-NSCLC	93.5 ± 5.4	94.0 ± 5.3	–
Proposed 2D model	20 Cases	**95.3 ± 0.1**	**98.1 ± 0.6**	**71.9 ± 01.5**
–	Task3-StructSeg	**98.5 ± 2.1**	96.5 ± 3.3	**–**
–	Task3-NSCLC	**94.5 ± 2.1**	**95.2 ± 2.4**	**–**

Table 6
The DSC for proposed and existing method using 5 cross-validation method. Best results are in bold.

Dataset	Methods	DSC Fold1	DSC Fold2	DSC Fold3	DSC Fold4	DSC Fold5
NSCLC	**DeepLabV3**	0.921	0.919	**0.935**	0.930	0.931
	Unet	0.919	0.909	**0.928**	0.921	0.922
	Proposed model	0.951	0.941	**0.952**	0.949	0.950
COVID19	**DeepLabV3**	0.841	**0.859**	0.839	0.850	0.853
	Unet	0.812	**0.823**	0.820	0.821	0.820
	Proposed model	0.880	**0.884**	0.879	0.881	0.884
StructSeg2019	**DeepLabV3**	0.915	0.923	0.921	**0.924**	0.922
	Unet	0.922	0.920	0.919	**0.923**	0.918
	Proposed model	0.966	0.970	0.970	**0.973**	0.971

Table 7
Performance comparison of the proposed method with different configurations. The best results are in bold.

Model	100 CT images Covid	20 cases Covid	Structseg 2019	NSCLC dataset
Base	0.862 ± 1.81	0.823 ± 3.71	0.943 ± 0.92	0.920 ± 1.49
Base + SE	0.878 ± 0.89	0.839 ± 2.43	0.964 ± 0.83	0.917 ± 0.64
Base + Multiscale	0.899 ± 0.85	0.876 ± 2.34	0.965 ± 0.71	0.936 ± 0.64
Base + KASSP	0.886 ± 1.39	0.862 ± 2.51	0.961 ± 0.78	0.942 ± 0.64
Proposed (Base + SE Multiscale + KASSP)	**0.901 ± 0.91**	**0.884 ± 2.26**	**0.975 ± 0.64**	**0.950 ± 0.64**

Table 8
Trained proposed model on 20 cases and validated on various other datasets.

| Model validation	Dice Similarity coefficients
100 slices dataset	0.861 ± 1.18
Structseg2019 dataset	0.946 ± 1.39
NSCLC dataset	0.905 ± 0.89

The visualization segmentation map in the zoom area clearly showed that our proposed model predicts more pixels in covid infection (green color) as compared to DeepLabV3 and 2d Unet model. The 2D Unet model failed to predict pixel values for covid infection is shown in the zoom area in Fig. 8 (last row, last column).

Fig. 9 shows the left and right lungs of a single case with different slices using the structseg2019 dataset. The green color represented the left and the red color represented the right lung. The visualization shows better segmentation for the proposed model. In Fig. 9 the zoom area represents in the first row shows that our proposed model predicts more pixels in the right lung as compared to other models such as DeepLabv3 and 2D Unet model.

Similarly, Fig. 10 shows the left and right lungs of a single case with different slices using the NSCLC dataset. The green color rep-

