FV-SEG-Net: Fully Volumetric Segmentation of COVID-19 Lesions in Medical Supply Chain

Mohamed Abdel-Basset; Hossam Hawash; Victor Chang

***Abstract*—Automated and precise pneumonia segmentation of COVID-19 extends the view of medical supply chains and offers crucial medical supplies to fight the COVID-19 pandemic. Deep learning plays a vital role in improving the COVID-19 segmentation from computed tomography (CT) scans. However, most of the current pneumonia segmentation approaches lack precision on small infection areas and often operate by partitioning the CT volumes into 2D slices or 3D patches, leading to the loss of contextual information. To address this, we propose an improved fully volumetric segmentation network, called FV-SEG-Net, that effectively exploits the local and global spatial information and enables the entire CT volume processing at once. The encoder network is implemented with 3D ResNeXt. The decoder is designed using a computationally efficient recalibrated anisotropic convolution (RAC) module that can acquire the 3D semantic representation of the CT volumes with anisotropic resolution. To avoid losing information during down-sampling, we reconstruct the skip-connection using a multi-level multi-scale pyramid aggregation (MPA) module and ensure more effective context fusion that improves the reconstruction capability of the decoder. Empirical investigations demonstrate the proposed FV-SEG-Net has an excellent performance in segmenting COVID-19 lesions with a Dice score of 78.58% and a surface-Dice score of 80.1% outperforming current cutting-edge approaches.**

***Index Terms*— Medical Supply chain; Volumetric Segmentation; Deep Learning; 3D CT Scans; COVID-19, Anistropic Convolution.**

I. Introduction

The emerging coronavirus pneumonia (COVID-19) became a worldwide epidemic in early 2020, by which the World Health Organization (WHO) reported the COVID-19 as a global health emergency of international interest. By June 2020, the world meter community[1] reported 7,630,715 cases of COVID-19 globally, with 424,472 deaths [1]. Several radiographic imagining tools (i.e., computed tomography (CT)) have established their effectiveness for the diagnosis of COVID-19 and the related healthcare tasks [2].

Mohamed Abdel-Basset is with the Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt (e-mail: mohamedbasset@zu.edu.eg).

Hossam Hawash is with the Computer Science, Zagazig University, Zagazig 44519, Egypt (e-mail: hossamreda@zu.edu.eg).

Victor Chang is with the Department of Operations and Information Management, Aston Business School, Aston University, Birmingham, UK (e-mail: ic.victor.chang@gmail.com)

In addition, several therapeutic studies with diverse numbers of patients have verified the proficiency of CT scan in terms of accuracy and precision of COVID-19 detection. These observations imply that radiological CT scans are likely valuable in recognizing COVID-19 at early stages [1], [2].

Based on the present serious situation, it turned to be essential to construct a suitable medical supply chain (MSC) network that enables delivering an applicable diagnosis to COVID-19 patients in a coordinated fashion. Thus, resource management and administration of diagnosis tasks are greatly improved in this state of affairs owing to the automated and early-stage diagnosis of patients. Additionally, the ambiguity in the MSC network of diagnosis services is an inherent component. The requirement in hospitals for different diagnosis services is usually distributed and has to be considered with ambiguity continuously. So, the compassion of work in realizing this MSC raises. In view of this, the location of diagnosis facilities might also speed up the diagnostic service to patients and decrease MSC costs. Moreover, giving more consideration to a viable MSC can lead to more elasticity and tighter dilemmas in the real world [3]. The urgent interest satisfies the communal and environmental requirements throughout the COVID-19 epidemic, forcing hospitals to consider the impact of sustainable MSC network design on the real-world environment. A major theory in sustainability is improving the public accountability of staff (nurses and doctors) in medical institutions, dispensaries, and research laboratories. Therefore, this study investigates highly efficient and automated segmentation of COVID-19 infection to improve diagnostic supplies in MSC [4].

Unlike X-ray, CT assessment is extensively recommended owing to its excellence in 3D visualization of the lung. Recent studies [1]-[5] reported that the distinctive indications of infection might be detected from CT scans (e.g., pulmonary effusion, ground-glass opacity (GGO), and consolidation). The specific estimation of pneumonia and longitudinal differences in 3D CT volumes might hence offer valuable and imperative information for understanding and monitoring COVID-19. Yet, the hand-annotation of lung lesions is a sophisticated task that requires a long time to perform and revise. Additionally, it suffers from inter-spectator and intra-spectator inconsistencies since it is an extremely subjective process that usually depends on personal knowledge and medical experience [5].

In the task of segmenting lesions of COVID-19 pneumonia, the majority of annotations are noisy, and it is difficult to aggregate perfect annotations because of several factors. First, several annotators could have a variety of annotation principles

that result in the problem of inter-observer diversity and high intra-observer diversity [6]. These variabilities are a large potential to bring up noise in the annotations. Second, several studies approved a human-in-the-loop paradigm to minimize the efforts made for annotation, where the radiologist only refined the labels generated by certain algorithms [7]. Hence, the labaels could be chiefly inclined to the model's outcomes and hence include strident pixel-wise annotations. Besides, other studies employ non-experts to aggregate less precise annotations to accommodate the restricted accessibility of specialists. However, these noisy annotations can negatively impact the learning capabilities of deep learning (DL) solutions [8].

A. Limitations and Challenges

Despite its vital role in diagnosis and treatment reports, the task of COVID-19 segmentation from lung CT volumes is difficult for a number of purposes. First, the infection regions have a diversity of composite manifestations such as GGO, consolidation, reticulation, and pulmonary effusion [2], [5]. Second, the magnitude and locations of the infected region differ mostly at various degrees of the infection and vary from one patient to another. Moreover, segmentation is complicated by the irregular structure of lesions and fuzzy margins, and some types of lesions, such as GGO, have low contrast with the enclosing areas [6]. Third, although Convolutional Neural Networks (CNNs) have realized superior results in a broad range of studies of medical segmentation, most of these studies did not entirely exploit the 3D representation of CT volumes (i.e. local information and global information), since they operate either on slice-bases or patch-based manner [7]. Fourth, the segmentation task depends on the perfect annotation of a huge set of CT volumes fulfilled by experienced radiologists, which is difficult to achieve and restricted by the availability of radiologists [8].

