Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
DMDF-Net: Dual multiscale dilated fusion network for accurate segmentation of lesions related to COVID-19 in lung radiographic scans

Muhammad Owais, Na Rae Baek, Kang Ryoung Park *

Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 04620, South Korea

ARTICLE INFO

Keywords:
DMDF-Net
Computer-aided diagnosis
Lung segmentation
COVID-19 lesions segmentation
Infection quantification

ABSTRACT

The recent disaster of COVID-19 has brought the whole world to the verge of devastation because of its highly transmissible nature. In this pandemic, radiographic imaging modalities, particularly, computed tomography (CT), have shown remarkable performance for the effective diagnosis of this virus. However, the diagnostic assessment of CT data is a human-dependent process that requires sufficient time by expert radiologists. Recent developments in artificial intelligence have substituted several personal diagnostic procedures with computer-aided diagnosis (CAD) methods that can make an effective diagnosis, even in real time. In response to COVID-19, various CAD methods have been developed in the literature, which can detect and localize infectious regions in chest CT images. However, most existing methods do not provide cross-data analysis, which is an essential measure for assessing the generality of a CAD method. A few studies have performed cross-data analysis in their methods. Nevertheless, these methods show limited results in real-world scenarios without addressing generality issues. Therefore, in this study, we attempt to address generality issues and propose a deep learning-based CAD solution for the diagnosis of COVID-19 lesions from chest CT images. We propose a dual multiscale dilated fusion network (DMDF-Net) for the robust segmentation of small lesions in a given CT image. The proposed network mainly utilizes the strength of multiscale deep features fusion inside the encoder and decoder modules in a mutually beneficial manner to achieve superior segmentation performance. Additional pre- and post-processing steps are introduced in the proposed method to address the generality issues and further improve the diagnostic performance. Mainly, the concept of post-region of interest (ROI) fusion is introduced in the post-processing step, which reduces the number of false-positives and provides a way to accurately quantify the infected area of lung. Consequently, the proposed framework outperforms various state-of-the-art methods by accomplishing superior infection segmentation results with an average Dice similarity coefficient of 75.7%, Intersection over Union of 67.22%, Average Precision of 69.92%, Sensitivity of 72.78%, Specificity of 99.79%, Enhance-Alignment Measure of 91.11%, and Mean Absolute Error of 0.026.

1. Introduction

Recently, the global disaster of coronavirus disease 2019 (COVID-19) has grieved millions of people and triggered a socioeconomic crisis worldwide. According to the given figures of the WHO (World Health Organization, WHO Coronavirus Disease (COVID-19) Dashboard, 2021), until June 18, 2021, approximately 176,693,988 positive cases of the COVID-19 virus, including 3,830,304 deaths (2.17% mortality rate), have been reported globally. Additionally, the advent of different variants of COVID-19 has further caused an alarming situation worldwide due to their more contagious impact. Regarding the treatment of COVID-19, different experimental vaccines (Kim, Marks, & Clemens, 2021) have finalized clinical assessments and are authorized by the European Medicines Agency (EMA) and/or the Food and Drug Administration (FDA). However, mass production and worldwide distribution of the COVID-19 vaccine is still a challenging and time-consuming task. Until now, preventive measures and early diagnosis have been the only solutions to prevent further spread of this deadly virus. In the context of diagnosis, molecular tests, such as the nucleic acid amplification test (NAAT) and reverse transcription polymerase chain reaction (RT-PCR) are being performed to identify positive cases (Ai et al., 2020). However, these subjective evaluations are performed under strict clinical conditions that can limit the use of these testing methods in outbreak regions.

Recent studies (Ai et al., 2020; Fang et al., 2020) have found that

* Corresponding author.
E-mail address: parkgr@dongguk.edu (K.R. Park).

https://doi.org/10.1016/j.eswa.2022.117360
Received 13 July 2021; Received in revised form 24 January 2022; Accepted 25 April 2022
Available online 2 May 2022
0957-4174/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
chest computed tomography (CT) is a cost-effective diagnostic tool for the identification of COVID-19 infection. Fig. 1 shows a few CT images of different patients infected with the COVID-19 virus. The infected regions are indicated by red boundary lines. The quantitative results presented in (Ai et al., 2020) show that the personal evaluation of COVID-19 infection reached 97% sensitivity, in contrast with RT-PCR testing results based on lung CT images. Similar results in (Fang et al., 2020) demonstrated the diagnostic potential of chest radiography in the initial evaluation of the COVID-19 virus. Moreover, the quantitative evaluation of infection progress inside lung lobes is an important measure for medical treatment (Zhang et al., 2020). Therefore, accurate segmentation of the infected regions is an important pre-processing step for assessing the severity of COVID-19 infection. However, manual evaluation of a large volume of CT scans is a time-consuming task and increases the workload requirements of healthcare professionals.

Recent advances in artificial intelligence technology, especially in the field of medical diagnostics (Hou et al., 2021; Mhiri, Khalifa, Mahjoub, & Rekik, 2020; Xu et al., 2020), have substituted several human-dependent diagnostic approaches with computer-aided diagnosis (CAD) tools. In the present outbreak of COVID-19, these CAD techniques can also support healthcare professionals in making timely and efficient diagnostic decisions using chest CT images. Generally, a CAD method applies a set of artificial intelligence algorithms to analyze the given data, such as CT images, and provides diagnostic results. Recently, a new set of artificial intelligence algorithms called deep learning has emerged that has significantly enhanced the diagnostic capabilities of numerous CAD techniques. These state-of-the-art algorithms can emulate the diagnostic capability of healthcare experts and make effective diagnostic decisions. Recently, convolutional neural networks (CNNs), a new variant of deep learning algorithms, have attained special attention in the development of CAD tools related to various medical domains. However, such CNN-based diagnostic methods need to be trained through supervised learning, which requires a large-scale annotated dataset. In the medical domain, data annotation is accomplished by healthcare experts, which requires sufficient time and resources. To substitute the requirement of a large-scale training dataset, transfer learning (Krizhevsky, Sutskever, & Hinton, 2012) was adopted to train a CNN-based CAD method. In this training approach, a pre-trained CNN (trained on a huge collection of natural images, such as ImageNet (Deng et al., 2009)) can be used in the medical domain. The internal structure of a CNN model consists of a series of convolutional and fully-connected (FC) layers, including other layers, such as batch normalization (BN), softmax, classification, and a rectified linear unit (ReLU) (detailed in (Heaton, 2015)). The convolutional and FC layers included learnable parameters that were initially trained with the training dataset.

In response to the current pandemic, various types of CAD methods exist in the literature. These existing methods mainly use chest X-ray and/or CT images to diagnose COVID-19. Initially, most of the existing methods (Li et al., 2020; Minaee, Kafieh, Sonka, Yazdani, & South, 2020; Owais, Baek, & Park, 2021; Rathod and Kahanja, 2021) used deep classification models to make diagnostic decisions. These methods (Li et al., 2020; Minaee et al., 2020; Owais, Baek, et al., 2021; Rathod and Kahanja, 2021) can only classify positive and negative patients without highlighting the infectious regions in a given radiographic scan. Later, new methods were proposed based on deep segmentation networks, which localized the infected regions in a given radiographic scan. However, most of the existing methods lack cross-data analysis, which is a prime indicator for assessing the effectiveness of a CAD method under real conditions. Limited studies (Ma et al., 2021; Owais et al., 2021; Zhang et al., 2020) have been conducted in which a cross-data analysis of the methods is performed. However, these methods showed limited results in cross-data analysis. Consequently, in this study, we address the limitations of the existing studies and develop a high-performance CAD method for the efficient and well-localized detection of COVID-19 related findings in chest CT images. The main contributions of our method are as follows.

- We propose a dual multiscale dilated fusion network (DMDF-Net) for the robust segmentation of small lesions in chest CT images. Our designed model utilizes the strength of grouped convolution and multiscale deep features fusion inside the encoder and decoder modules using multiscale dilated convolution to achieve better segmentation results with a reduced number of training parameters.
- Additional pre- and post-processing steps are introduced in the proposed method to address the generality issues and obtain superior performance in a real-world setting. Moreover, the post-processing step also provides a way to accurately estimate the proportion of the infected area of the lung (PIAL), which is an essential measure for quantifying the severity of COVID-19 infection.
- Our proposed method achieves state-of-the-art results for the case of cross-data analysis and outperforms various existing methods and recent deep segmentation networks.

Fig. 1. Example CT images of different patients infected with COVID-19 virus from (a) MosMed data and (b) COVID-19-CT-Seg data. The infectious regions of the lungs are shown inside the red boundary lines.
- Finally, we make the proposed framework (including implementation of DMDF-Net, pre-, and post-processing steps) openly available through (https://github.com/Owais786786/DMDF-Net.git, accessed on 18 January 2022) for fair comparisons by other researchers.

The remainder of this article is structured as follows. In Section 2, we briefly review the various existing CAD methods for diagnosing COVID-19 infection using chest radiographic scans. Section 3 presents our selected datasets and proposed method. The training/validation settings and quantitative results are provided in Section 4. Finally, a brief discussion and the conclusions for the proposed framework are presented in Sections 5 and 6, respectively.

2. Related work

In recent literature, various types of diagnostic methods have been proposed to automatically diagnose COVID-19 from chest radiographic scans. These methods primarily focus on CNN-based classification and segmentation models to make diagnostic decisions. In (Li et al., 2020; Minaee et al., 2020; Owais et al., 2021), the authors proposed CNN-based CAD frameworks that mainly classify the given radiographic scan as either positive or negative. Additionally, different training schemes were proposed in (Li et al., 2020; Minaee et al., 2020; Owais et al., 2021) to perform the optimal training of their models using a limited number of training samples. However, the methods in (Li et al., 2020; Minaee et al., 2020; Owais et al., 2021) were trained to classify only positive and negative cases of COVID-19 without detecting and localizing accurate lesion regions in a given radiographic image. In contrast, the semantic segmentation networks performed well in finding infected regions with COVID-19 in each radiographic image. However, pixel-level annotated ground truths are required to properly train and validate these segmentation networks. Such data annotation is performed by healthcare professionals, which requires sufficient time and resources.