(a) COVID19 dataset

(b) StructSeg2019 dataset
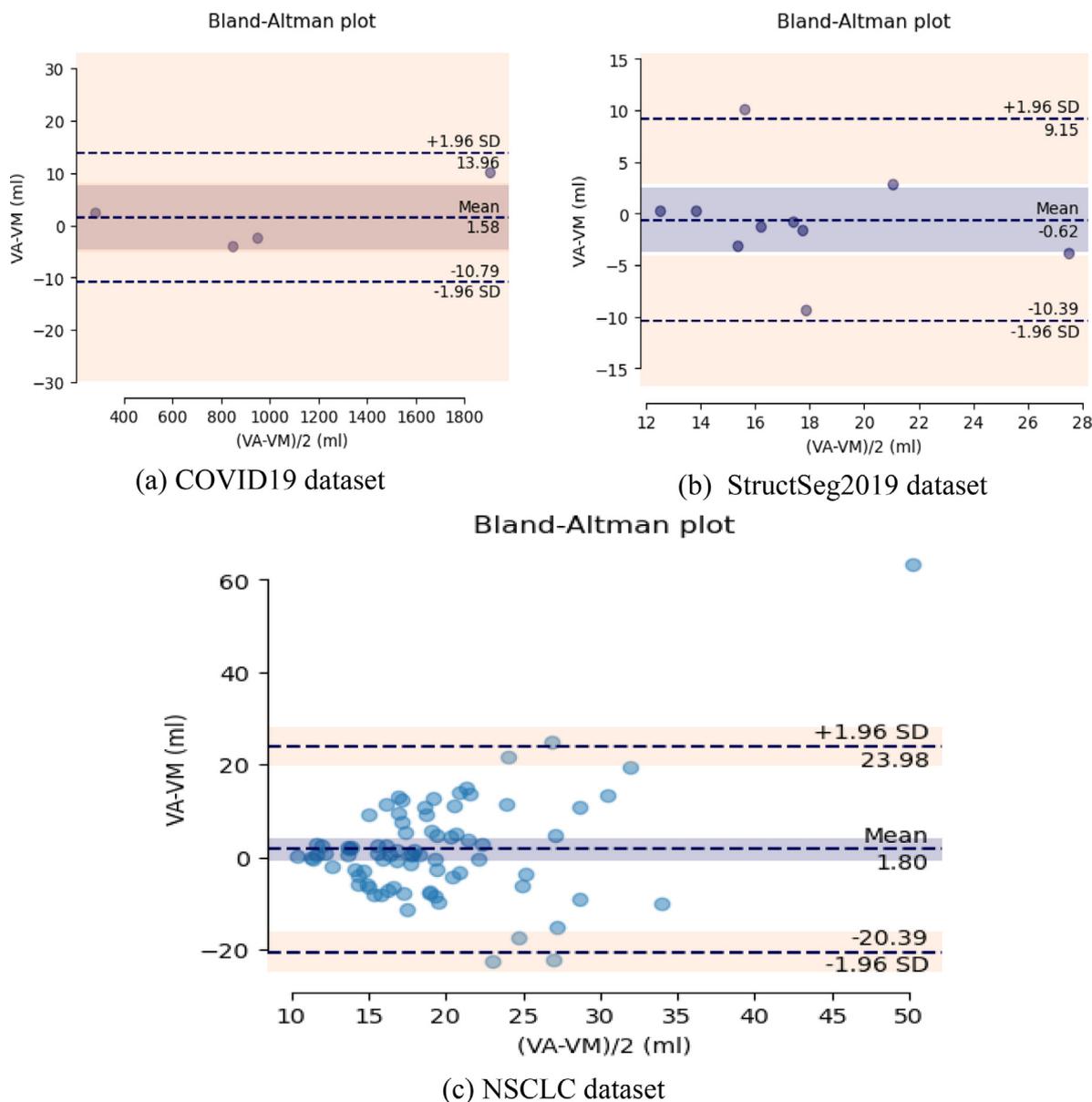
(c) NSCLC dataset

Fig. 11. Band-Altman plots based on the proposed model for all datasets, (a) COVID19, (b) StructSeg2019 dataset, (c) NSCLC dataset.

resents the left and the red color represents the right lung. The visualization shows better segmentation for the proposed model.

The proposed model with existing 3D and 2D UNet that are based on COVID19 and NoN-COVID19 datasets is shown in Table 5. The results indicate that our proposed model produced better dice coefficients in the left, right lung, and infection area and slightly lower than 3D UNet in structseg2019 for the right lung. The 5-fold cross-validation method was used to validate our proposed model for all datasets. We have chosen the best DSC for proposed and existing deep learning as shown in Table 6.

4.2. Ablation study

To demonstrate the impact of each block on the performance of the proposed model, an ablation study was done. We firstly trained a baseline model with proposed encoder and decoder blocks without using SE block in the decoder, multiscale configuration, and KASSP block.

Next, we have added SE to the baseline model (baseline + SE). Besides, the KASSP block was added separately to the baseline model (baseline + KASSP). Further, we used a multiscale approach in our proposed baseline model, and also several configurations were investigated, such as baseline + Multiscale and baseline + K ASSP + Multiscale. Finally, we studied the performance of the proposed model with all modules (baseline + SE + KASSP + Multiscale).