B. Contributions

This study contributes to tackling the problem of volumetric segmentation of COVID-19: The fully volumetric convolutional-based lung and lesion segmentation for COVID-19 CT data with an architecture designed to permit the processing of an entire CT volume without any partitioning. This offers better exploitation of both local (voxel-level) and global (spatial) information.

- An adaptive feature recalibration is performed after every block of encoder and decoder using the project and excites (PE) module [47].
- A novel skip-connection is designed using a multi-scale pyramid aggregation (MPA) module to cooperatively capture the global and local representations and concurrently tackle the problems of voxel-wise classification and gland localization.
- An improved decoder is constructed using a novel recalibrated anisotropic convolution (RAC) module to address the problem of lesion size variability and anisotropic resolution issues.

II. RELATED STUDIES

Recent literature presents DL approaches for lesion segmentation and detection of 3D medical images that can be categorized into portion-dependent approaches and in-box segmentation approaches guided by the locations of ROIs [23]. Primarily, as gullible traditions, portion-dependent fully CNN learn from portions of two dimensional slices [9], [10], small 3-D patches [11]–[13], or 2.5-D slices [14] and realize the entire volume inference by sliding over all parts, which is time-consuming and vulnerable to inaccuracy and failures associated with target imperfection. Additionally, portion-dependent approaches experience incomplete active receptive fields [30]. In order to expand the operative receptive in case portion-dependent, Crossbar-Net [15] was introduced to use patches with un-squared shape and with diverse aspect ratios to exploit the global/local context representations efficiently during training.

Concerning 3D approaches like 3D U-Net [11] and 3D V-Net [12], which are developed based on 3D Convs to learn spatial information from image patches. These patches could be categorized as overlapped or non-overlapped. Zhu et al. [16] developed a novel domain adaptive network that uses a boundary-weighted loss and transfers learning method to improve the model vulnerability of the network to boundaries during segmentation. Yet, that approach fails to exploit the entire global/local contexts representation fully. A fast, fully volumetric approach is required to tackle this, and Bontempi et al. [7] introduced an efficient encoder-decoder for brain segmentation and heart atrium segmentation.

Moreover, researchers tend to employ 2D networks to enhance and accelerate the segmentation process of the 3D network in volumetric segmentation tasks. In the literature, approaches that acquire a sequence of 2D segmentation approaches to segment the 3D data are known as 2.5D models. In [17], [18], [13], [10], [14], the authors proposed the fusion of several 2D networks to optimize the space and time complexity on volumetric medical images. They concluded that by exploiting 2D models of segmentation, the relevant 2.5D networks occasionally overcome the 3D networks such as 3D U-Net [9]. Zhou et al. [14] proposed to resolve the segmentation of COVID-19 pneumonia into triple 2D segmentation by depending on the regularity possessions of the lung's tissues, aiming to attain a small number of model parameters and little computational complexity. Generally, several advantages are observed for 2.5D models, including training simplicity, fewer hyper-parameters compared to 3D models, and hence lower complexity, and rapid convergence and prediction time. Yet, such models can still not capture the entire (local and global) semantic representation of CT volume.

To summarize, despite the superiority of the before-mentioned DL approaches suffering two main drawbacks: first, the inability to fully exploit the local and global semantic information of 3D CT volume; and second, the failure to consider the anisotropic spatial resolution of medical volumes, i.e., low inter-slice resolution and large intra-slice resolution. Jia et al. [19] tried to address this issue in magnetic resonance

images using traditional convolutions. However, more effective techniques are still required. Further, these approaches are inclined to suffer from a spatial cutout in the production label maps due to independent predictions of every label, which potentially prevents the effective localization of COVID-19 lesion edges.

III. Proposed Framework

reestablished using a novel MPA module to enable multi-level and multi-view feature aggregation. Third, the decoder is designed using novel RAC modules to anisotropic resolution characteristics.

A. *FV-SEG-Net*

To facilitate dealing with our volumetric data and make it tractable, we prudently adjust the network architecture to deal

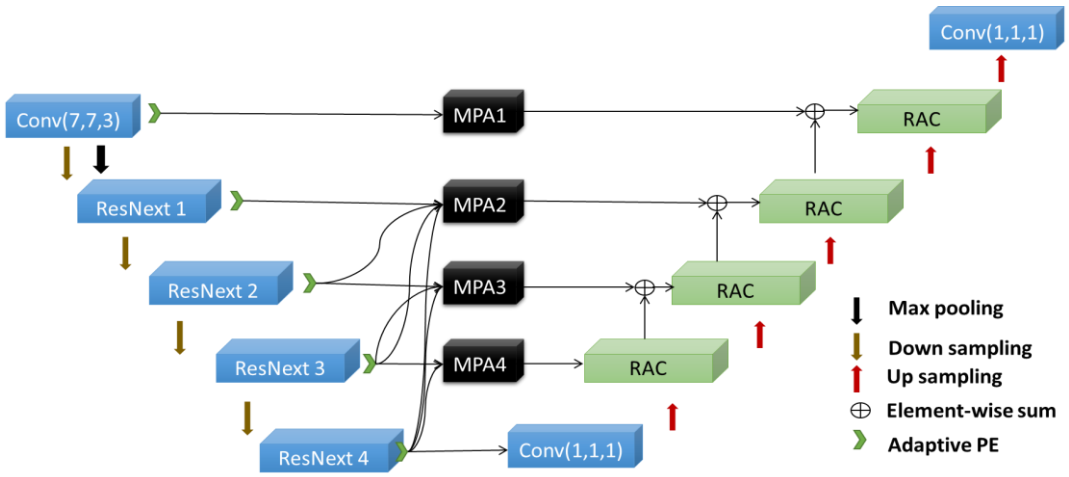

Fig. 1. The structure of the FV-SEG-Net. The whole CT image passed into (left blue boxes) a 3D ResNeXt as a feature encoder, the decoder path is implemented using RAC modules (green boxes), and the skip connection is designed by multi-level multi-scale MPA (black boxes) by aggregating multi-scale semantic information lost through the down-sampling operation of encoder and minimizing the dangers of vanishing gradients.