To substitute the constraints of large-scale datasets, different semi-supervised learning and data synthesis methods were proposed in literature (Fan et al., 2020; Jiang, Chen, Loew, & Ko, 2021; Zhang et al., 2020). These methods can effectively train deep networks with limited training data. For example, (Jiang et al., 2021) proposed an image synthesis framework based on a conditional generative adversarial network (C-GAN) that can generate radiographic data samples (including both COVID-19 positive and negative CT images) for adequate training of deep networks. In addition, the conventional U-Net segmentation model (Ronneberger, Fischer, & Brox, 2015) was trained with and without using synthesized data to demonstrate the efficiency of their data synthesis approach. Subsequently, (Zhang et al., 2020) presented a new version of C-GAN, called CoSinGAN, that can synthesize high-quality CT images by learning from a single data sample. The experimental results show superior segmentation performance for 2D and 3D U-Net compared to previous reference methods based on the synthesized data of CoSinGAN. In addition, (Fan et al., 2020) proposed a semi-supervised training scheme that effectively trains the proposed deep segmentation model (Inf-Net) using unlabeled data. A novel randomly selected propagation algorithm was adopted to perform the training of Inf-Net using labeled and unlabeled training data. Moreover, the aggregation of high-level features was performed inside Inf-Net to exploit the diverse representations of the lesion regions. Later, (Ma et al., 2021) presented benchmarks for lung lobes and infection segmentation using two radiographic datasets, including CT images. Different segmentation models were trained and evaluated to achieve the best results. 3D U-Net was ranked as the best model among the different reference models. In a comparative study, (Ouichikhi, Aghian, Trongtirakul, & Laouar, 2021) presented a detailed analysis of traditional machine learning techniques in response to the automated diagnosis of COVID-19. Based on a limited number of data samples, the first-ranked machine learning method showed comparable results to a deep CNN model. However, recent comparative studies (Jiang et al., 2021; Li et al., 2020; Owais et al., 2021; Zhang et al., 2020) have proved that deep learning models outperform traditional machine learning methods using multi-source radiographic datasets. Furthermore, (El-Bana, Al-Kabbany, & Sharsak, 2020) developed a multitasking CAD method that comprises a classification and segmentation model to identify and segment certain types of infections in a given CT image. Initially, a pre-trained CNN model was configured to recognize the positive and negative cases of COVID-19. Subsequently, a deep segmentation network (DeepLabV3+ (Chen, Zhu, Papandreou, Schroff, & Adam, 2018)) was included to segment the infectious regions in a given CT image. Similarly, (Zheng et al., 2020) presented a multiscale identification network (MSD-Net) to segment multi-class lesions of different sizes.

In a recent study, (Abdel-Basset, Chang, Hawash, Chakrabortty, & Ryan, 2021) presented a novel segmentation network, FSS-2019-nCov, to substitute the constraints of large-scale training datasets. FSS-2019-nCov contains a dual-path encoder–decoder design that mainly extracts high-level features without changing the channel information. A pre-trained residual network (ResNet54) was configured as an encoder. Later, (Selvaraj, Venkatasesan, Mahesh, & Raj, 2021) developed a CAD framework based on the joint connectivity of a classification and segmentation network, similar to (El-Bana et al., 2020). Additional handcrafted features (i.e., lesion texture and structure information) were also used to efficiently train both networks. Subsequently, (Owais et al., 2021) and (Zhou, Canu, & Ruan, 2021) proposed segmentation-based CAD solutions for the effective detection of minor infectious regions in CT images caused by the COVID-19 virus.

To deal with multi-plane CT data, (Kesavan, Al Naimi, Al Attar, Rajnikanth, & Kadry, 2021) applied a pre-trained ResUNET model for identifying the COVID-19 related lesion regions from the lung CT images with various 2D planes (such as axial, coronal, and sagittal orientations). In another study, (Munusamy, Muthukumar, Gnanaprakasam, Shanmugakani, & Sekar, 2021) proposed a novel CNN model (FractalCovNet) for detecting COVID-19 infection from heterogenous radiographic data (i.e., X-ray and CT images). The proposed model was configured to perform the following two tasks: 1) classifying the lung X-ray images into COVID-19 positive and negative cases; 2) recognizing the COVID-19 related infectious regions from lung CT images. Subsequently, (Voulodimos, Protopapadakis, Katsamenis, Doulamis, & Doulamis, 2021) performed the comparative analysis of two known segmentation models (U-Net and fully convolutional networks (FCNs)) using CT data from COVID-19 patients. Comparative results indicate the following distinctive aspects of FCNs over U-Nets: 1) achieve accurate segmentation despite the class imbalance on the dataset; 2) perform well even in case of annotation errors on the boundaries of symptom manifestation areas. (Zeng, Zheng, & Dong-Ye, 2021) performed the volumetric segmentation of the whole 3D chest CT-scan using an enhanced version of U-Net named 3D CU-Net. An attention mechanism was mainly included in the encoder part of the proposed 3D CU-Net to obtain different levels of the feature representation. Additionally, a pyramid fusion module with expanded convolutions was introduced at the end of the encoder to combine multiscale context information from high-level features. Similarly, (Zhao et al., 2021) proposed a dilated dual attention U-Net (namely D2A U-Net) for accurate detection of COVID-19 related lesion regions in chest CT images. The proposed D2A U-Net utilized the strength of the dual attention strategy to improve feature maps and decrease the semantic gap between different levels of feature maps. Additionally, the hybrid dilated convolutions are included in the decoder part to achieve larger receptive fields, which improves the decoding process. Finally, Table 1 presents a comparative summary of our proposed framework and various existing methods to highlight the superior aspects and limitations of each study.
related to COVID-19 infection are marked as white ‘1.
Moscow Health Care Department. In each CT image, all the findings
All segmentation masks were annotated by medical experts from the
2,049 images and corresponding ground truths as segmentation masks.
son. The MosMed dataset was made available by municipal hospitals in
Moscow, Russia, which includes a total of 50 CT scans of different pa
patients with COVID-19 infection. This dataset includes a total of 3,520
in the annotated mask.
right lung ROI, and infectious regions) were annotated by junior data
for each CT image. All segmentation masks (including left lung ROI,
corresponding annotated mask. Similarly, all the remaining pixels (other
than the lesion regions) are marked as black ‘0.


Table 1
Comparative review summary of the proposed and existing methods on the infection segmentation related to COVID-19.

| Method                              | #SL. (#Pat.) | Strengths                                                                 | Limitations                                                                                     |
|-------------------------------------|--------------|---------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| C-GAN and U-Net (Jiang et al., 2021)| 829(9)       | C-GAN overcomes the underfitting issue                                     | - Lack of cross-data analysis                                                                  |
| CoSinsGAN and 2D U-Net (Zhang et al., 2020)| 5,569(70)   | - CoSinsGAN overcomes the underfitting issue                              | - Data synthesis requires high computation power                                              |
| Inf-Net and Semi-inf-Net (Fan et al., 2020)| 100(40)      | - Detect the infected areas of various sizes                              | - Limited cross-data performance                                                              |
| 3D U-Net (Ma et al., 2021)          | 5,569(70)    | - Semi-supervised learning improves the performance                        | - Lack of cross-data analysis                                                                  |
| Modified local contrast enhancement (Ouleffi et al., 2021)| 275(22)   | - Visualize the progression of disease                                     | - Limited cross-data performance                                                              |
| InceptionV3 and DeepLabV3+ (El-Bana et al., 2020)| 100(40)    | - Computationally cheap                                                   | - Lack of ablation study                                                                       |
| MSD-Net (Zheng et al., 2020)        | 4,780(36)    | - Improve small lesion segmentation                                        | - Lack of cross-data analysis                                                                  |
| FSS-2019-nCov (Abdel-Basset et al., 2021)| 939(69)     | Perform optimal training with a limited dataset                           | - Lack of cross-data analysis                                                                  |
| CNN (Selvaraj et al., 2021)         | 80(N/A)      | Perform optimal training with a limited dataset                           | - Lack of data synthesis required high computation                                            |
| DAL-Net (Owais et al., 2021)        | 5,569(70)    | - Address generality issue                                                | - Lack of cross-data analysis                                                                  |
| U-Net + Attention mechanism (Zhou et al., 2021)| 473(69)     | - Improve small lesion segmentation                                        | - Include pre-processing stage                                                                |
| Res-U-Net (Kesavan et al., 2021)    | 200(10)      | Detect the infected areas in various 2D planes of lung CT slices           | - Limited dataset                                                                            |
| FractalConvNet (Munusamy et al., 2021)| 473(N/A)    | Detection of COVID-19 cases using both chest X-ray and CT images           | - Lack of cross-data analysis                                                                  |
| U-Net and FCNs (Voulodimos et al., 2021)| 939(10)    | - Deal with small lesion segmentation                                      | - Lack of ablation study                                                                       |
| Improved 3D CU-Net (Zheng et al., 2021)| 5,569(70)   | Overcome the effect of class imbalance and annotation errors.             | - Lack of cross-data analysis                                                                  |
| D2A U-Net (Zhao et al., 2021)       | 1,765(N/A)   | Perform well in case of uneven distribution of lesions                     | - Lack of cross-data analysis                                                                  |
| DMDF-Net (Proposed)                | 5,569(70)    | - Computationally efficient                                               | - A large number of parameters                                                                 |

#SL.: Number of CT scan slices; #Pat.: Number of patients.

3. Proposed method

3.1. Datasets and experimental setup

A total of two openly available datasets, MosMed (Morozov, Andreychenko, Pavlov, Vladzymyrskyy, Ledikhova, & Gombolevskiy, 2020) and COVID-19-CT-Seg (Jun, Cheng, Yixin, Xingle, Jiashao, & Ziqi, 2021; Ma et al., 2021), were selected to assess the performance of the proposed DMDF-Net and various baseline networks for a fair comparison. The MosMed dataset was made available by municipal hospitals in Moscow, Russia, which includes a total of 50 CT scans of different patients with COVID-19 infection. The entire dataset includes a total of 2,049 images and corresponding ground truths as segmentation masks. All segmentation masks were annotated by medical experts from the Moscow Health Care Department. In each CT image, all the findings related to COVID-19 infection are marked as white ‘1’ pixels in the corresponding annotated mask. Similarly, all the remaining pixels (other than the lesion regions) are marked as black ‘0’ in the annotated mask. The second dataset, COVID-19-CT-Seg, includes 20 CT scans of different patients with COVID-19 infection. This dataset includes a total of 3,520 images and corresponding ground truths as separate segmentation masks of the left lung ROI, right lung ROI, and infectious regions. Thus, COVID-19-CT-Seg includes three separate segmentation ground truths for each CT image. All segmentation masks (including left lung ROI, right lung ROI, and infectious regions) were annotated by junior data annotators and validated by three medical professionals. Fig. 2 presents a few CT images and their corresponding ground truths for both datasets.

MATLAB (version R2020b), which is a well-known coding framework, was used to implement and simulate the proposed DMDF-Net and other baseline models. All the experiments were performed using a personal desktop computer with an Intel Core i7 CPU with a Nvidia GeForce GPU (GTX 1070), 16-GB RAM, and a Windows 10 operating system.