Table 7 presents the results of different configurations of the proposed models. The baseline model achieved DSC scores of 86.2% for 100 CT images Covid, 82.3% for 20 cases Covid, 94.3% for Structseg 2019 and 92.0% for NSCLC dataset. Baseline + Multi scale achieved a better DSC score when compared by adding KASSP with baseline. When we added KASSP with Multiscale, the performance was further achieved in the DSC score for all proposed datasets.

4.3. Performance evaluation on different validation datasets

Our proposed model performs better on the small data sample that can be shown in Table 8. With only 20 cases, our proposed solution achieved better performance as compared to existing state-of-the-art performance. We have performed an additional

experiment in this revised manuscript and we used our proposed model trained on 20 cases Covid dataset and validate on small 100 (slices) Covid dataset to validate the generalization ability of the proposed solution. We keep the same input processing setups for validation of our proposed model on different datasets like the 100 slices dataset. The proposed model achieved better performance and produced 4.1% less performance as compared to when we used the proposed model for training and testing on 100 slices dataset. Similarly, we have used our proposed model to validate other proposed datasets and we have compared the performance that degraded little as compared to training and testing on individual datasets. The proposed model degraded DSC score 2.9% for the Struckseg2019 dataset and 4.5% for the NSCLC dataset.

4.4. Statistical evaluation

We also used Bland–Altman plots to analyze the agreement between the segmentation volumes determined by proposed models for COVID19 and other datasets and manual segmentation. We used a Bland–Altman plot, which graphs the mean difference of measured predicted volume versus manual volume and constructs limits of agreement. A Bland-Altman analysis has been carried out between ground truth and automatic segmentation maps as shown in Fig. 11.

4.5. Discussion and state-of-the-art comparison

The current AI methods are widely used in segmentation problems using a small dataset. Dataset needs to be increased to make the model more generalizable and to minimize the overfitting effect. Most of the existing studies used U-Net for COVID19 segmentation. It is worth noting that interpretability has been a core issue for AI applications in health care. Explainable Artificial Intelligence (XAI) methods [52,53] in most of the AI-based studies have been proposed than the traditional class activation mapping (CAM) method to extract the relevant regions very close to the predicted results. This could be used to diagnose the CVOID-19 in clinical applications. Deep learning has been used in CVOID19 segmentation and classification but, these models could not perform well due to incomplete, inexact, and inaccurate labels and training of

Table 9

The state-of-the-art deep learning model for COVID-19 segmentation using different public and private datasets.

Author and Year	Methods	dataset	Public and private dataset	Overall Dice score	P-values
FeiShan et al. [16]	VB-Net	268 cases	Private	0.916	–
Yu-Huan Wu et al. [17]	JSC model	750 cases	Private	0.783	–
Yu Qiu et al. [54]	MiniSeg DL model	100 axial CT image	Public	0.770	–
Qingsen Yan, et al. [55]	3D COVID-SegNet	861 cases	Private	0.980 lung	–
				0.720 infact	
Narges Saeedizadeh et al. [56]	TV-UNet	900 images	Public	0.863	–
Amine Amyar et al. [57]	Multi-task deep learning	100 CT scans	Public	0.880	–
Zhou et al. [58]	U-Net + DL	100 CT scans and 9 CT volume	Public	0.610	–
Zhou et al. [59]	U-Net + FTL	100 CT scans and 9 CT volume	Public	0.831	–
Adnan Saood et al. [60]	UNet + SegUNet	100 CT scans	Public	0.741	–
Guotai Wang et al. [61]	COPLE-Net	558 cases	Private	0.807	–
Deng-Ping Fan et al. [31]	Inf-Net	100 CT scans (public)	Public	0.739	–
Chen et al. [33]	U-net, M-A, M-R	100 CT scans (public)	Public	0.820	–
				0.850	
				0.840	
Omar Elharrouss at, [62]	SegNet with encoder and decoder	100 axial CT images	Public	0.786	–
Pei, Hong-Yang et al. [63]	MPS-Net	100 CT scans (public)	public	0.832	–
Kumar Singh et al. [64]	Lunginfseg	20 cases	public	0.803	–
Müller, Dominik et al. [65]	Patch based UNet	20 cases	Public	0.858	–
Proposed Model	**The multi-scale and multilevel deep learning model**	**100 axial CT images**	Public	**0.901**	**1.322**
		20 cases		**0.884**	**2.356**
		Structseg2019		**0.975**	**3.191**
		NSCLC dataset		**0.950**	**4.516**

these models is challenging for segmentation and diagnostic network. Moreover, more data gathering and annotation are expensive and time-consuming that could be encouraged to move and investigate deep transfer learning methods and self-supervised deep learning [25,26]. With noisy labels, deep learning models performed well for segmentation and classification tasks [53] and could be potentially used for COVID-19 diagnosis.