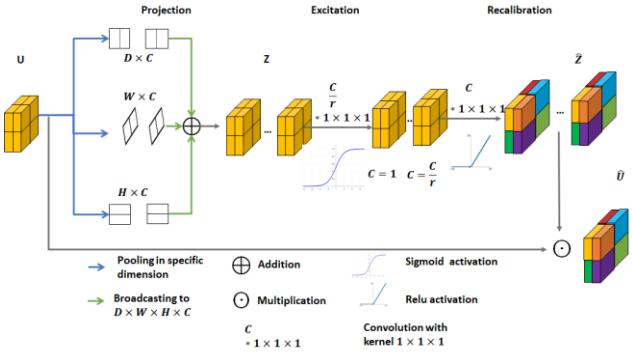

Fig.2. Illustration of PE block for volumetric feature recalibration.

In this part, we discuss the methodology and structure of the proposed FV-SEG-Net for the segmenting COVID-19 infection in MSC. First, we propose a 3D fully volumetric network (FV-SEG-Net) designed to process the whole CT volume without inducing any partitioning and hence enable the exploitation of both local and global semantic information, as presented in Fig.1. Second, the network skip connection linking is

Table I. The configurations of encoder implemented with 3d ResNeXt. The abbreviations 'A' and 'B' are 1 × 1 × 1 *and* 3 × 3 × 1 Convs, correspondingly.

Layer	Output size	Component
Conv 1	48 × 48 × 32	7 × 7 × 3, stride (2, 2, 1)
Pool 1	48 × 48 × 16	7 × 7 × 3, max pool, stride (1, 1, 2)
Pool 2	24 × 24 × 8	7 × 7 × 3, max pool, stride 2
ResNext-block 1	24 × 24 × 8	$((A, 64), (B, 64), (A, 256)) \times 3$
ResNext-block 2	12 × 12 × 4	$((A, 64), (B, 64), (A, 256)) \times 3$
ResNext-block 3	6 × 6 × 2	$((A, 64), (B, 64), (A, 256)) \times 3$
ResNext-block 4	3 × 3 × 1	$((A, 64), (B, 64), (A, 256)) \times 3$

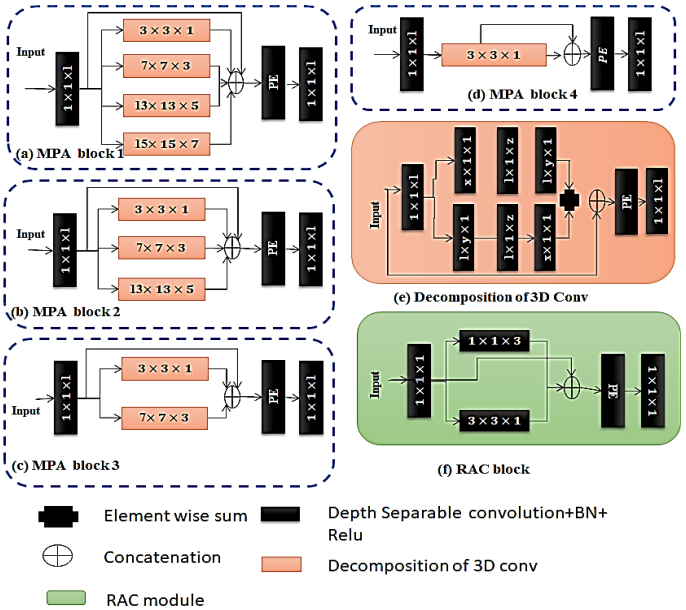

Fig. 3. Architectures of MPA blocks and RAC blocks: the subsections (a)-(d) present the architecture of MPA module 1, 2, 3, and 4 implemented with 3d depth separable convolution(DS-Conv), (e)the decomposition of the 3D convolution size $x \times y \times z$ in MPA blocks, and (f)present the architecture of the proposed RAC block.

with the restrained GPU memory efficiently. A machine comprising three GeForce® GTX 2080 Ti is used for training, and diverse portions of the model were allocated to diverse GPUs. Motivated by the recent work in [7], we introduce a 3D encoder-decoder architecture with eight convolutional modules.

Meanwhile, we consider the whole CT volume as an input. The convolutional module learns to extract the feature maps

Table II. The configuration of the proposed decoder architecture.

Layer	Output size	Component
Conv 1	$3 \times 3 \times 1$	$1 \times 1 \times 1$, 1024
Up sample 1	$6 \times 6 \times 2$	$\times$ 2 tri-linear up sample
RAC block 1	$6 \times 6 \times 2$	I, M, O: 1024, 256, 1024
Up sample 2	$12 \times 12 \times 4$	$\times$ 2 tri-linear up sample
RAC block 2	$12 \times 12 \times 4$	I, M, O: 512, 192, 512
Up sample 3	$24 \times 24 \times 8$	$\times$ 2 tri-linear up sample
RAC block 3	$24 \times 24 \times 8$	I, M, O: 256, 128, 2563
Up sample 4	$48 \times 48 \times 16$	$\times$ 2 tri-linear up sample
RAC block 4	$48 \times 48 \times 16$	I, M, O: 128, 64, 128
Conv 2	$48 \times 48 \times 32$	$1 \times 1 \times 1$, 2

'I=input','M=middle channel', 'O=output channel'

that are not restricted to patches and extend over the complete volume. Every module captures the content of the entire Lung CT, which in turn enables acquiring both global and local information by exploiting the spatial semantics that is broadcasted to every successive layer. The competence of FV-SEG-Net to learn all semantic features is intelligible in the ending layer of the network with a hypothetical receptive field of $100 \times 100 \times 100$. In order to extract more informative feature maps, we perform feature extraction using a pre-trained ResNeXt [20]. We use only the first three blocks for computational efficiency and omit the fourth and the average pooling operations and fully connected layers (FCLs). The main factor that motivated us to adopt this architectural design is that the shortcut mechanism help accelerating model convergence and evade gradient vanishing [12], [16]. Additionally, it leverages a "*split-transform-merge*" mechanism that enables aggregation of multi-view transformations that empirically improve representations of the power of the network. In order to empower the encoder capabilities, a feature recalibration operation is applied to the output of every block in encoder architecture. Table I present the parameters of the 3D ResNeXt encoder.