3.2. Overview of proposed method

As shown in Fig. 3, the proposed CAD framework mainly consists of the following four stages: 1) data pre-processing step; 2) lung segmentation network (DMDF-Net-1); 3) infection segmentation network (DMDF-Net-2); and 4) post-processing step. In the first stage, the color and contrast of the input CT image $I$ are adjusted according to the training dataset by applying a simple Reinhard transformation (RT) (Reinhard, Adhikmin, Gooch, & Shirley, 2001). Mathematically, each testing image $I$ is transformed into an enhanced image $I'$ by applying the transformation $I' = r(I, \phi)$, where $r(\cdot)$ presents the RT as a mapping function and $\phi$ is the mapping parameter that incorporates the color and contrast information of training image. Subsequently, the second and third stages process the enhanced image $I'$ (obtained after pre-processing) using two independent DMDF-Nets, and generate the segmented image of the lung region of interest (ROI) $I_1$ and infectious...
regions $I_2$, respectively. Our proposed DMDF-Nets (named as DMDF-Net-1 and DMDF-Net-2 in Fig. 3) mainly perform semantic segmentation and classify each pixel of input CT image either as black ‘0’ or white ‘1’. In the output of DMDF-Net-1, the white ‘1’ pixels represent the “lung region” and black ‘0’ pixels corresponding to “background.” Similarly, the output of DMDF-Net-2 presents the “infectious region” and “normal/background region” as white ‘1’ and black ‘0’ pixels, respectively. Finally, the post-processing stage further refines $I_2$ (the output of DMDF-Net-2 in the third stage) and generates the final output $I$ by performing post-ROI fusion (i.e., $I = (I_1 \cap I_2) \cup I_1$) of both networks, as shown in Fig. 3. The final output provides well-localized information about the infectious regions inside the lung lobes as $I$ that can be further used for the severity assessment of COVID-19 infection. The addition of the post-processing stage reduces the false-positive pixels in $I_2$ (the output of DMDF-Net-2) and further provides a way to accurately quantify the severity of COVID-19 infection in terms of PIAL score. The PIAL score is calculated by dividing the area of the infected region (i.e., the total number of red pixels in final output image $I$ ) over the total area of lung lobes (i.e., the total number of red and green pixels in final output image $I$ ) as shown in Fig. 3. The subsequent sections present the detailed design, workflow, and selected training loss of the proposed DMDF-Net.

### 3.3. Design of the proposed DMDF-Net

The architecture of our proposed DMDF-Net is designed to meet the following objectives: 1) efficient memory consumption; 2) low number of trainable parameters; and 3) minimum performance degradation in terms of segmentation results. To accomplish these milestones, we primarily utilize the strength of the grouped-convolutional (G-Conv) and dilated convolutional (D-Conv) layers to develop the overall structure of the proposed network. The use of the G-Conv layer results in efficient memory consumption and fast processing speed owing to the decreased number of learnable parameters (Heaton, 2015). In detail, a conventional convolutional layer (Heaton, 2015) processes an input tensor $I_k \in \mathbb{R}^{H \times W \times C}$, respectively.
processing cost of exploiting the deep features from depth of 3) Projection layer: a 1 kernel

Blocks consist of the following three layers. 1) Expansion layer: a 1 (labeled as C-Block in Fig. 4 and Table 2) was also included to exploit in addition, a set of four multiscale D-Conv layers (Chen et al., 2018) (labeled as A-Block and B-Block in Fig. 4 and Table 2) to develop an efficient encoder design. (Sandler, Howard, Zhu, Zhmoginov, 2018) detailed explanation of the encoder/decoder structure and workflow.

The complete layer-wise design of the proposed DMDF-Net is shown in Fig. 4. The network design mainly comprises an encoder part, followed by a decoder module. The encoder part mainly exploits the multiscale deep features without substantially affecting the computation cost (Chen et al., 2018).

Fig. 4. Overall design and workflow diagram of the proposed DMDF-Net (including both encoder and decoder modules).
Table 2
Complete layer-wise structure, configuration, and parametric information of the proposed DMDF-Net.

| Layer       | Input Size | Kernel Size | Kernel Depth | Stride  | Output Size | #Par. |
|-------------|------------|-------------|--------------|---------|-------------|-------|
| Encoder     |            |             |              |         |             |       |
| Image Input | 288 × 352 × 3 | –           | –            | –       | –           | –     |
| Conv 1      | 288 × 352 × 3 | 3^2         | 32           | 2       | 144 × 176 × 32 | 960   |
| G-Conv 1    | 144 × 176 × 32 | 3^2         | 32           | 1       | 144 × 176 × 32 | 384   |
| Conv 2      | 144 × 176 × 32 | 1^2         | 16           | 1       | 144 × 176 × 16 | 560   |
| A-Block 1   | 144 × 176 × 16 | 1^2, 3^2, 1^2 | 96,96,24     | 1,2,1  | 72 × 88 × 24  | 5,352 |
| B-Block 1   | 72 × 88 × 24 | 1^2, 3^2, 1^2 | 144,144,24   | 1,1,1  | 72 × 88 × 24  | 9,144 |
| A-Block 2   | 72 × 88 × 24 | 1^2, 3^2, 1^2 | 144,144,32   | 1,2,1  | 36 × 44 × 32  | 10,520|
| B-Block 2   | 36 × 44 × 32 | 1^2, 3^2, 1^2 | 192,192,32   | 1,1,1  | 36 × 44 × 32  | 15,264|
| A-Block 3   | 36 × 44 × 32 | 1^2, 3^2, 1^2 | 192,192,64   | 1,2,1  | 22 × 88 × 24  | 21,504|
| B-Block 3   | 18 × 22 × 64 | 1^2, 3^2, 1^2 | 384,384,64   | 1,1,1  | 22 × 88 × 24  | 55,104|
| A-Block 4   | 18 × 22 × 64 | 1^2, 3^2, 1^2 | 384,384,320  | 1,1,1  | 18 × 22 × 320 | 154,176|
| C-Block 1   | 18 × 22 × 320 | 1^2, 3^2, 3^2(1,6,12,18)** | 320,320,320,320 | 1,1,1,1 | 18 × 22 × 1024 | 340,992|
| Conv 3      | 18 × 22 × 1024 | 1^2         | 256          | 1       | 18 × 22 × 256 | 262,912|
| Decoder     |            |             |              |         |             |       |
| TP-Conv 1   | 18 × 22 × 256 | 8^2         | 256          | 4       | 72 × 88 × 256 | 4,194,560|
| Conv 4      | 18 × 22 × 144 | 1^2         | 48           | 1       | 72 × 88 × 48  | 7,056 |
| Depth Concatenation | 72 × 88 × 256 | –           | –            | –       | 72 × 88 × 304 | –     |
| G-Conv 2    | 72 × 88 × 304 | 3^2         | 304          | 1       | 72 × 88 × 304 | 3,040 |
| Conv 5      | 72 × 88 × 304 | 1^2         | 256          | 1       | 72 × 88 × 256 | 78,592|
| G-Conv 3    | 72 × 88 × 256 | 1^2         | 256          | 1       | 72 × 88 × 256 | 2,560 |
| Conv 6      | 72 × 88 × 256 | 1^2         | 320          | 1       | 72 × 88 × 320 | 82,880|
| C-Block 2   | 72 × 88 × 320 | 1^2, 3^2, 3^2(1,6,12,18)** | 320,320,320,320 | 1,1,1,1 | 72 × 88 × 1024 | 340,992|
| Conv 7      | 72 × 88 × 1024 | 1^2         | 256          | 1       | 72 × 88 × 256 | 262,912|
| Conv 8      | 72 × 88 × 256 | 1^2         | 2            | 1       | 72 × 88 × 2   | 514   |
| TP-Conv 2   | 72 × 88 × 2   | 8^2         | 2            | 4       | 288 × 352 × 2 | 258   |
| Softmax     | 288 × 352 × 2 | –           | –            | –       | –            | –     |

*Output tensors of these layers are fed to depth concatenation layer; * dilation rate (DR); #Par.: Total number of parameters; \( n \times n \): Not applicable; \( n^2 = n \times n \).

3D tensor of size 18 × 22 × 256. In detail, the first 3 × 3 convolutional layer (labeled as Conv 1 in Table 2) explores the input image \( I^2 \) in both horizontal and vertical directions and converts it into an output tensor of size 144 × 176 × 32. Subsequently, the second and third convolutional layers (labeled as G-Conv 1 and Conv 2 in Table 2) process the output of the previous layer and transform the output of Conv 1 into a tensor of size 144 × 176 × 16. Consequently, a stack of seven structural blocks (labeled as A-Blocks 1,2,3,4 and B-Blocks 1,2,3 in Table 2) consecutively process the output of the previous layer/block to obtain a more diverse representation of the input image as a high-level abstraction. Eventually, these seven structural blocks convert the output tensor of Conv 2 into an output of size 18 × 22 × 320. Additionally, C-Block 1 applies the strength of multiscale dilated convolution and further explores the output of A-Block 4 at four different scales (with DR factors of 1, 6, 12, and 18) and provides diversified multiscale feature maps of size 18 × 22 × 1024 after performing multiscale deep features fusion. For efficient computation on the decoder side, a projection layer (labeled as Conv 3 in Table 2) further transforms these high-level features (output of C-Block 1) in a low-dimensional space. In detail, the Conv 3 layer reduces its (output of C-Block 1) depth by a factor of 4 and gives a final output tensor of size 18 × 22 × 256, which contains diverse semantic information.

3.3.2. Decoder design and workflow

The decoder part of the proposed DMDF-Net mainly includes two transposed convolutional (TP-Conv) layers, one C-Block (labeled as C-Block 2 in Fig. 4 and Table 2), some other layers named Conv and G-Conv, softmax, and pixel classification in Table 2. Our main contribution in the decoder part is the addition of multiscale D-Conv layers (Chen et al., 2018) (labeled as C-Block-2 in Fig. 4 and Table 2), which captures the multiscale representation of deep features from the unsampled output of the TP-Conv 1 layer by performing multiscale deep features fusion and provides additional performance gain. Moreover, a residual connection (extracted from B-Block 1 of the encoder module) was included in the decoded output of the TP-Conv 1 layer (before C-Block 2) to enrich the edge information. Most of the existing studies (Chen et al., 2018; Owais et al., 2021) have shown a better performance of residual skip connection (from encoder to decoder), particularly in the case of minor output objects.

The detailed layer-wise configuration of the decoder module is presented in Table 2. Initially, an 8 × 8 TP-Conv layer (labeled as TP-Conv 1 in Table 2) bilinearly upsamples the encoder final output (output tensor of size 18 × 22 × 256 after Conv 3 layer) with an upsampling factor of 4 and generates an upsampled tensor of size 72 × 88 × 256. Subsequently, a depth concatenation layer combines a residual tensor of size 72 × 88 × 48 (obtained from B-Block 1 and further processed by Conv 4) with the output tensor of TP-Conv 1 and gives a concatenated tensor of size 72 × 88 × 304. Furthermore, a total of four convolutional layers (labeled as G-Conv 2, Conv 5, G-Conv 3, and Conv 6 in Table 2) further transform the output of the previous layer into a new output tensor of size 72 × 88 × 320. Consequently, C-Block 2 applies the strength of multiscale dilated convolution and further processes the output of the preceding layer (Conv 6) at four different scales (with DR factors of 1, 6, 12, and 18) and provides diversified multiscale feature maps of size 72 × 88 × 1024 after performing multiscale deep features fusion. These feature maps are further projected into a low-dimensional space from 72 × 88 × 1024 to 72 × 88 × 2 by processing through two convolutional layers (labeled as Conv 7 and Conv 8 in Table 2). Next, a second 8 × 8
TP-Conv layer (labeled as TP-Conv 2 in Table 2) further upsamples the output of Conv 8 using an upsampling factor of 4 and gives a final output tensor of size $288 \times 352 \times 2$. Finally, a pixel classification layer in conjunction with the softmax layer generates the pixel-wise prediction of the given input as the final output of our model. The softmax layer applies a softmax function (Heaton, 2015) that transforms the output of the TP-Conv 2 layer into a probability. Subsequently, the pixel classification layer provides a class label (either black ‘0’ or white ‘1’) to each pixel of the input CT image and generates a binary image as a final output.