A few works have been provided on infection segmentation based on CT slices [24–31]. The variation in texture, size, and position of infections in CT slices is challenging for the detection of COVID-19 infection. Furthermore, the inter-class variance is also small with fewer contrast regions at the boundaries that make the task for detection and segmentation difficult. Due to the busy schedule of the radiologist, another difficulty is arising for big data acquisitions with a smaller number of times. Due to the visual characteristics and special structure, the boundaries of COVID-19 infection regions are hard to differentiate from the chest wall, making it difficult for accurate COVID-19 infection segmentation tasks. Recently, various authors [63–65] proposed 2D and 3D deep learning models for Covid segmentation. The comparison of the results obtained with the proposed model against state-of-the-art models on different datasets is reported in Table 9. The p-value greater than 0.005 shows the similarity between ground truth and segmentation mask. We have computed p-values for all datasets used in this paper shown in Table 9.

5. Conclusion

Automated segmentation and detection of lung infections from computed tomography (CT) had a great interest and could be used to help in understanding COVID-19 infection. The various difficulties such as high variation in infection characteristics and low-intensity contrast between infections and normal tissues have been arising for CT-based lung infection detection. The proposed model based on novel components in encoder and decoder with multiscale and multilevel feature extraction approach produced better segmentation results in lung infection disease detection particularly and globally produced better segmentation on validation of non-covid19 datasets. Extensive experiments based on COVID19 and NON-COVID19 datasets demonstrated that the proposed model outperforms the cutting-edge segmentation models and advances state-of-the-art performances. Our system has great potential to be applied in assessing the diagnosis of COVID-19, e.g., quantifying the infected regions, monitoring the longitudinal disease changes, and mass screening processing. Note that the proposed model can detect objects with low-intensity contrast between infections and normal tissues. This phenomenon is often occurring in nature to disguise objects. In the future, our proposed model could be used for other segmentation tasks with additional features and can be used as an extension of the 3D model for volumetric segmentation.

CRediT authorship contribution statement

Abdul Qayyum: Conceptualization, Methodology, Software, Writing – original draft. **Alain Lalande:** Conceptualization, Resources, Validation, Writing – original draft, Writing – review & editing. **Fabrice Meriaudeau:** Conceptualization, Methodology, Software, Resources, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. H. Organization et al., "Coronavirus disease 2019 (covid-19): situation report, 94," 2020.

[2] R.A. Neher, R. Dyrdak, V. Druelle, E.B. Hodcroft, J. Albert, Potential impact of seasonal forcing on a sars-Nov-2 pandemic, Swiss Medical Weekly 150 (1112) (2020).

[3] S.M. Kissler, C. Tedijanto, E. Goldstein, Y.H. Grad, M. Lipsitch, Projecting the transmission dynamics of sars-cov-2 through the postpandemic period, Science (2020).

[4] A.M. Foust, G.S. Phillips, W.C. Chu, P. Daltro, K.M. Das, P. Garcia-Peña, T. Kilborn, A.J. Winant, E.Y. Lee, International expert consensus statement on chest imaging in pediatric covid-19 patient management: imaging findings, imaging study reporting and imaging study recommendations, Radiology: Cardiothoracic Imaging 2 (2) (2020) e200214.

[5] M. Oudkerk, H.R. Blller, D. Kuijpers, N. van Es, S.F. Oudkerk, T.C. McLoud, D. Gommers, J. van Dissel, H. ten Cate, E.J. van Beek, Diagnosis, prevention, and treatment of thromboembolic complications in covid-19: report of the national institute for public health of the netherlands, Radiology (2020) 201629.

[6] P.J. Mazzone, M.K. Gould, D.A. Arenberg, A.C. Chen, H.K. Choi, F.C. Detterbeck, F. Farjah, K.M. Fong, J.M. Iaccarino, S.M. Janes, J.P. Kanne, E.A. Kazerooni, H. MacMahon, D.P. Naidich, C.A. Powell, S. Raoof, M.P. Rivera, N.T. Tanner, L.K. Tanoue, A. Tremblay, A. Vachani, C.S. White, R.S. Wiener, G.A. Silvestri, Management of lung nodules and lung cancer screening during the covid-19 pandemic: Chest expert panel report, Radiology: Imaging Cancer 2 (3) (2020) e204013.