Project and Excite (PE) block [21] was adopted to perform volumetric feature recalibration. An illustration of the structure of the *PE* module is illustrated in Fig. 2. The output convolutional maps of the earlier layer are passed as input U to the PE module. Then, this input is compressed using function $C(\cdot)$ into a representation Z with tiny dimensions. Follow, the function $P(\cdot)$ is employed to generate recalibrated map $\hat{Z}$ by capturing the inherent mapping from the compressed feature maps Z. After that, the recalibration function $R(\cdot,\cdot)$ is performed, which primarily applies a gating mechanism to scale up $\hat{Z}$ and multiply the input maps U with $\hat{Z}$, generating a recalibrated representation $\hat{U}$. Different approaches can be employed to compress the feature maps involving parametric layers (i.e. Conv layers) and non-parametric pooling layers. The function $P(\cdot)$ is usually designed via a parametric dense or convolution layer.

Motivated on the notion that the global pooling layer could not appropriately acquire the pertinent positional representation (spatial) of the 3D CT scan, the PE operation is employed to empower the network to model comprehensive enlightening positional features beyond the project and excite operation. The inter-relationships among the projections through the various

Table III. The configuration of the skip-connection.

Layer	Output size	Component
MPA block 1	$48 \times 48 \times 16$	I, M, O: 64, 64, 64
MPA block 2	$24 \times 24 \times 8$	I, M, O: 256, 64, 256
MPA block 3	$12 \times 12 \times 4$	I, M, O: 512, 64, 512
MPA block 4	$6 \times 6 \times 2$	I, M, O: 1024, 256, 1024

'I=input','M=middle channel', 'O=output channel'

channels are modeled by means of the excitation operation. Accordingly, PE combines and exploits both spatial and channel information during the feature recalibration operation. The function $C(\cdot)$ is divided into threesome projection procedures, namely $C_H(\cdot)$, $C_W(\cdot), and\ C(\cdot)$, along the spatial dimensions with outputs $z_{h_c} \in \mathbb{R}^{C\times H}$, $z_{w_c} \in \mathbb{R}^{C\times W}$, and $z_{d_c} \in \mathbb{R}^{C\times H}$. Herein we propose to use adaptive average pooling operation to designate the projection procedure. So, we perform spatial dimensions averaging according to equations (1-3).

$$C_H: \quad z_{h_c}(\mathrm{i}) = \frac{1}{W}\frac{1}{D}\sum_{j=1}^{W}\sum_{k=1}^{D} u_c(i,j,k) \tag{1}$$

$$C_W: \quad z_{w_c}(j) = \frac{1}{H}\frac{1}{D}\sum_{i=1}^{H}\sum_{k=1}^{D} u_c(i,j,k) \tag{2}$$

$$C_D: \quad z_{d_c}(k) = \frac{1}{H}\frac{1}{W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j,k) \tag{3}$$

whereas $i \in \{1,\cdots,H\}$, $j \in \{1,\cdots,W\}$, and $k \in \{1,\cdots,D\}$, then the outcome $z_{h_c}, z_{w_c}, and\ z_{d_c}$ are transported to the dimension $H \times W \times D \times C$ and added to obtain Z that followingly fed to the function $P(\cdot)$ consisting of two convolutions ($kernel\ size = 1 \times 1 \times 1$) accompanied with $Relu$ activation; which enables efficient utilization of interrelations among various channels. On convolution layer is responsible for lowering the count of channels by r, while the other one retrieves the original dimensions of feature maps. The $P(\cdot)$ and $R(\cdot,\cdot)$ procedures are computed according to the (4) and (5):

$$P: \quad \hat{Z} = \mathrm{V}_2 * \delta(\mathrm{V}_1 * \mathrm{Z}), \tag{4}$$

$$R: \quad \hat{\mathrm{U}} = \sigma(\hat{Z}) \odot \mathrm{U}, \tag{5}$$

whereas $*$ defines the convolutional layer and $\odot$ designates Hadamard product, $\mathrm{V}_1 \epsilon \mathbb{R}^{1\times1\times1\times\frac{C}{r}}$ and $\mathrm{V}_2 \epsilon \mathbb{R}^{1\times1\times1\times C}$ represent the corresponding parameters.

Feature Decoder: For a fast and efficient reconstruction of high-resolution feature maps, several improved decoding modules are introduced in the feature decoding branch. The decoder reestablishes the volumetric representation via improved repreentations captured with the encoder and progressively carries out global context fusion using skip connection modules, to be discussed later.

From the investigation of the conventional convolutional layer, several shortcomings were observed. First, it exhibits a fixed scale that unavoidably produces static receptive fields across the image. Nevertheless, the lesions in the CT image often have diverse sizes with various shapes. Hence the fixed scale confines the feature fusion capability. Second, it suffers

from information loss [19], owing to the down-sampling procedure and the increased network depth that complicate the optimization and cause gradient disappearance [22]. Third, it fails to handle anisotropic voxel resolution [19]. Thus, a perfect convolution design can adaptively select kernels with multiple adequate scales that match variability in lesion sizes. To tackle the before-mentioned limitations, we construct a decoder based on the Recalibrated Anisotropic Convolutional (RAC) instead of just employing the conventional isotropic 3D convolutions. All the layers of the RAC block are implemented using depth separable convolutional (DS-Conv) for reduced computational complexity.