3.4. Loss function

To perform better segmentation of minor lesion regions, a balanced cross-entropy (BCE) loss was selected for the training of the proposed DMDF-Net. BCE shows better performance than conventional cross-entropy (CE) loss (Heaton, 2015; Jadon, 2020), mainly in the case of small segmentation objects or lesion regions (Li et al., 2021; Ni, Wu, Tong, Chen, & Zhao, 2020; Owais et al., 2021; Roh et al., 2018).

Additionally, we took advantage of transfer learning (Krizhevsky et al., 2012) to perform timely and efficient training of the proposed network. The basic structural units of MobileNetV2 (labeled as A-Block and B-Block in Fig. 4) were used to develop the encoder design. Therefore, the initial training parameters of our encoder module (backbone network) were obtained from the MobileNetV2 network that was initially trained with an ImageNet dataset (Deng et al., 2009) using the conventional CE loss function (Heaton, 2015; Jadon, 2020) and stochastic gradient descent (SGD) optimization method (Li, 2018). Therefore, a related variant of the conventional CE loss function named as BCE was selected to perform sufficient training of the proposed DMDF-Net for the target domain. The mathematical interpretation of our selected BCE loss function is given as follows:

$$\arg \min_{\theta} \mathcal{L}_{BCE} = -\frac{1}{N} \sum_{k=1}^{N} \left[ \beta \cdot \log(\psi(I_k, \theta)) + (1 - \beta) \cdot \log(1 - \psi(I_k, \theta)) \right]$$

where $I_k$ and $M_k$ are the $k^{th}$ training data sample and its ground-truth mask, respectively. Subsequently, $\psi(\cdot)$, $N$, and $\theta$ represent the proposed DMDF-Net as a transfer function, the total number of data samples, and the total initial training parameters, respectively. Finally, $\beta$ is the class balancing factor between black ‘0’ and white ‘1’ pixels, and is calculated as the fraction of dominant pixels (i.e., black ‘0’ pixels) in the entire training dataset (Jadon, 2020).

4. Results

In this section, we present a detailed explanation of training, validation, and quantitative results of our method, including a detailed ablation study. Finally, we compare the performance of the proposed DMDF-Net (including both DMDF-Net-1 and DMDF-Net-2) with various state-of-the-art methods.

4.1. Training and validation

Based on existing studies (Kandel & Castelli, 2020; Prabowo & Herwanto, 2019), an SGD optimizer with a small learning rate factor of 0.001 was selected to perform efficient training of the proposed model. Generally, a small learning rate factor can achieve the global minimum; however, it requires many epochs to perform sufficient training of a segmentation network (Johnson & Zhang, 2013). However, a large value of the learning rate factor can skip the global minimum (Johnson & Zhang, 2013). Therefore, a small learning rate factor of 0.001 was selected to achieve optimal convergence of the proposed DMDF-Net. Additionally, we used the default settings provided by MATLAB R2020b for other hyperparameters. The overall training procedure of the proposed DMDF-Net is given in Algorithm 1 as pseudo-code.

**Algorithm 1:** Training procedure of our DMDF-Net (pseudo-code)

| Input: | $(I_k, M_k) \forall k$, $(M_k, M_k)$: $N$ training data samples; $(I_k, M_k)$: $V$ validation data samples; $\theta$: initial training parameters; $\alpha$: initial learning rate; $K$: maximum number of epochs; $B$: mini-batch size; $\beta$: class balancing factor |
|---------|-------------------------------------------------|
| **Initial:** | Initialize training parameters; $\theta$ (ImageNet pretrained weights) |
| **2:** | Split the whole training data into $N/B$ mini-batches of size $B$: $(I_k, M_k), (I_k, M_k), \ldots, (I_k, M_k)$; $(I_k, M_k), (I_k, M_k), \ldots, (I_k, M_k)$ |
| **3:** | for $i = 1, 2, 3, \ldots, N/B$ do // Number of iterations |
| **4:** | $\theta = \theta - \alpha \nabla_{\theta} \mathcal{L}_{BCE}((I_k, M_k), (\hat{M}_k, \psi(I_k, \theta)))$ |
| **5:** | // Loss as Eq. (4) |
| **6:** | Obtain: $(\hat{M}_k)$; Transfer function of our DMDF-Net |
| **7:** | // Check validation accuracy after every epoch to avoid overfitting |
| **8:** | if validation Accuracy$_{\theta, \gamma}$ converged (i.e., not increasing) do // Stop training process for remaining epochs |
| **9:** | end |
| **Output:** | finally trained parameters; $\theta$ // optimal weights of our model |

In the first experiment (hereinafter referred to as Exp#1), we performed lung segmentation using 70% (14/20), 10% (2/20), and 20% (4/20) of the COVID-19-CT-Seg data for training, validation, and testing, respectively. For a fair evaluation in Exp#1, a five-fold cross-validation was performed. Subsequently, in the second experiment (hereinafter referred to as Exp#2), we performed COVID-19 infection segmentation using two different datasets for training, validation, and testing. Such a cross-data analysis highlights the generality of the proposed framework. In Exp#2, we used 80% (16/20) and 20% (4/20) of the COVID-19-CT-Seg dataset for training and validation, respectively, and 100% (50/50) of the MosMed dataset for testing. In the MosMed dataset, the ground truths for lung segmentation are not given; therefore, we did not perform cross-validation in Exp#2 (i.e., using MosMed in training and COVID-19-CT-Seg in testing). Fig. 5 shows the training/validation accuracies and losses of the proposed DMDF-Net for lung segmentation (Exp#1) and COVID-19 infection segmentation (Exp#2). To avoid the overfitting problem, we included independent validation datasets in the training procedure for both Exp#1 and Exp#2. Consequently, we selected the best models based on the maximum validation accuracies for Exp#1 and Exp#2. Thus, the training of DMDF-Net-1 was continued till 15 epochs with a mini-batch size of eight. Subsequently, the second DMDF-Net-2 was trained till 13 epochs with the same mini-batch size as eight. To make a fair comparison, the same datasets and training protocols (such as training hyperparameter and loss function) were adopted to evaluate the results of different baseline models in both experiments. Finally, the quantitative results of the proposed and different baseline models were evaluated using the following seven performance evaluation metrics: the average Dice similarity coefficient (DICE), Intersection over Union (IoU), Average Precision (AP), Sensitivity (SEN), Specificity (SPE), Enhance-Alignment Measure (EAM), and Mean Absolute Error (MAE) (Fan et al., 2018; Fan et al., 2020; Fan, Ji, Qin, & Cheng, 2021; Margolin, Zelnik-Manor, & Tal, 2014, Zhaoxin, Wang, & Zhu, 2020).

4.2. Testing results (Ablation Studies)

Our proposed framework mainly includes a lung segmentation network (DMDF-Net-1 in Fig. 3) and an infection segmentation network (DMDF-Net-2 in Fig. 3) to segment lung lobes and infectious regions in a given CT image, respectively. Primarily, the output of DMDF-Net-1 is used in the post-processing step to refine the output of DMDF-Net-2 and generate well-localized information about the infectious regions in the CT image. Therefore, accurate segmentation of the lung lobes is important in the proposed framework. Thus, we evaluated the quantitative results of both DMDF-Net-1 and DMDF-Net-2, along with a comprehensive analysis of the pre- and post-processing stages as an
addition, these results (Table 3) also highlight the significance of the left, right, and both lung lobes (Exp#1) based on DMDF-Net-1. In particular, the use of C-Blocks and transfer learning gave an average gain with DICE scores of 0.8% [94.86%–94.06%], 3.03% [98.52%–95.49%], and 2.4% [98.66%–96.26%], IoU scores of 1.62% [90.59%–88.97%], 1.78% [97.11%–95.33%], and 1.17% [97.38%–96.21%], and Eφ scores of 0.6% [94.67%–94.07%], 1.07% [98.78%–97.71%], and 0.66% [98.73%–98.07%] for left, right, and both lung lobes, respectively. Moreover, the proposed encoder design (in DMDF-Net-1) contains fewer training parameters (specifically, 0.88 million) than the original MobileNetV2 (specifically, 2.24 million). Table 4 also includes a performance comparison of our selected BCE loss versus conventional CE loss function. In comparison of two different losses (i.e., our BCE loss versus conventional CE loss), our BCE gave an average gain with DICE scores of 0.24% [94.86%–94.62%], 0.13% [98.52%–98.39%], and 0.15% [98.66%–98.51%] and IoU scores of 0.4% [90.59%–90.19%], 0.25% [97.11%–96.86%], and 0.29% [97.38%–97.09%] for left, right, and both lung lobes, respectively. Owing to a slight class imbalance problem between two classes (lung ROI and background), the performance difference of our BCE loss is minimal in comparison with conventional CE loss. However, we observed a significant performance gain of our BCE loss in COVID-19 infection segmentation (Exp#2).

Table 5 presents the quantitative results of the proposed DMDF-Net-2 for COVID-19 infection segmentation (Exp#2), along with a detailed ablation study. It can be observed from Table 5 that the proposed framework shows the best results (75.7%, 67.22%, 69.92%, 72.78%, 99.79%, 91.11%, and 0.026 for average DICE, IoU, AP, SEN, SPE, Eφ, and MAE, respectively) with the addition of pre- and post-processing stages after performing transfer learning. After training the final proposed DMDF-Net-2 (including C-Blocks) for Exp#2 via transfer learning, both pre- and post-processing stages resulted in additional gains of 39.11% [75.7%–36.59%], 38.53% [67.22%–28.69%], 19.89% [69.92%–50.03%], 22.51% [72.78%–50.27%], 42.59% [99.79%–57.2%], 46.07% [91.11%–45.04%], 4.255 [4.281–0.026] for average DICE, IoU, AP, SEN, SPE, Eφ, and MAE, respectively. These results show that transfer learning, pre-processing, post-processing, and multiscale deep features fusion applying multiscale dilated convolution (C-Blocks) work in a mutually beneficial way to enhance the overall performance of the proposed diagnostic framework. Additionally, Fig. 6 shows the visual outputs of the proposed framework with and without including the pre- and post-processing stages for COVID-19 infection segmentation (Exp#2). It can be observed (Fig. 6) that both pre- and post-processing stages mutually contribute to reducing the number of false-positive and false-negative pixels and correctly segment the lesion regions in a given CT image.