[7] G. Chassagnon, M. Vakalopoulou, E. Battistella, S. Christodoulidis, T.- N. Hoang-Thi, S. Dangeard, E. Deutsch, F. Andre, E. Guillo, N. Halm, S. El Hajj, F. Bompard, S. Neveu, C. Hani, I. Saab, A. Campredon, H. Koulakian, S. Bennani, G. Freche, A. Lombard, L. Fournier, H. Monnier, T. Grand, J. Gregory, A. Khalil, E. Mahdjoub, P.-Y. Brillet, S. Tran Ba, V. Bousson, M.-P. Revel, and N. Paragios, Ai-driven ct-based quantification, staging and short-term outcome prediction of covid-19 pneumonia, medRxiv, 2020.

[8] S. Chaganti, A. Balachandran, G. Chabin, S. Cohen, T. Flohr, B. Georgescu, P. Grenier, S. Grbic, S. Liu, F. Mellot, et al., Quantification of tomographic patterns associated with covid-19 from chest ct, arXiv preprint arXiv:2004.01279, 2020.

[9] M.-Y. Ng, E.Y.P. Lee, J. Yang, F. Yang, X. Li, H. Wang, M.-S. Lui, C.-Y. Lo, B. Leung, P.-L. Khong, C.-M. Hui, K.-Y. Yuen, M.D. Kuo, Imaging profile of the COVID-19 infection: radiologic findings and literature review, Radiology: Cardiothoracic Imaging 2 (1) (2020) e200034.

[10] N. Zhu, D. Zhang, W. Wang, X. Li, B.o. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019, New England J. Med. 382 (2020) 727–733.

[11] L. Huang et al., Serial quantitative chest CT assessment of COVID-19: deep-learning approach, Radiol., Cardiothoracic Imag., 2(2), 2020, e200075.

[12] J. Lei, J. Li, X. Li, X. Qi, CT imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia, Radiology 295 (1) (2020) 200236.

[13] L. Li et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, Radiology (2020), https://doi.org/10.1148/radiol.2020200905 200905.

[14] F. Shi et al., Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19, IEEE Rev. Biomed. Eng., early access, Apr. 16, 2020, doi: 10.1109/RBME.2020.2987975.

[15] Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, H. Shi, Longitudinal assessment of COVID-19 using a deep learning–based quantitative CT pipeline: illustration of two cases, Radiology: Cardiothoracic Imaging 2 (2) (2020) e200082.

[16] F. Shan et al., "Lung infection quantification of COVID-19 in CT images with deep learning," 2020, arXiv:2003.04655. [Online]. Available: http://arxiv.org/abs/2003.04655.

[17] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, R.-G. Zhang, M.-M. Cheng. JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation." arXiv preprint arXiv:2004.07054 (2020).

[18] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, L. Xia, Serial quantitative chest ct assessment of covid-19: deep-learning approach, Radiology: Cardiothoracic Imaging 2 (2) (2020) e200075.

[19] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and Computer-assisted Intervention, 2015, pp. 234–241.

[20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation, in Proc. MICCAI Workshop DLMIA, in Lecture Notes in Computer Science, vol. 11045, 2018, pp. 3–11.

[21] J. Chen et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study, medRxiv, pp. 1–27, 2020, doi: 10.1101/2020.02.25.20021568.

[22] S. Jin et al., AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system in four weeks, medRxiv, pp. 1–22, 2020, doi: 10.1101/2020.03.19.20039354.

[23] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in Proc. ICDV, 2016, pp. 565–571.

[24] X. Xu, X. Jiang et al., Deep learning system to screen coronavirus disease 2019 pneumonia, arXiv, 2020.

[25] V. Rajinikanth, N. Dey, A.N.J. Raj, A.E. Hassanien, K.C. Santosh, N.S.M. Raja, Harmony-search and Otsu based system for coronavirus disease (COVID-19) detection using lung CT scan images," arXiv, 2020.

[26] S. Chaganti, A. Balachandran et al., Quantification of tomographic patterns associated with COVID 19 from chest CT, arXiv, 2020.

[27] F. Shan, Y. Gao et al., Lung infection quantification of COVID-19 in CT images with deep learning, arXiv, 2020.