As presented in Fig. 3, the feature maps passed to the RAC block are firstly passed into $[1 \times 1 \times 1]$, the generated output concurrently fed into $[3 \times 3 \times 1]$ and $[1 \times 1 \times 3]$, and their outcomes are then fused along with the input using a concatenation operation. Feature recalibration is applied to the concatenated output through the PE block and followingly passed to the $[1 \times 1 \times 1]$ base Conv. The main purpose of the $[1 \times 1 \times 1]$ layer is to regulate the channel numbers of features and to eschew the ungovernable growth of computational complexity. According to this architectural design, the DS-Conv $[3 \times 3 \times 1]$ ultimately utilizes the 2D representation embedded in the $X - Y$ planes, and simultaneously the DS-Conv $[1 \times 1 \times 3]$ emphasizes learning inter-slices features. For every fusion scale, the up-sampling procedure is applied to rescale up the feature maps to original size utilizing the tri-linear interpolation, and Conv $[1 \times 1 \times 1]$ is also employed to match the numbers of a channel of feature maps of two neighboring scale phases. Finally, at the termination of this decoding path, another Conv $[1 \times 1 \times 1]$ operation is applied to produce the volumetric estimate of the COVID-19 lesion segmentation for the input volume. In our RAC block, the size of the kernel was experimentally set to 3 and did not differ from the spatial resolution of CT volumes for the subsequent factors. First, before training, all of the CT volumes were resized to the same size, so one kernel is employed for all CT scans. Second, a couple of anisotropic convolutions were used for feature extraction from $z - direction$ on the $x - y$ flat to mitigate the influence of the anisotropic characteristics. Table II shows the parameters of the DAC.

Skip links: Skip linking between encoder and decoder modules has been broadly deployed in a variety of segmentation models [9], [18], as it could extenuate the gradient vanishing problem through the process of errors backpropagation, which enables keeping shallow locative representations of CT image in feature maps. Nevertheless, these kinds of representation might be gradually diminished once they are progressively connected to low-level blocks [9]. In addition, the standard skip-links often provide unrelated disorders and exhibit a semantic break owing to the discrepancy of receptive fields. For 3D segmentation of COVID-19 lesion from CT volumes, it is extremely beneficial to aggregate the local and global representation at several scales. Thus, we propose to reestablish the skip connection using a novel Multi-scale Pyramid aggregation (MPA) module by merging the convolutional maps of the same phase with the convolution maps of all upper-level phases as presented in Fig.3. The input for the MPA module is resampled into the same dimension fed into a sequence of DS-Conv. The planned MPA module empowers the model to gain two advantages. First, for every scale, a moderately large kernel convolution is utilized, emulating the densely linked classification network [19], which can fine-tune the voxel classification performance of the model in addition to its original auspicious localization capability. Second, in addition to these large-scale kernels, another depth-wise separable convolution with a variety of small-scale kernels are also created concurrently to establish a pyramidical convolutional module, displayed as MPA 1- 4 in Fig. 3, which effectively perform feature fusion on diverse areas and also avoid the loss of semantic information at various volume scales. Third, DS-Convs bring the advantage of reduced model parameters compared to standard convolutions. Nevertheless, utilizing large-scale kernels could enlarge the computational complexity inclined to be confined by the hardware capacity. To tackle this problem, we perform a 3D convolutional decomposition inspired by the 2D decomposition deployed in [19], [22]. Unambiguously, a high dimension kernel 3D Conv $x \times y \times z$ is decaying into a full connected mixture of $[x \times 1 \times 1], [1 \times y \times 1], [1 \times 1 \times z]$ and $(1 \times y \times 1), (1 \times 1 \times z), (x \times 1 \times 1)$ convolutions, as presented in Fig.3. Accordingly, this enables reducing the total parameters from O^3 to $3O$. Table III shows the specifications of skip linking blocks.

IV. Experiments and Results

A. Implementation Setup

For experimenting with FV-SEG-Net, the model was trained to optimize both CEL and DIL loss utilizing the optimizer of Adam, for 50 training epochs, with 0.00005 as initial rate of learning. The training process follow the 5-fold cross-validation. Overall, the implementation of this work is performed in Python using the PyTorch library.

B. Data Set

To assess the the proposed FV-SEG-Net, a public COVID-19 CT dataset [23] is employed. The dataset comprises twenty public COVID-19 CT volumes with more than 1800 annotated slices. The data was partitioned into 16 and 4 volumes for training and testing, respectively. In our experiments, we evaluate the models using a 5-folded stratified cross-validation strategy. Additionally, 50 CT volumes published in the MosMedData dataset [24] are also employed for experimental evaluations.

C. Evaluation metrics

Inspired by their recent and extensive use for the evaluation of segmentation techniques, we propose to assess the segmentation performance of FV-SEG-Net using two metrics. First, the Dice similarity coefficient (DSC) and Normalized surface Dice (NSD), which are computed as follow:

$$DSC = \frac{2|S \cap G|}{|S| + |G|} \tag{16}$$

$$NSD = \frac{2|\partial S \cap B^{\tau}_{\partial S}| + 2|\partial S \cap B^{\tau}_{\partial G}|}{|\partial S| + |\partial G|} \tag{17}$$

where $B_{\partial S}^{\tau}, B_{\partial G}^{\tau}$ designate the border area of GT as well as surface at τ acceptance. Herein, the tolerance τ is set to $1mm$ and $3mm$ for segmenting lung and lesion, correspondingly. Moreover, the Mean Absolute Error (MAE) is adopted to calculate the pixel-wise divergence between segmentation S_p and GT outcomes as formulated in (18).

$$MAE = \frac{1}{w \times h \times d} \sum_{x}^{w} \sum_{y}^{h} \sum_{z}^{d} \left| S_p(x,y,z), GT(x,y,z) \right| \quad (18)$$