We also compared the performance of the proposed encoder design (in DMDF-Net-2) with that of the original MobileNetV2 as a backbone network for COVID-19 infection segmentation (Exp#2). Table 6 presents the comparative results of the proposed versus original MobileNetV2 for Exp#2. It can be observed (Table 6) that both pre- and post-processing steps show higher results for our proposed and original MobileNetV2 compared to those without including any pre- and post-processing steps. However, our model (including both pre- and post-processing stages) still outperforms MobileNetV2, yielding average gains of 2.3% [75.7%–73.4%], 2.04% [67.22%–65.18%], 2.5% [69.92%–67.42%], 1.06%
Table 3
Average five-fold performance of the proposed lung segmentation network (DMDF-Net-1) for lung segmentation (Exp#1). These results also highlight the significance of multiscale deep features fusion using multiscale dilated convolution (C-Blocks) and transfer learning in Exp#1 as an ablation study. (unit: %).

| Training Option | Option (#Par.)  | Dataset | DICE (Std) | IoU (Std) | AP (Std) | SEN (Std) | SPE (Std) | Eφ (Std) | MAE (Std) |
|------------------|-----------------|---------|------------|-----------|----------|-----------|-----------|----------|-----------|
| Without Transfer Learning | C-Blocks (4.71 M) | Right Lung | 43.86 (5.42) | 33.24 (5.38) | 52.38 (1.84) | 63.39 (7.42) | 60.12 (7.93) | 36.84 (4.79) | 3.971 (0.788) |
|                      | Both Lung       | 48.52 (3.85) | 35.75 (3.97) | 54.08 (1.89) | 62.34 (5.29) | 60.3 (5.87) | 43.21 (4.37) | 3.95 (0.578) |
|                      | Left Lung       | 41.88 (3.61) | 31.37 (3.72) | 51.52 (1.12) | 59.0 (5.28) | 57.63 (6.12) | 35.17 (3.9) | 4.231 (0.608) |
|                      | Right Lung      | 44.56 (5.11) | 33.82 (5.24) | 52.43 (1.74) | 62.4 (7.8) | 60.93 (8.3) | 38.2 (5.05) | 3.899 (0.826) |
|                      | Both Lung       | 52.65 (4.69) | 39.74 (4.55) | 56.28 (2.38) | 67.32 (4.53) | 65.28 (5.8) | 47.6 (5.01) | 3.451 (0.564) |
| With Transfer Learning | C-Blocks (4.71 M) | Right Lung | 94.45 (0.65) | 89.92 (1.08) | 90.75 (0.97) | 99.15 (0.17) | 98.89 (0.24) | 94.36 (1.7) | 0.11 (0.023) |
|                      | Both Lung       | 97.76 (0.84) | 95.71 (1.55) | 96.3 (1.11) | 99.1 (1.41) | 99.57 (0.14) | 97.81 (1.22) | 0.046 (0.02) |
|                      | Left Lung       | 98.41 (0.37) | 96.91 (0.57) | 97.6 (0.59) | 99.07 (1.02) | 99.48 (0.08) | 98.29 (0.05) | 0.057 (0.01) |
|                      | Right Lung      | 98.52 (0.24) | 97.11 (0.45) | 97.42 (0.43) | 96.64 (0.28) | 99.71 (0.04) | 98.78 (0.24) | 0.03 (0.004) |
|                      | Both Lung       | 98.66 (0.17) | 97.38 (0.33) | 97.82 (0.26) | 99.57 (0.12) | 99.51 (0.08) | 98.73 (0.35) | 0.048 (0.008) |

#Par.: Number of parameters; M: Million; Std: Standard deviation; The best results are shown in boldface.

Table 4
Comparative results of the proposed encoder design versus original MobileNetV2 as backbone networks and adopted BCE loss versus conventional CE loss function for lung segmentation (Exp#1). (unit: %).

| Backbone/Loss Networks | Dataset | DICE (Std) | IoU (Std) | AP (Std) | SEN (Std) | SPE (Std) | Eφ (Std) | MAE (Std) |
|------------------------|---------|------------|-----------|----------|-----------|-----------|----------|-----------|
| MobileNetV2            | Left Lung | 94.86 (0.4) | 90.59 (0.66) | 91.29 (0.66) | 99.43 (0.15) | 98.98 (0.1) | 94.67 (1.54) | 0.1 (0.01) |
| (Sandler et al., 2018) | Right Lung| 95.49 (1.88) | 95.33 (0.7) | 95.83 (0.5) | 99.4 (0.43) | 99.5 (0.11) | 97.71 (0.51) | 0.05 (0.013) |
| Proposed               | Both Lung | 98.52 (0.24) | 97.11 (0.45) | 97.42 (0.43) | 96.64 (0.28) | 99.71 (0.04) | 98.78 (0.24) | 0.03 (0.004) |
| MobileNetV2            | Left Lung | 98.46 (0.4) | 90.59 (0.66) | 91.29 (0.66) | 99.43 (0.15) | 98.98 (0.1) | 94.67 (1.54) | 0.1 (0.01) |
| (Sandler et al., 2018) | Right Lung| 98.99 (1.63) | 96.26 (0.46) | 96.86 (0.27) | 99.28 (0.39) | 99.51 (0.15) | 98.07 (0.41) | 0.072 (0.016) |
| Proposed               | Both Lung | 98.66 (0.17) | 97.38 (0.33) | 97.82 (0.26) | 99.57 (0.12) | 99.51 (0.08) | 98.73 (0.35) | 0.048 (0.008) |

Loss Functions

| CE (Eleiston, 2015; Jadon, 2020) | Left Lung | 94.62 (0.48) | 90.19 (0.79) | 90.94 (0.83) | 99.25 (0.23) | 98.93 (0.12) | 94.48 (1.6) | 0.105 (0.01) |
| BCE (our)                    | 94.86 (0.4) | 90.59 (0.66) | 91.29 (0.66) | 99.43 (0.15) | 98.98 (0.1) | 94.67 (1.54) | 0.1 (0.01) |
| CE (Eleiston, 2015; Jadon, 2020) | Right Lung| 98.39 (0.24) | 96.86 (0.44) | 97.16 (0.42) | 99.69 (0.25) | 99.68 (0.24) | 98.71 (0.03) | 0.015 (0.005) |
| BCE (our)                    | 98.52 (0.24) | 97.11 (0.45) | 97.42 (0.43) | 96.64 (0.28) | 99.71 (0.04) | 98.78 (0.24) | 0.03 (0.004) |
| CE (Eleiston, 2015; Jadon, 2020) | Both Lung | 98.51 (0.12) | 97.09 (0.22) | 97.55 (0.18) | 99.61 (0.15) | 99.44 (0.15) | 98.75 (0.02) | 0.048 (0.009) |
| BCE (our)                    | 98.66 (0.17) | 97.38 (0.23) | 97.82 (0.26) | 99.57 (0.12) | 99.51 (0.08) | 98.73 (0.35) | 0.048 (0.008) |

#Fold: Fold number; Avg.: Average results; Std: Standard deviation; The best results are shown in boldface.
comparison of pre-ROI fusion versus post-ROI fusion with and without applying the pre-processing step for Exp#2. After training the proposed DMDF-Net-2 through transfer learning (without including the data pre-processing stage), post-ROI fusion outperforms pre-ROI fusion (Table 7), yielding the average gains of 21.04% \( [73.13\% - 52.09\%] \), 15.5% \( [64.96\% - 49.46\%] \), 17.09% \( [68.74\% - 51.65\%] \), 4.16% \( [99.8\% - 95.64\%] \), 15.37% \( [91.13\% - 75.76\%] \), and 0.412 \( [0.439 - 0.027] \) for average DICE, IoU, AP, SEN, SPE, \( E_{\phi} \), and MAE, respectively. In conjunction with the pre-processing stage, post-ROI fusion also outperforms pre-ROI fusion (Table 7), yielding the average gains of 8.24% \( [75.7\% - 67.46\%] \), 6.77% \( [67.22\% - 60.45\%] \), 7.27% \( [69.92\% - 62.65\%] \), 15.84% \( [72.78\% - 56.94\%] \), 0.11% \( [99.79\% - 99.68\%] \), 1.72% \( [91.11\% - 89.39\%] \), and 0.015 \( [0.041 - 0.026] \) for average DICE, IoU, AP, SEN, SPE, \( E_{\phi} \), and MAE, respectively. Similar performance disparities can be observed in Table 7 after training the proposed DMDF-Net-2 from scratch (i.e., without transfer learning). Performance differences of our DMDF-Net-2 of no pre- and post-processing (the 1st and 7th rows of Table 7), only pre-processing (the 2nd and 8th rows of Table 7), combined pre- and post-processing (the 3rd and 9th rows of Table 7), pre- and post-processing with C-Blocks (the 4th and 10th rows of Table 7), and with C-Blocks (the 5th and 11th rows of Table 7) in Exp#2 as an ablation study.

### Table 5

| Training Option | Option(#Par. (M)) | Pre-Processing | DICE | IoU | AP | SEN | SPE | \( E_{\phi} \) | MAE |
|----------------|------------------|----------------|------|-----|----|-----|-----|-------------|-----|
| Without Transfer Learning | Without | × | × | 36.59 | 28.69 | 50.03 | 50.27 | 57.2 | 45.04 | 4.281 |
| | C-Blocks | ✓ | × | 43.37 | 38.06 | 50.02 | 27.14 | 76 | 57 | 2.409 |
| | (4.71) | ✓ | ✓ | 36.48 | 28.53 | 50.05 | 55.5 | 56.86 | 44.87 | 4.314 |
| | With | × | × | 43.42 | 38.11 | 50.04 | 31.84 | 76.06 | 57.04 | 2.403 |
| | C-Blocks | × | ✓ | 33.62 | 25.2 | 50.01 | 53.56 | 50.23 | 40.59 | 4.976 |
| | (5.85) | ✓ | × | 44.99 | 40.65 | 50.02 | 21.61 | 81.2 | 40.5 | 4.948 |
| | ✓ | ✓ | 33.76 | 25.34 | 50.02 | 54.7 | 50.52 | 40.75 | 4.948 |
| | With | × | × | 44.72 | 40.12 | 50.06 | 29.62 | 80.07 | 59.92 | 2.003 |
| | C-Blocks | × | ✓ | 71.94 | 63.98 | 68.99 | 52.13 | 99.84 | 89.98 | 0.026 |
| | (4.71) | ✓ | × | 71.63 | 63.73 | 69.13 | 49.94 | 99.84 | 89.98 | 0.025 |
| | With | × | × | 72.91 | 64.78 | 68.38 | 61.07 | 99.8 | 91.02 | 0.028 |
| | C-Blocks | ✓ | ✓ | 73.13 | 64.96 | 68.74 | 60.66 | 99.8 | 91.13 | 0.027 |
| | (5.85) | ✓ | × | 75.3 | 66.86 | 69.43 | 72.95 | 99.78 | 91.01 | 0.027 |
| | ✓ | ✓ | 75.7 | 67.22 | 69.92 | 72.78 | 99.79 | 91.11 | 0.026 |

×: Not included; ✓: Included; #Par.: Number of parameters; M: Million; The best results are shown in boldface.