[28] F. Shi, L. Xia et al., Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification, arXiv, 2020.

[29] Z. Tang, W. Zhao et al., Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images, arXiv, 2020.

[30] X. Chen, R. Zhang, P. Yan, Feature fusion encoder decoder network for automatic liver lesion segmentation, in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), April 2019, pp. 430–433.

[31] L.-C. Chen, Y. Yang, J. Wang, W. Xu, A. L. Yuille, Attention to scale: scale-aware semantic image segmentation, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3640–3649, 2015.

[32] S. Dong, Z. Gao, S. Sun, X. Wang, M. Li, H. Zhang, G. Yang, H. Liu, S. Li, Holistic and deep feature pyramids for saliency detection, in 29TH British Machine Vision Conference (BMVC), 2018.

[33] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikainen, Deep learning for generic object detection: a survey, Int. J. Comp. Vision 128 (2) (2020) 261–318.

[34] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.

[35] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa et al., Attention u-net: Learning where to look for the pancreas, arXiv preprint arXiv:1804.03999, 2018.

[36] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu, Deeply-supervised nets, in: Int. Conf. Artificial Intelligence and Statistics, 09–12 May, 2015, pp. 562–570.

[37] R. LaLonde, U. Bagci, "Capsules for object segmentation," arXiv preprint arXiv:1804.04241, 2018.

[38] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in CVPR. IEEE Computer Society, July 2017, pp. 6230–6239. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.660.

[39] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848, https://doi.org/10.1109/TPAMI.2017.2699184.

[40] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, et al., Ce-net: Context encoder network for 2d medical image segmentation, IEEE Trans. Med. Imaging 38 (10) (2019) 2281–2292.

[41] K. Kamnitsas, C. Ledig, V.F.J. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Medical Image Anal. 36 (2017) 61–78.

[42] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, IEEE Trans. Pattern Anal. Machine Intel. 35 (8) (2013) 1915–1929.

[43] S.E.A. Raza, L. Cheung, D. Epstein, S. Pelengaris, M. Khan, and N. M. Rajpoot, MIMO-Net: A multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images, in IEEE 14th International Symposium on Biomedical Imaging (ISBI), April 2017, pp. 337–340.

[44] H. Fu, J. Cheng, Y. Xu, D.W.K. Wong, J. Liu, X. Cao, Joint optic disc and cup segmentation based on multi-label deep network and polar transformation, IEEE Trans. Med. Imaging 37 (7) (2018) 1597–1605.

[45] Ma, Jun, Yixin Wang, Xingle An, Cheng Ge, Ziqi Yu, Jianan Chen, Qiongjie Zhu et al. Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation. arXiv preprint arXiv:2004.12537 (2020).

[46] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, F. Huazhu, J. Shen, L. Shao, Inf-Net: automatic COVID-19 lung infection segmentation from CT images, IEEE Trans. Med. Imag. (2020).

[47] K. Kiser, S. Ahmed, S. Stieb, A. Mohamed, H. Elhalawani, P. Park, N. Doyle, B. Wang, A. Barman, C. Fuller, L. Giancardo, Data from thethoracic volume and pleural effusion segmentations in diseased lungs for benchmarking chest ct processing pipelines, Cancer Imag. Arch. (2020).

[48] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017).

[49] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.

[50] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[51] S. Nikolov, S. Blackwell, R. Mendes, J. De Fauw, C. Meyer, C. Hughes, H. Askham, B. Romera-Paredes, A. Karthikesalingam, C. Chu, et al., Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy, arXiv preprint arXiv:1809.04430, 2018.

[52] T. Zhou, H. Fu, G. Chen, J. Shen, L. Shao, Hi-net: hybrid-fusion network for multi-modal MR image synthesis, IEEE Trans. Med. Imag. (2020).

[53] J. Zhang, D.-P. Fan, et al., UC-Net: uncertainty inspired RGB-D saliency detection via conditional variational autoencoders, CVPR, 2020.

[54] Qiu, Yu, Yun Liu, Jing Xu. MiniSeg: an extremely minimum network for efficient COVID-19 segmentation. arXiv preprint arXiv:2004.09750 (2020).

[55] Yan, Qingsen, Bo Wang, Dong Gong, Chuan Luo, Wei Zhao, Jianhu Shen, Qinfeng Shi, Shuo Jin, Liang Zhang, Zheng You. COVID-19 Chest CT image segmentation–a deep convolutional neural network solution. arXiv preprint arXiv:2004.10987 (2020).