D. Results and Comparisons

techniques with DSC of 79.4±15.2 and DSC of 80.80±16.3 on the left and right lung segmentation, respectively. It can be noted that 2D-based approaches [9],[25] attain the greatest performance compared with the other approaches, which can be explained because they are trained on slice level, so it uses 1800 slices, which are enough to make the model converge. On the other hand, 3D patch-based models [11] exhibit lower performance than the before-mentioned 2D models with a 10% and 8% drop in DSC of left and right lung segmentation correspondingly. The 2.5 H-DUnet [13] achieved similar results

*Table IV. The results of comparative analysis (mean ± standard deviation) for left and right lung segmentation using the first dataset[23]. The asterisk * indicates the statistical significance of the result.*

		Left Lung segmentation			Right lung segmentation		
Methods	**Data**	**DSC↑**	**NSD↑**	**MAE↓**	**DSC↑**	**NSD↑**	**MAE↓**
2D-U-Net [9]	*L*	95.1±7.9	84.6±12.7	0.106	95.60±7.4	85.5±12.8	0.109
2D-FCN-8[25]	*L*	94.2±8.1	85.1±8.2	0.110	95.10±8.4	85.6±9.8	0.115
3D-U-Net [11]	*L*	85.8±10.5	71.2±13.8	0.136	87.90±9.3	74.8±11.9	0.138
3D-V-Net [12]	*L*	87.1±9.2	74.2±13.2	0.142	88.80±13.3	76.8±15.2	0.129
H-DUnet [13]	*L*	86.6±9.2	75.2±10.1	0.147	87.40±9.3	74.8±16.2	0.131
MPUnet [10]	*L*	79.4±15.2	68.4±11.6	0.141	80.80±16.3	69.7±12.5	0.141
*** FV-SEG-Net**	***L***	**96.01±5.1***	**86.5±7.7***	**0.099**	**96.17±7.20***	**0.865±8.9***	**0.088***

*Table V. The results of comparative analysis (mean ± standard deviation) for for COVID-19 lesion segmentation the first dataset [23]. The asterisk * indicates the statistical significance of the result.*

Methods	**Data**	**DSC↑**	**NSD↑**	**MAE↓**
2D-U-Net [9]	*L*	60.9±24.5	61.5±27.0	0.123
2D-FCN-8[25]	*L*	61.5±18.11	60.1±21.1	0.118
3D-U-Net [11]	*L*	67.3±22.3	70.0±24.4	0.107
3D-V-Net [12]	*L*	68.8±19.4	69.3±17.3	0.099
H-DUnet [13]	*L*	68.6±14.2	67.2±18.1	0.109
MPUnet [10]	*L*	67.4±15.2	66.4±13.9	0.111
*** FV-SEG-Net**	***L***	**74.25±8.1***	**75.81±9.30***	**0.059***

Table VI. Model comparison for infection segmentation on MoSMedData with real CT volumes

Methods	**DSC↑**	**NSD↑**	**MAE↓**
H-DUnet [13]	45.16	43.0	0.167
2D-U-Net [9]	35.81	37.24	0.214
3D-U-Net [11]	62.41	64.46	0.106
Inf-Net [26]	57.11	58.12	0.054
Model [14]	70.59	96.09	0.034
***FV-SEG-Net**	**75.11**	**74.91**	**0.029**

In this part, we present the segmentation results to prove the proposition that refraining from partitioning the lung CT volume enables the FV-SEG-Net to learn global spatial representation and consequently improve the segmentation performance. Table V presents the quantitative results of the proposed FV-SEG-Net in comparison with other cutting-edge architectures. Specifically, FV-SEG-Net is compared against the common 2D-patch based U-Net [9] and FCN-8[25] trained using three primary views (i.e., axial, longitudinal, and coronal), the 3D U-Net [11] and 3D V-Net [12] with 3D patches dimensions $64 \times 64 \times 64$, and the 2.5 base approaches H-DUnet [13] and MPUnet [10], which perform view-accumulation from 2D-patch architectures. All the models are trained on the same number of CT volumes for fifty epochs to optimize the same loss, using the comparable learning rates (with some alteration to warrant the greatest performance).

Overall, despite having far fewer parameters, the FV-SEG-Net outperforms the existing approaches on lung and lesion segmentation, and it exhibits a minor variability (indicating higher consistency). Moreover, we calculate the $p - values$ for these results via a $paired\ t - test$ using a SciPy library. In Table IV and Table V, results that exhibit statistically significant estimated with paired t-test are annotated with asterisks.

Specifically, Table IV presents the quantitative results of our models compared with the existing approaches. We observe that MPUnet [10] achieves the worst results among all to the 3D models. It is obvious that our proposed FV-SEG-Net outperforms the existing approaches on all performance measures using only the same labeled data with 1% improvement in DSC ($p - value$; 0.022) and 2% improvement in NDS ($p - value$,0.035).

More importantly, Table V presents the quantitative results of our models against the existing approaches for COVID-19 lesion segmentation. We observe that 2D-U-Net [9] and 2D-FCN-8[25] achieve the lowest performance with DSC of 60.9 and 61.5, respectively. 3D-U-Net [11] and 3D-VNet [12] attain around 7% and 8% improvement over 2D models. The H-DUnet [13] exhibits similar DSC performance, yet it achieves a 2% lower NSD measure. It can be clearly noted that the proposed FV-SEG-Net attains around 5% DSC improvement ($p - value$,0.037) over the 3D models and 13% DSC improvement ($p - value$,0.019) over the 2D model. For the NSD measure, FV-SEG-Net attains 4% improvement ($p - value$,0.031), 13% improvement ($p - value$,0.018) over the 2D model, and 7% improvement ($p - value$, 0.027) over H-DUnet [13]. The before-mentioned results demonstrate the superiority of the proposed FV-SEG-Net in the supervised segmentation of pneumonia lesions. Some of the segmentation outcomes generated by the FV-SEG-Net on axial images are shown in Fig. 4.