Fig. 6. Visual output results of the proposed framework with and without including the pre- and post-processing stages for COVID-19 infection segmentation (Exp#2).
and 11.71% for average DICE, IoU, AP, and SEN, respectively. These
case of the same dataset (COVID-19-CT-Seg). In comparison, the effect of
pre- and post-processing stages gave a marginal gain of 0.48%, 0.5%
post-processing steps. It can be observed ( Table 8 ) that the addition of
posed DMDF-Net-2 by performing the five-fold cross-validation. Table 8
Evaluating the effect of pre- and post-processing stages on
COVID-19 infection segmentation (Exp#2). (unit: %).

| Backbone/Loss | Option                        | DICE | IoU  | AP   | SEN  | SPE  | E₀   | MAE   |
|---------------|-------------------------------|------|------|------|------|------|------|-------|
| MobileNetV2   | Without Pre- and Post-Processing | 73.2 | 65.02| 69   | 59.79| 99.81| 88.44| 0.027 |
| MobileNetV2   | With Pre-Processing | 72.91| 64.78| 68.38| 61.07| 99.8 | 91.02| 0.028 |
| MobileNetV2   | With Pre- and Post-Processing | 73.4 | 65.18| 67.42| 71.72| 99.74| 89.54| 0.031 |
| MobileNetV2   | BCE (our)                      | 75.3 | 66.86| 69.43| 72.95| 99.78| 91.01| 0.027 |
| MobileNetV2   | BCE (our)                      | 73.7 | 67.22| 69.92| 72.78| 99.79| 91.11| 0.026 |

The best results are shown in boldface.

Table 6
Comparative results of the proposed encoder design versus original MobileNetV2 as backbone networks and adopted BCE loss versus conventional CE loss function for COVID-19 infection segmentation (Exp#2). (unit: %).

| Backbone/Loss | Option                        | DICE | IoU  | AP   | SEN  | SPE  | E₀   | MAE   |
|---------------|-------------------------------|------|------|------|------|------|------|-------|
| MobileNetV2   | Without Pre- and Post-Processing | 73.2 | 65.02| 69   | 59.79| 99.81| 88.44| 0.027 |
| MobileNetV2   | With Pre-Processing | 72.91| 64.78| 68.38| 61.07| 99.8 | 91.02| 0.028 |
| MobileNetV2   | With Pre- and Post-Processing | 73.4 | 65.18| 67.42| 71.72| 99.74| 89.54| 0.031 |
| MobileNetV2   | BCE (our)                      | 75.3 | 66.86| 69.43| 72.95| 99.78| 91.01| 0.027 |
| MobileNetV2   | BCE (our)                      | 73.7 | 67.22| 69.92| 72.78| 99.79| 91.11| 0.026 |

Table 7
Quantitative performance comparison of pre-ROI fusion versus post-ROI fusion with and without applying pre-processing step for COVID-19 infection segmentation (Exp#2). (unit: %).

| Training Option | Pre-Processing | Pre-ROI Fusion | Post-ROI Fusion | DICE | IoU  | AP   | SEN  | SPE  | E₀   | MAE   |
|-----------------|----------------|---------------|-----------------|------|------|------|------|------|------|-------|
| Without Transfer Learning | ×   | ×             | ×               | 33.62| 25.2 | 50.01| 53.56| 50.23| 40.59| 4.976 |
|                   | ×               | ×             |                 | 33.76| 25.34| 50.02| 54.7 | 50.52| 40.75| 4.948 |
|                   | ×               | ✓             |                 | 36.08| 28.09| 50.0 | 43.67| 56.05| 44.29| 4.398 |
|                   | ×               | ×             | ✓               | 44.99| 40.65| 50.02| 21.61| 81.2 | 60.5 | 1.892 |
|                   | ✓               | ✓             | ×               | 36.3 | 28.35| 50.0 | 43.38| 56.57| 44.59| 4.345 |
|                   | ✓               | ×             | ✓               | 44.72| 40.12| 50.06| 29.62| 80.07| 59.92| 2.003 |
| With Transfer Learning | ×   | ×             | ×               | 72.91| 64.78| 68.38| 61.07| 99.8 | 91.02| 0.028 |
|                   | ✓               | ×             | ×               | 75.3 | 66.86| 69.43| 72.95| 99.78| 91.01| 0.027 |
|                   | ×               | ✓             | ×               | 52.09| 49.46| 51.65| 77.89| 95.64| 75.76| 0.439 |
|                   | ×               | ✓             | ✓               | 73.13| 64.96| 68.74| 60.66| 99.8 | 91.13| 0.027 |
|                   | ✓               | ✓             | ×               | 67.46| 60.45| 62.65| 56.94| 99.68| 89.39| 0.041 |
|                   | ✓               | ×             | ✓               | 75.7 | 67.22| 69.92| 72.78| 99.79| 91.11| 0.026 |

×: Not included; ✓: Included; The best results are shown in boldface.

processing (the 6th and 12th rows of Table 7) versus only post-processing (the 4th and 10th rows of Table 7) can be observed in Table 7. As shown in this table, we confirm that the combined pre- and post-processing, only pre-processing, only post-processing, and no pre- and post-processing show the 1st-4th highest accuracies, respectively. Subsequently, Fig. 7 presents the visual output results of pre-ROI fusion versus post-ROI fusion (for Exp#2) after applying the pre-processing step and training DMDF-Net-2 through transfer learning. It can be observed (Fig. 7) that the post-ROI fusion-based post-processing step effectively contributes to reducing the number of false-positive and false-negative pixels and correctly segmenting the lesion regions in a given CT image.

We further analyze the effect of pre- and post-processing stages on the same dataset (COVID-19-CT-Seg) to show the comparative results of two different data distributions. In this experiment, we evaluated the average performance of the COVID-19-CT-Seg dataset with our proposed DMDF-Net-2 by performing the five-fold cross-validation. Table 8 shows these comparative results with and without applying pre- and post-processing steps. It can be observed (Table 8) that the addition of pre- and post-processing stages gave a marginal gain of 0.48%, 0.5%, 0.55%, and 0.06% for average DICE, IoU, AP, and SPE, respectively, in case of the same dataset (COVID-19-CT-Seg). In comparison, the effect of pre- and post-processing stages is significantly high in case of cross-dataset (MosMed) with the average gains of 2.79%, 2.44%, 1.54%, and 11.71% for average DICE, IoU, AP, and SEN, respectively. These comparative results (Table 8) show the significant contribution of pre- and post-processing stages in cross-dataset (having different data distributions) and validate the generality of our proposed solution.

Our first MosMed dataset comprises a total of 2,049 images (having a total of 785 infected and 1264 non-infected slices). Owning to a small number of data samples, the number of infected versus non-infected slices influenced the training of our proposed model by causing an under-fitting problem. To address this problem, we utilized the strength of transfer learning (Krizhevsky et al., 2012) to perform timely and efficient training of our model using a small dataset such as COVID-19-CT-Seg in all the experiments. Tables 3 and 5 show the significant gains of transfer learning in Exp#1 and Exp#2, respectively. In addition, we also observed a class imbalance problem (particularly in case of the MosMed dataset) owning to a small ratio of infected lung regions in each infected slice, which ultimately resulted in poor testing results (Table 6). This problem was further addressed by using a BCE loss in all the experiments. Tables 4 and 6 show the comparative performance gains of our BCE loss over the conventional CE loss function in Exp#1 and Exp#2, respectively.

4.3. Comparisons with the State-of-the-Art methods

In this section, we perform a detailed comparative analysis of the
proposed method with state-of-the-art segmentation networks proposed for COVID-19 (Ma et al., 2021; Owais et al., 2021; Zhang et al., 2020) and general image segmentation domains (Badrinarayanan, Kendall, & Cipolla, 2017; Chen et al., 2018; Long, Shelhamer, & Darrell, 2015; Ronneberger et al., 2015; Sandler et al., 2018). The proposed diagnostic framework includes DMDF-Net-1 and DMDF-Net-2 to extract lung ROI (Exp#1), and segment infectious regions (Exp#2) in a given CT image, respectively. Consequently, the performance of both DMDF-Net-1 and DMDF-Net-2 is separately compared with different baseline models. In (Ma et al., 2021; Owais et al., 2021; Zhang et al., 2020), the authors used the same datasets as selected in our method; therefore, we directly compared our method with the given results in (Ma et al., 2021; Owais et al., 2021; Zhang et al., 2020) for both lung segmentation (Exp#1) and COVID-19 infection segmentation (Exp#2). Whereas in the case of other segmentation networks (Badrinarayanan et al., 2017; Chen et al., 2018; Long et al., 2015; Ronneberger et al., 2015; Sandler et al., 2018) proposed for general image segmentation applications, a direct comparison was not possible. Therefore, to make a fair comparison, we evaluated the segmentation results of these models (Badrinarayanan et al., 2017; Chen et al., 2018; Long et al., 2015; Ronneberger et al., 2015; Sandler et al., 2018) using the same datasets as those selected in this study. These baseline models include U-Net (Ronneberger et al., 2015), DeepLabV3+ (based on ResNet (Chen et al., 2018)), MobileNetV2 (Sandler et al., 2018), SegNet (based on VGG16 and VGG19 (Badrinarayanan et al., 2017)), and FCNs (Long et al., 2015).

Table 9 shows the comparative results of the proposed DMDF-Net-1 and all these baseline models for the lung segmentation task (Exp#1). It can be observed (Table 9) that the proposed DMDF-Net-1 shows superior results (in terms of average DICE and IoU scores), including a lower number of training parameters compared to the other models. DAL-Net (Owais et al., 2021) ranked as the second-best network based on the second highest DICE and IoU scores among all the baseline methods. The proposed DMDF-Net-1 outperforms (Owais et al., 2021) (second-best), yielding the average gains of 0.36% [97.35–96.99%], 0.64% [95.03–94.39%], 0.45% [95.51–95.06%], 0.37% [99.55–99.18%], 0.07% [99.4%–99.33%], 0.25% [97.39%–97.14%], and 0.01 [0.07–0.06] for average DICE, IoU, AP, SEN, SPE, $E_\phi$, and MAE, respectively. These results show the average gains of the performance of “left lung”, “right lung”, and “both lung” datasets. Additionally, in a t-test analysis (proposed versus (Owais et al., 2021)), we obtain an average p-value of less than 0.05 (specifically, a p-value of 0.013) that distinguishes our model from (Owais et al., 2021) with a 95% confidence score. In addition, the
Table 9
Quantitative results of the proposed DMDF-Net-1 compared with different state-of-the-art segmentation networks for lung segmentation task (Exp#1). (unit: %).