[56] Saeedizadeh, Narges, Shervin Minaee, Rahele Kafieh, Shakib Yazdani, and Milan Sonka. COVID TV-UNet: segmenting COVID-19 chest CT images using connectivity imposed U-Net. arXiv preprint arXiv:2007.12303 (2020).

[57] Amyar, Amine, Romain Modzelewski, Su Ruan. Multi-task deep learning based CT imaging analysis for COVID-19: classification and segmentation. medRxiv (2020).

[58] Zhou, Tongxue, Stéphane Canu, Su Ruan. An automatic covid-19 ct segmentation network using spatial and channel attention mechanism. arXiv preprint arXiv:2004.06673 (2020).

[59] Saood, Adnan. COVID-19 lung CT image segmentation using deep learning methods: UNET Vs. SegNET. (2020).

[60] G. Wang, X. Liu, C. Li, X. Zhiyong, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, S. Zhang, A noise-robust framework for automatic segmentation of covid-19 pneumonia lesions from ct images, IEEE Trans. Med. Imag. 39 (8) (2020) 2653–2663.

[61] O. Elharrouss, N. Subramanian, S. Al-Maadeed. An encoder-decoder-based method for COVID-19 lung infection segmentation." arXiv preprint arXiv:2007.00861 (2020).

[62] O. Kherad, M. Moret Bochatay, T. Fumeaux, Utilité du CT-scan thoracique pour le diagnostic et le triage des patients suspects de COVID-19, Rev. Med. Suisse 16 (692) (2020) 955–957.

[63] H.-Y. Pei, D. Yang, G.-R. Liu, L. Tian, MPS-Net: multi-point supervised network for CT image segmentation of COVID-19, IEEE Access 9 (2021) 47144–47153.

[64] V. Kumar Singh, M. Abdel-Nasser, N. Pandey, D. Puig, Lunginfseg: Segmenting covid-19 infected regions in lung ct images based on a receptive-field-aware deep learning framework, Diagnostics 11 (2) (2021) 158.

[65] D. Müller, I. Soto-Rey, F. Kramer, Robust chest CT image segmentation of COVID-19 lung infection based on limited data, Inf. Med. Unlocked 25 (2021) 100681.

ABDUL QAYYUM is currently working as a Postdoc Researcher in ImViA laboratory, University of Bourgogne Franche-Comté, Dijon, France. He received his Ph. D in electrical and electronics engineering with specializaion in deep learning and image processing in 2017 from Universiti Teknologi Petronas Malaysia. He developed prototype for vital sign estimation, assessment of stroke and arterial fibrillation using deep learning approach based on face video analytical approach. His area of interest is machine learning and deep learning for signal processing e.g., EEG, ECG, biomedical images (Cardiac MRI, CT, Prostate, Kidney Tumor) segmentation and classification.

ALAIN LALANDE is currently an Associate Professor in biophysics and medical image processing with the University of Burgundy, Dijon, France, where he is also with the ImViA Laboratory (Image and Artificial Intelligence). His research interests concern the medical imaging domain, initially mainly the imaging of the cardiovascular system (particularly MRI), and cover the topics of image acquisition, image post-processing, and protocol design. His current research interests include more diversified with also augmented reality in robotic surgery, post-processing of CT-scans in LDR prostate cancer brachytherapy, and artificial intelligence in medical imaging.

FABRICE MERIAUDEAU (Member, IEEE) was born in Villeurbanne, France, in 1971. He received the master's degree in physics, the Engineering degree (first) in material sciences, and the Ph.D. degree in image processing from Dijon University, France, in 1994 and 1997, respectively. He held a postdoctoral position with the Oak Ridge National Laboratory for one year. From 2011 to 2016, he was the Director of Le2i (UMR CNRS), which has more than 200 staff members, where he is currently the Professeur des Universites. He has coordinated an Erasmus Mundus Master in the field of computer vision and robotics, from 2006 to 2010. He was the Vice President of the International Affairs for the University of Burgundy from 2010 to 2012. He has authored or coauthored more than 150 international publications. He holds three patents. Since 2016, he has been a Professor with the Universiti Teknologi Petronas. His research interests include image processing for non-conventional imaging systems (UV, IR, and polarization) and, more recently, the medical biomedical imaging.