In practical experiments, every 3D CT volume consists of several 2D slices; wherein the vast majority of slices potentially have no ROI. In order to provide extra validation for the

Table VII. the results of ablation experiments (mean ± standard deviation) for leaion segmentation on the first dataset[23].

Methods	DSC↑	NSD↑
Baseline	67.6±22.30	71.09±24.4
Baseline+ PE	70.09±21.40	72.80±19.0
Baseline+ PE+ IC	70.10±14.50	72.95±18.0
Baseline+ PE+ RAC	71.54±18.11	73.58±21.1
Baseline +PE+ MPA	71.30±22.30	73.84±16.4
Baseline +PE+ RAC+ MPA	72.18±11.20	73.99±13.9
Baseline +PE+ RAC+ MPA +PT (FV-SEG-Net)	**74.25±8.1**	**75.81±9.30**

performance of the proposed FV-SEG-Net, we compare the performance of the proposed FV-SEG-Net against the before-mentioned competing methods using the MosMedData dataset. The results of this comparison are presented in Table VI. It is notable that the 2D U-Net realizes the lowest segmentation performance with 35.81 of DSC. The Inf-Net [26] attain great performance improvement over the 2D U-Net. Among the competing methods, the model [14] attains the best segmentation performance (DSC: 70.59, NSD: 96.09). More importantly, the proposed FV-SEG-Net overcame all the competing methods with great performance improvement (DSC: 70.59, NSD: 96.09) over the model [14] ($p-value$, 0.039). This further validates the efficiency and the generalizability of the proposed FV-SEG-Net thanks to its ability to capture contextual information in CT volume during training. In spite of comprising non-infected slices, our approach still outperforms the existing approach by a considerable margin. This further explains the superiority of the proposed segmentation approach for tackling the problem of limited annotated data and the heterogeneous dataset problem and for effectively dealing with the situation of having non-infected images as an input.

V. Discussion and analysis

Following the obtained results, in this section, an ablation study was introduced to assess the contribution of the proposed MPA blocks, RAC blocks, and PE blocks to the overall performance of the proposed FV-SEG-Net. For suitability, we use the term baseline model for the original basic U-shape model with the ResNeXt backbone. Eight versions of FV-SEG-Net were implemented as follows. First, the baseline architecture was implemented by attaching PE blocks between the convolution block of the encoder and decoder. Second, we implemented the FV-SEG-Net by replacing RAC modules with traditional isotropic convolutional (Is-Conv) modules [19]. Third, we implemented the FV-SEG-Net by using RAC modules to construct the decoder rather than traditional convolution (with a traditional skip connection). Fourth, we implemented the FV-SEG-Net by replacing the conventional skip-connection with MPA skip-connection. Fifth, we developed one more version to indicate the effect of pretraining (PT). In all of these experiments, all parameter configurations are the same, and only the structure changes each time.

The achieved results of applying FV-SEG-Net and its four versions on the COVID-19-CT-Seg dataset [23] utilizing five-fold cross-validation are presented in Table VII, from which it can be seen that the PE module for feature recalibration has a

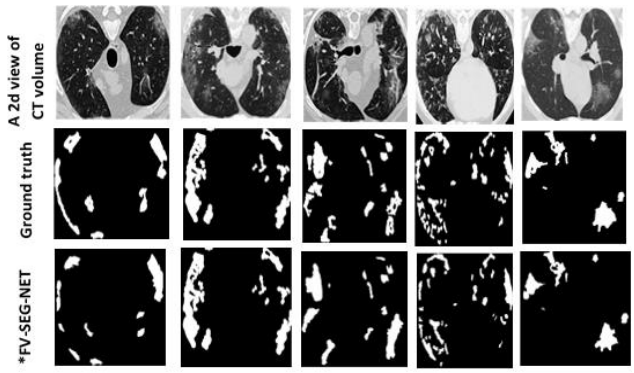

Fig.4 An illustrative representation for FV-SEG-NET segmentation outcomes against the ground truth

3% improvement on DSC and NSC. This demonstrates the effectiveness of applying PE blocks in our model. The Is-Conv attains comparable performance as traditional convolutional layers. Meanwhile, the RAC module attains 1.4% and 0.7% improvements on DS and NSD compared to the Is-Conv-based architecture, which demonstrates the effectiveness of the RAC modules for implementing the decoding path of the network. Further, it can be noted that adopting the MPA module to reconstruct the model's skip-connection resulted in 1.2% and 1.1% improvements respectively on DSC and NSD measure over Basleine+PE architecture. Additionally, it is obvious that applying all the before-mentioned modules to the non-pre-trained Baseline (Baseline +PE+ RAC+ MPA) results in around 5% and 3% improvement on DSC and NSD correspondingly. This further explains the cooperative effect of these modules in capturing local and global context information from input CT volumes. This experiment is performed again with pre-trained ResNeXt as the encoder (Baseline +PE+ RAC+ MPA+PT); the result shows the pretraining further performance improvement over the previous experiment (Baseline +PE+ RAC+ MPA) with 2% higher DSC and NSD. Also, we observed that pretraining exhibits a much lower standard deviation of the result, which indicates more stability gain. Making these improvements are highly crucial for MSC and quality of medical diagnosis.

VI. Conclusion and future work

In this study, a novel FV-SEG-Net framework is presented, an encoder-decoder alike approach for segmentation of lung and/or COVID-19 pneumonia in a fully volumetric manner. The main aim is to improve the quality of the automatic segmentation process to empower the diagnostic services over MSC networks. The architectural design of FV-SEG-Net is cautiously optimized to produces a segmentation outcome in a reasonable time for desktop GPU. The experimental results show that the FV-SEG-Net yields better segmentation results than other models on two COVID-19 test sets. The proposed RAC blocks and MPA blocks effectively contribute to the improved performance.