| Methods                                | #Par. (M) | FLOPs (G) | PT (FPS) | DICE (Std) | IoU (Std) | AP (Std) | SEN (Std) | SPE (Std) | Fp (Std) | MAE (Std) |
|-----------------------------------------|-----------|-----------|----------|------------|-----------|----------|-----------|-----------|----------|-----------|
| 3D U-Net (Ma et al., 2021)              | 85.8      | -         | -        | 88.7       | 97.7     | 98.61    | 93.16    | 0.143     |          |          |
| CoSinGAN + 2D U-Net (Zhang et al., 2020) | 93.9      | -         | -        | 93.9       | 93.3    | 94.06    | 96.46    | 0.072     |          |          |
| SegNet (VGG16) (Badrinarayanan et al., 2017) | 29.44  | 123.91    | 10.31    | 96.47     | 93.36   | 99.27    | 96.54    | 0.072     |          |          |
| SegNet (VGG19) (Badrinarayanan et al., 2017) | 40.07  | 157.54    | 9.62     | 96.44     | 93.3   | 94.06    | 98.28    | 0.072     |          |          |
| U-Net (Encoder Depth:4) (Ronneberger et al., 2015) | 31.03  | 155.69    | 23.26    | 94.78     | 91.01   | 96.69    | 87.89    | 0.091     |          | 0.091 |
| FCN (Up Sampling Factor: 32) (Long et al., 2015) | 134.29 | 187.34    | 18.52    | 91.39     | 85.07   | 86.97    | 96.15    | 0.175     |          | 0.175 |
| DeepLabv3+ (ResNet) (Chen et al., 2018) | 20.61  | 58.42     | 31.25    | 95.78     | 89.97   | 90.89    | 98.85    | 0.107     |          | 0.107 |
| DAL-Net (Owais et al., 2021)            | 6.65     | 35.81     | 23.26    | 95.42     | 90.03   | 90.89    | 99.05    | 0.183     |          | 0.183 |
| DAL-Net (Owais et al., 2021)            | 6.65     | 35.81     | 23.26    | 94.86     | 90.59   | 91.29    | 99.43    | 0.94       |          | 0.94       |
| DMDF-Net-1 (Proposed)                   | 5.85     | 37.45     | 25.64    | 98.66     | 97.38   | 97.82    | 99.57    | 0.096     |          | 0.096     |
| DMDF-Net-1 (Proposed)                   | 5.85     | 37.45     | 25.64    | 98.66     | 97.38   | 97.82    | 99.57    | 0.096     |          | 0.096     |

Right Lung

| Methods                                | #Par. (M) | FLOPs (G) | PT (FPS) | DICE (Std) | IoU (Std) | AP (Std) | SEN (Std) | SPE (Std) | Fp (Std) | MAE (Std) |
|-----------------------------------------|-----------|-----------|----------|------------|-----------|----------|-----------|-----------|----------|-----------|
| 3D U-Net (Ma et al., 2021)              | 87.9      | -         | -        | 89.97     | 90.89    | 98.85    | 94.79    | 0.107     |          | 0.107     |
| CoSinGAN + 2D U-Net (Zhang et al., 2020) | 94.6      | -         | -        | 93.35     | 94.06    | 98.28    | 93.56    | 0.128     |          | 0.128     |
| SegNet (VGG16) (Badrinarayanan et al., 2017) | 29.44  | 123.91    | 10.31    | 96.47     | 93.36   | 99.27    | 96.54    | 0.072     |          | 0.072     |
| SegNet (VGG19) (Badrinarayanan et al., 2017) | 40.07  | 157.54    | 9.62     | 96.44     | 93.3   | 94.06    | 98.28    | 0.072     |          | 0.072     |
| U-Net (Encoder Depth:4) (Ronneberger et al., 2015) | 31.03  | 155.69    | 23.26    | 94.78     | 91.01   | 96.69    | 87.89    | 0.091     |          | 0.091     |
| FCN (Up Sampling Factor: 32) (Long et al., 2015) | 134.29 | 187.34    | 18.52    | 91.39     | 85.07   | 86.97    | 96.15    | 0.175     |          | 0.175     |
| DeepLabv3+ (ResNet) (Chen et al., 2018) | 20.61  | 58.42     | 31.25    | 95.78     | 89.97   | 90.89    | 98.85    | 0.107     |          | 0.107     |
| DAL-Net (Owais et al., 2021)            | 6.65     | 35.81     | 23.26    | 95.42     | 90.03   | 90.89    | 99.05    | 0.183     |          | 0.183     |
| DAL-Net (Owais et al., 2021)            | 6.65     | 35.81     | 23.26    | 94.86     | 90.59   | 91.29    | 99.43    | 0.94       |          | 0.94       |
| DMDF-Net-1 (Proposed)                   | 5.85     | 37.45     | 25.64    | 98.66     | 97.38   | 97.82    | 99.57    | 0.096     |          | 0.096     |
| DMDF-Net-1 (Proposed)                   | 5.85     | 37.45     | 25.64    | 98.66     | 97.38   | 97.82    | 99.57    | 0.096     |          | 0.096     |

#Par.: Number of parameters; M: Million; G: Giga; PT: Processing time; FPS: Frame per second; ”-“: Not available; Std: Standard deviation; The best results are shown in boldface.

number of training parameters of the proposed network is lower than (Owais et al., 2021). To be specific, our DMDF-Net-1 includes approximately 0.8 million fewer parameters than (Owais et al., 2021) (i.e., 5.85 million [proposed] vs. 6.65 million [Owais et al., 2021]). Table 9 also includes the floating point operations per second (FLOPs) and execution speed of our DMDF-Net-1 and other baseline models. Our proposed DMDF-Net-1 requires 37.45 Giga FLOPs with an average execution speed of 25.64 frames per second (FPS).

Figure 8 shows the final segmentation results of the proposed lung segmentation model (DMDF-Net-1) versus different state-of-the-art segmentation networks (Badrinarayanan et al., 2017; Chen et al., 2018; Long et al., 2015; Owais et al., 2021; Ronneberger et al., 2015; Sandler et al., 2018). It can be observed in Fig. 8 that our proposed DMDF-Net-1 and the methods of (Chen et al., 2018; Owais et al., 2021; Sandler et al., 2018) show comparable visual results and outperform the other three baseline methods (Badrinarayanan et al., 2017; Long et al., 2015; Ronneberger et al., 2015). However, the overall quantitative performance of Table 10 of our DMDF-Net-1 is better than all the
baseline methods.

Furthermore, Table 10 shows the comparative results of the proposed DMDF-Net-2 versus different baseline models for the infection segmentation task (Exp#2). It can be observed from Table 10 that the proposed DMDF-Net-2 also provides superior performance and includes a lower number of training parameters compared to other methods. In contrast, DeepLabV3+ (MobileNetV2) (Sandler et al., 2018) is ranked as the second-best network among the other networks. With the addition of only the pre-processing step, the proposed DMDF-Net-2 outperforms (Sandler et al., 2018) (second-best), yielding the average gains of 2.46% [75.3%-72.84%], 2.16% [66.86%-64.7%], 2.67% [69.43%-66.76%], 0.74% [72.95%-72.21%], 0.06% [99.78%-99.72%], 1.66% [91.01%-89.35%], and 0.006 [0.033-0.027] for average DICE, IoU, AP, SEN, SPE, Eρ, and MAE, respectively. Similarly, after including both pre- and post-processing steps (applying the same lung segmentation network of DMDF-Net-1), the proposed DMDF-Net-2 outperforms (Sandler et al., 2018) (second-best), yielding the average gains of 2.14% [75.7%-73.56%], 1.9% [67.22%-65.32%], 2.35% [69.92%-67.57%], 0.8% [72.78%-71.98%], 0.05% [99.79%-99.74%], 1.57% [91.11%-89.54%], and 0.005 [0.031-0.026] for average DICE, IoU, AP, SEN, SPE, Eρ, and MAE, respectively. However, without the pre- and post-processing steps, (Sandler et al., 2018) shows better results than our DMDF-Net-2 with small increases of 0.29% [73.2%-72.91%], 0.24% [65.02%-64.78%], 0.62% [69%-68.38%], 0.01% [99.81%-99.8%], and 0.001 [0.028v0.027] for average DICE, IoU, AP, SPE, and MAE, respectively. Still, our model shows high SEN and Eρ in comparison with (Sandler et al., 2018) (second-best), yielding an average gain in SEN score of 1.28% [61.07%-59.79%], and the Eρ score of 2.58% [91.02%-88.44%], respectively. Additionally, in a t-test analysis (proposed versus (Sandler et al., 2018)), we obtained an average p-value of less than 0.01 (specifically, a p-value of 0.0072) that distinguishes our model from (Sandler et al., 2018) with a 99% confidence score. Similarly, the number of training parameters of the proposed network is also lower than (Sandler et al., 2018). To be specific, our DMDF-Net-2 includes approximately 1.6 million fewer parameters than (Sandler et al., 2018) (i.e., 11.7 million [proposed] ≤ 13.56 million (Sandler et al., 2018)). Consequently, these results (Tables 9 and 10) highlight the superior performance of our model. Table 10 also includes the FLOPs and execution speed of our DMDF-Net-2 and other baseline models. After including both pre- and post-processing steps, our final infection segmentation framework (DMDF-Net-1 and DMDF-Net-2) requires 74.9 Giga FLOPs with an average execution speed of 7.35 frames per second.

Figure 9 presents the visual comparative results of our proposed framework with the other state-of-the-art deep segmentation models. Fig. 9a presents the comparative results without including pre- and post-processing stages. Fig. 9b shows the visual outputs of all the methods by including only pre-processing stage. Finally, Fig. 9c visualizes the comparative performance by including both pre- and post-processing stages (applying the same lung segmentation network). It can be observed (Fig. 9) that the proposed network generates well-localized segmentation outputs for the input CT images. However, several reference models (Badrinarayanan et al., 2017; Chen et al., 2018; Long et al., 2015; Ronneberger et al., 2015) generate inadequate segmentation results, which are marked as false-positive (i.e., incorrectly recognize the normal regions as infectious regions) and false-negative (i.e., not recognizing the infected regions) pixels in Fig. 9. Nevertheless, (Owais et al., 2021) and (Sandler et al., 2018) showed better performance than (Badrinarayanan et al., 2017; Chen et al., 2018; Long et al., 2015; Ronneberger et al., 2015). However, the average segmentation results (Tables 9 and 10) show a higher performance of our method than that of (Owais et al., 2021) and (Sandler et al., 2018). Primarily, the superior performance of our method is attained by the addition of C-Blocks in both encoder and decoder modules, which mainly exploit diverse representations of lung/lesion patterns from the given data by performing multiscale deep features fusion. Moreover, a residual connection (extracted from B-Block 1 of the encoder module) further contributes to the low-level contextual information in the decoding part to refine the edge information of the final output.

5. Discussion

This section describes the distinctive aspects of the proposed method, with possible limitations that can influence the diagnostic performance of our system. Finally, it includes a brief roadmap for future work to address these constraints and improve the overall performance.
lung lobes and to find suitable personalized treatments (Zhang et al., 2020) in the CT image. Accurate detection and quantification of infected areas in the CT image. The output of DMDF-Net-1 is mainly used in a later post-processing step to improve the infection segmentation results of the intermediate outputs (in Fig. 10, after pre-processing, DMDF-Net-1, and post-processing) further present the diagnostic workflow of the proposed framework. In Fig. 10, the PIAL score represents the quantification of the infectious regions in each CT image, which is calculated by dividing the area of the infected region by the total area of lung lobes (i.e., \( \text{PIAL} = \frac{\text{infected lung area}}{\text{total lung area}} \)).

Furthermore, in our network design, we mainly utilized the strength of grouped convolution and multiscale deep features fusion using multiscale dilated convolution (C-Blocks) to achieve better segmentation results with a reduced number of training parameters (specifically, 5.85 million). The encoder design of the proposed DMDF-Net contains fewer training parameters (specifically, 0.88 million) than the original MobileNetV2 (which showed the second-best accuracies in Table 2). The average execution time of our DMDF-Net is approximately 25.64 frames per second, while that of the original MobileNetV2. Specifically, the average execution time of our DMDF-Net is 2.24 million). Owing to the optimal size of the encoder design of the proposed DMDF-Net contains fewer training parameters (specifically, 5.85 million). The encoder design of the proposed DMDF-Net contains fewer training parameters (specifically, 0.88 million) than the original MobileNetV2. Consequently, the optimal design of our model achieves state-of-the-art performance and utilizes low-cost hardware resources without influencing the overall diagnostic performance.

Consequently, the optimal design of our model achieves state-of-the-art performance and utilizes low-cost hardware resources without influencing the overall diagnostic performance.

To visualize the internal workflow of the proposed DMDF-Net, we also show the multiscale class activation maps (CAMs) (Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016) extracted from five different layers (labeled as Conv 2, B-Block 1, B-Block 2, Conv 3, and Conv 8 in Table 2).
Fig. 9. Visual comparison of COVID-19 infection segmentation results of the proposed framework with the other state-of-the-art deep segmentation models (a) without pre- and post-processing, (b) with pre-processing, and (c) with pre- and post-processing (applying the same lung segmentation network).
of the network. The input image of size $288 \times 352$ was downsampled into four spatial sizes (i.e., $144 \times 176$, $72 \times 88$, $36 \times 44$, and $18 \times 22$) after passing through the encoder module. Subsequently, the encoded output of size $18 \times 22$ is upsampled into two spatial sizes (i.e., $72 \times 88$ and $288 \times 352$) after passing through the decoder module. The decoded output with a size of $288 \times 352$ is the final output of the proposed network. Therefore, a total of five layers were selected for multiscale CAM visualization based on the distinctive spatial sizes of their outputs inside the encoder and decoder modules. **Fig. 11** shows the multiscale CAM visualization of the proposed DMDF-Net-1 (lung segmentation network) and DMDF-Net-2 (infection segmentation network) for testing the CT images. It can be observed (Fig. 11) that the class-specific regions (lung ROIs or infectious regions) become increasingly discriminative after passing through successive layers. Finally, a binary image is obtained as the final output that presents the “Lung/infectious region” and “normal/background region” as white ‘1’ and black ‘0’ pixels, respectively.

Although several online datasets are available related to COVID-19,
most datasets are related to the classification problem. A few segmentation datasets related to COVID-19 are publicly available, which only include the segmentation masks for infectious regions. This study presents a CAD framework for automatic detection and quantification of COVID-19 related findings in lung CT scans. The proposed method includes an additional post-processing step that also requires a lung segmentation mask for accurate segmentation and quantification of infected lung areas. Therefore, we selected the COVID-19 CT Seg dataset that includes the ground truths for both lung and infection regions of each slice. Secondly, we aimed to develop a CAD tool to segment trivial infected regions in the lung efficiently. Therefore, we selected the MosMed dataset that includes trivial lung tissue abnormalities with COVID-19 (pulmonary parenchymal involvement $< 25\%$) (Morozov et al., 2020). In our future work, we will explore additional...
supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

References

Abdel-Rassoul, M., Chang, V., Hawash, H., Chakraborty, R. K., & Ryan, M. (2021). FSS-2019 arGo: A deep learning architecture for semi-supervised low-shot segmentation of COVID-19 infection. Knowledge-Based Systems, 212, Article 106647.

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., et al. (2020). Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. Radiology, 296, e23-e40.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 2481-2495.

Chen, L. C., Zha, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (pp. 801-818).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255).

El-Bana, S., Al-Kabbany, A., & Sharkas, M. (2020). A multi-task pipeline with specialized streams for classification and segmentation of infection manifestations in covid-19 scans. PeerJ Computer Science, 6, Article e303.

Fan, D. P., Gong, C., Cao, Y., Ren, B., Cheng, M.-M., & Borji, A. (2018). Enhanced-alignment measure for binary foreground map evaluation. In Proceedings of the 17th International Joint Conference on Artificial Intelligence (pp. 698-704).

Fan, P. D., Zhou, T., Ji, G. P., Zhou, Y., Chen, G., Fu, H., et al. (2020). InfNet: Automatic covid-19 lung infection segmentation from CT images. IEEE Transactions on Medical Imaging, 39, 2625-2637.

Fan, D. P., Li, G. P., Qin, X., & Cheng, M. M. (2021). Cognitive vision inspired object segmentation metric and loss function. Scientia Sinica Informationis, 6.

Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., et al. (2020). Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. Radiology, 296, E115-E117.

Heaton, J. (2015). Artificial Intelligence for Humans, Deep learning and neural networks. An IEEE Press Book.

Hou, W., Wang, L., Gai, S., Lin, Z., Yu, R., & Qin, J. (2021). Early neoplasm identification in Barrett’s esophagus via attentive hierarchical aggregation and self-distillation. Medical Image Analysis, 72, Article 102992.

Jadon, S. (2020). A survey of loss functions for semantic segmentation. In Proceedings of the IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology, Via del Mar, Chile, 27-29 October, pp. 1–7.

Jang, Y. V., Chen, H., Lowy, M. H., & Ko, H. (2021). COVID-19 CT image synthesis with a conditional generative adversarial network. IEEE Journal of Biomedical and Health Informatics, 25, 441-452.

Johnson, R., & Zhang, T. (2013). Accelerating stochastic gradient descent using predictive variance reduction. Advances in Neural Information Processing Systems, 26, 315-323.

Jun, M., Cheng, G., Yinxi, W., Xingle, A., Jianfao, G., Ziqi, Y., et al. (2021). COVID-19 CT lung and infection segmentation dataset (Version 1.0).Zenodo. http://doi.org/10.5281/zenodo.3757476.

Kadali, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. ICT Express, 6, 312-315.

Kesavan, S. M., Al Naimi, L., Al Attar, F., Rajeswaran, Y., & Kedry, S. (2021). Res-UNet supported segmentation and evaluation of COVID19 lesion in lung CT. In Proceedings of International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1–4).

Kim, J. H., Marks, F., & Clemens, J. D. (2021). Looking beyond COVID-19 vaccine phase 3 trials. Nature Medicine, 27, 205-211.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097-1105.

Li, X., Yu, L., Chen, H., Fu, C. W., Xing, L., & Heng, P. A. (2021). Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Transactions on Medical Imaging, 32, 521-534.

Li, X. L. (2018). Preconditioned stochastic gradient descent. IEEE Transactions on Neural Networks and Learning Systems, 29, 1454-1466.

Li, Y., Wei, D., Chen, J., Cao, S., Zhan, H., Zha, Y., et al. (2020). Efficient and effective training of COVID-19 classification networks with self-supervised dual-track learning to rank. IEEE Journal of Biomedical and Health Informatics, 24, 2787-2797.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3431-3440).

Ma, J., Wang, V., An, X., Ge, C., Yu, Z., Chen, J., et al. (2021). Towards data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation. Medical Physics, 48, 1197-1210.

Margolin, R., Zelnik-Manor, L., Tal, A., 2014. How to evaluate foreground maps? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA, 23-28 June, pp. 248-255.

Mhiri, I., Khalifa, A. B., Mahjoub, M. A., & Rekik, I. (2020). Brain graph super-resolution for boosting neurological disorder diagnosis using unsupervised multi-topology convolutional brain template learning. Medical Image Analysis, 65, Article 102768.

Minea, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, G. J. (2020). Deep-oncot: Predicting covid-19 from chest x-ray images using deep transfer learning. Medical Image Analysis, 65, Article 101794.
Morozov, S.P., Andreychenko, A.E., Pavlov, N.A., Vladymyrs'kyi, A.V., Led'kova, N.V., Gombolevskiy, V.A., et al., 2020. Mosmeddata: Chest CT scans with covid-19 related findings dataset. arXiv preprint arXiv:2005.06465.

Munusamy, H., Muthukumar, K. J., Ganaaprabakam, S., Shannugakani, T. R., & Sekar, A. (2021). FractalCovNet architecture for COVID-19 Chest X-ray image classification and CT-scan image segmentation. Biocybernetics and Biomedical Engineering, 41, 1025–1038.

Ni, J., Wu, J., Tong, J., Chen, Z., & Zhao, J. (2020). GC-Net: Global context network for medical image segmentation. Computer Methods and Programs in Biomedicine, 190, Article 105121.

Oulefki, A., Agaian, S., Trongtirakul, T., & Laouar, A. K. (2021). Automatic COVID-19 lesions and its classification from CT images: A survey. In 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 25–27 June, pp. 1–8.

Rathod, S.R., Khanuja, H.K., 2021. Automatic segmentation of COVID-19 pneumonia lesions and its classification from CT images: A survey. In 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 25–27 June, pp. 1–8.

Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. IEEE Computer Graphics and Applications, 21, 34–41.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234–241).

Roth, H. R., Oda, H., Zhou, X., Shimizu, N., Yang, Y., Hayashi, Y., et al. (2018). An application of cascaded 3D fully convolutional networks for medical image segmentation. Computerized Medical Imaging and Graphics, 66, 90–98.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June, pp. 4510–4520.

Selvaraj, D., Venkatesan, A., Mahesh, V. G., & Raj, A. N. J. (2021). An integrated feature frame work for automated segmentation of COVID-19 infection from lung CT images. International Journal of Imaging Systems and Technology, 31, 28–46.

Voulodimos, A., Protopapadakis, E., Katsamakis, I., Douiamas, A., & Doulamis, N. (2021). Deep learning models for COVID-19 infected area segmentation in CT images. In Proceedings of the 14th IFIP/IEEE Technologies Related to Assistive Environments Conference (pp. 404–411).

World Health Organization, WHO Coronavirus Disease (COVID-19) Dashboard. Available online: https://covid19.who.int/ (accessed on 18 June 2021).

Xu, X., Wang, C., Guo, J., Gan, Y., Wang, J., Bai, H., et al. (2020). MSCS-DeepLN: Evaluating lung nodule malignancy using multi-scale cost-sensitive neural networks. Medical Image Analysis, 65, Article 101772.

Zhang, J., Meng, G., Li, W., Shi, B., Dong, H., Su, Z., et al. (2020). Relationship of chest CT score with clinical characteristics of 108 patients hospitalized with COVID-19 in Wuhan, China. Respiratory Research, 21, 180.

Zheng, R., Zheng, Y., & Dong-Ye, C. (2021). Improved 3D U-Net for COVID-19 chest CT image segmentation. Scientific Programming, 2021, 9999968.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June, pp. 2921–2929.

Zhou, T., Canu, S., & Ruan, S. (2021). Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. International Journal of Imaging Systems and Technology, 31, 16–27.