The performance of the proposed model is evaluated and compared with the cutting-edge 2D, 2.5D, and 3D-patch-based approaches that have a similar structure. The achieved results demonstrate the potential of the FV-SEG-Net framework since it outperforms the existing 2D, 2.5D, and 3D-patch-based

models by means of appropriate parameters. Eliminating the CT volume segregating, as theorized, enables the model to capture both local and spatial features effectively.

In future work, we intend to address variability in scale and shape of lesions simultaneously; a recently proposed Anisotropic convolution could be used to achieve this. Further, we intend to develop a distributed learning scheme that enables improved medical supply chain to deliver diagnosis and medicine to COVID-19 patients.

REFERENCES

[1] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*. 2020, doi: 10.1016/S0140-6736(20)30185-9.

[2] T. Ai *et al.*, "Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases," *Radiology*, 2020, doi: 10.1148/radiol.2020200642.

[3] S. K. Panda and S. C. Satapathy, "Drug traceability and transparency in medical supply chain using blockchain for easing the process and creating trust between stakeholders and consumers," *Pers. Ubiquitous Comput.*, 2021, doi: 10.1007/s00779-021-01588-3.

[4] S. Khan, A. Haleem, S. G. Deshmukh, and M. Javaid, "Exploring the impact of COVID-19 pandemic on medical supply chain disruption," *J. Ind. Integr. Manag.*, 2021, doi: 10.1142/S2424862221500147.

[5] C. Bao, X. Liu, H. Zhang, Y. Li, and J. Liu, "Coronavirus Disease 2019 (COVID-19) CT Findings: A Systematic Review and Meta-analysis," *J. Am. Coll. Radiol.*, 2020, doi: 10.1016/j.jacr.2020.03.006.

[6] M. Li, W. Hsu, X. Xie, J. Cong, and W. Gao, "SACNN: Self-Attention Convolutional Neural Network for Low-Dose CT Denoising with Self-Supervised Perceptual Loss Network," *IEEE Trans. Med. Imaging*, 2020, doi: 10.1109/TMI.2020.2968472.

[7] D. Bontempi, S. Benini, A. Signoroni, M. Svanera, and L. Muckli, "CEREBRUM: a fast and fully-volumetric Convolutional Encoder-decodeR for weakly-supervised sEgmentation of BRain strUctures from out-of-the-scanner MRI," *Med. Image Anal.*, 2020, doi: 10.1016/j.media.2020.101688.

[8] G. Wang *et al.*, "A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions from CT Images," *IEEE Trans. Med. Imaging*, 2020, doi: 10.1109/TMI.2020.3000314.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015, doi: 10.1007/978-3-319-24574-4_28.

[10] J. M. J. Valanarasu, R. Yasarla, P. Wang, I. Hacihaliloglu, and V. M. Patel, "Learning to Segment Brain Anatomy from 2D Ultrasound with Less Data," *IEEE J. Sel. Top. Signal Process.*, 2020, doi: 10.1109/JSTSP.2020.3001513.

[11] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," 2016, doi: 10.1007/978-3-319-46723-8_49.

[12] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," 2016, doi: 10.1109/3DV.2016.79.

[13] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes," *IEEE Trans. Med. Imaging*, 2018, doi: 10.1109/TMI.2018.2845918.

[14] L. Zhou *et al.*, "A Rapid, Accurate and Machine-Agnostic Segmentation and Quantification Method for CT-Based COVID-19 Diagnosis," *IEEE Trans. Med. Imaging*, 2020, doi: 10.1109/TMI.2020.3001810.

[15] Q. Yu, Y. Shi, J. Sun, Y. Gao, J. Zhu, and Y. Dai, "Crossbar-Net: A Novel Convolutional Neural Network for Kidney Tumor Segmentation in CT Images," *IEEE Trans. Image Process.*, 2019, doi: 10.1109/TIP.2019.2905537.

[16] Q. Zhu, B. Du, and P. Yan, "Boundary-Weighted Domain Adaptive Neural Network for Prostate MR Image Segmentation," *IEEE Trans. Med. Imaging*, 2020, doi: 10.1109/TMI.2019.2935018.

[17] N. Savioli, G. Montana, and P. Lamata, "V-FCNN: Volumetric Fully Convolution Neural Network for Automatic Atrial Segmentation," 2019, doi: 10.1007/978-3-030-12029-0_30.

[18] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-UNet: A Dimension-Fusion U Shape Network for Chronic Stroke Lesion Segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, 2021, doi: 10.1109/TCBB.2019.2939522.

[19] H. Jia *et al.*, "3D APA-Net: 3D Adversarial Pyramid Anisotropic Convolutional Network for Prostate Segmentation in MR Images," *IEEE Trans. Med. Imaging*, 2020, doi: 10.1109/TMI.2019.2928056.

[20] O. Kopuklu, N. Kose, A. Gunduz, and G. Rigoll, "Resource efficient 3D convolutional neural networks," 2019, doi: 10.1109/ICCVW.2019.00240.

[21] A. M. Rickmann, A. Guha Roy, I. Sarasua, and C. Wachinger, "Recalibrating 3D ConvNets with Project Excite," *IEEE Trans. Med. Imaging*, 2020, doi: 10.1109/TMI.2020.2972059.

[22] W. Li *et al.*, "Anisotropic Convolution for Image Classification," *IEEE Trans. Image Process.*, 2020, doi: 10.1109/TIP.2020.2985875.

[23] J. Ma *et al.*, "Towards efficient COVID-19 CT annotation: A benchmark for lung and infection segmentation," *arXiv*, 2020.

[24] S. P. Morozov *et al.*, "MosMedData: data set of 1110 chest CT scans performed during the COVID-19 epidemic," *Digit. Diagnostics*, 2020, doi: 10.17816/dd46826.

[25] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015, doi: 10.1109/CVPR.2015.7298965.

[26] D. P. Fan *et al.*, "Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images," *IEEE Trans. Med. Imaging*, 2020, doi: 10.1109/TMI.2020.2996645.