ADID-UNET - A segmentation model for COVID-19 infection from lung CT scans

Alex Noel Joseph Raj [Equal first author, 1], **Haipeng Zhu** [Equal first author, 1], **Asiya Khan** [2], **Zhemin Zhuang** [Corresp., 1], **Zengbiao Yang** [1], **Vijayalakshmi G V Mahesh** [3], **Ganesan Karthik** [4]

[1] Department of Electronic Engineering, College of Engineering, Shantou University, Shantou, China

[2] School of Engineering, Computing and Mathematics University of Plymouth Plymouth PL4 8AA, Plymouth, UK

[3] Department of Electronics and Communication, BMS Institute of Technology and Management, Bangalore, India

[4] COVID CARE - Institute of Orthopedics and Traumatology, Madras Medical College, Chennai, India

Corresponding Author: Zhemin Zhuang
Email address: zmzhuang@stu.edu.cn

Currently, the new coronavirus disease (COVID-19) is one of the biggest health crises threatening the world. Automatic detection from computed tomography (CT) scans is a classic method to detect lung infection, but it faces problems such as high variations in intensity, indistinct edges near lung infected region and noise due to data acquisition process.Therefore, this paper proposes a new COVID-19 pulmonary infection segmentation depth network referred as the ADID-UNET (Attention Gate-Dense Network- Improved Dilation Convolution-UNET). The dense network replaces convolution and maximum pooling function to enhance feature propagation and solves gradient disappearance problem. An improved dilation convolution is used to increase the receptive field of the encoder output to further obtain more edge features from the small infected regions. The integration of attention gate into the model suppresses the background and improves prediction accuracy. The experimental results show that the ADID-UNET model can accurately segment COVID-19 lung infected areas, with performance measures greater than 80% for metrics like Accuracy, Specificity and Dice Coefficient (DC). Further when compared to other state-of-the-art architectures, the proposed model showed excellent segmentation effects with a high DC and F1 score of 0.8031 and 0.82 respectively.

ADID-UNET - A segmentation model for COVID-19 infection from lung CT scans

Alex Noel Joseph Raj[1], Haipeng Zhu[1], Asiya Khan[2], Zhemin Zhuang[1], Zengbiao Yang[1], Vijayalakshmi G V Mahesh[3], and Ganesan Karthik[4]

[1]**Department of Electronic Engineering, College of Engineering, Shantou University, Shantou, China**
[2]**School of Engineering, Computing and Mathematics, University of Plymouth Plymouth PL4 8AA**
[3]**Department of Electronics and Communication, BMS Institute of Technology and Management, Bangalore**
[4]**COVID CARE - Institute of Orthopedics and Traumatology, Madras Medical College, Chennai India**

Corresponding author:
Zhemin Zhuang[1]

Email address: zmzhuang@stu.edu.cn

ABSTRACT

Currently, the new coronavirus disease (COVID-19) is one of the biggest health crises threatening the world. Automatic detection from computed tomography (CT) scans is a classic method to detect lung infection, but it faces problems such as high variations in intensity, indistinct edges near lung infected region and noise due to data acquisition process.Therefore, this paper proposes a new COVID-19 pulmonary infection segmentation depth network referred as the ADID-UNET (Attention Gate-Dense Network- Improved Dilation Convolution-UNET). The dense network replaces convolution and maximum pooling function to enhance feature propagation and solves gradient disappearance problem. An improved dilation convolution is used to increase the receptive field of the encoder output to further obtain more edge features from the small infected regions. The integration of attention gate into the model suppresses the background and improves prediction accuracy. The experimental results show that the ADID-UNET model can accurately segment COVID-19 lung infected areas, with performance measures greater than 80% for metrics like accuracy, specificity and Dice Coefficient (DC). Further when compared to other state-of-the-art architectures, the proposed model showed excellent segmentation effects with a high DC and F1 score of 0.8031 and 0.82 respectively.

1 INTRODUCTION

COVID-19 has caused a worldwide health crisis. The World Health Organization (WHO) announced COVID-19 as a pandemic on March 11, 2020. The clinical manifestations of COVID-19 range from influenza-like symptoms to respiratory failure (i.e. diffuse alveolar injury) and its treatment requires advanced respiratory assistance and artificial ventilation. According to the global case statistics from the Center for Systems Science and Engineering (CSSE) of Johns Hopkins University (JHU) (Wang et al., 2020) (updated August 30, 2020), 24,824,247 confirmed COVID-19 cases, including 836,615 deaths, have been reported so far with pronounced effect in more than 180 countries. COVID-19 can be detected and screened by Reverse Transcription Polymerase Chain Reaction (RT-PCR). However, the shortage of equipment and the strict requirements on the detection environment limit the rapid and accurate screening of suspected cases. Moreover, the sensitivity of RT-PCR is not high enough, resulting in a large number of false-negatives (Ai et al., 2020), which presents early detection and treatment of patients with presumed COVID-19 (Fang et al., 2020). As an important supplement to RT-PCR, CT scans clearly describe the characteristic lung manifestations related to COVID-19 (Chung et al., 2020), the early Ground Glass Opacity (GGO), and late lung consolidation are shown in Figure 1. Nevertheless, CT scans also show

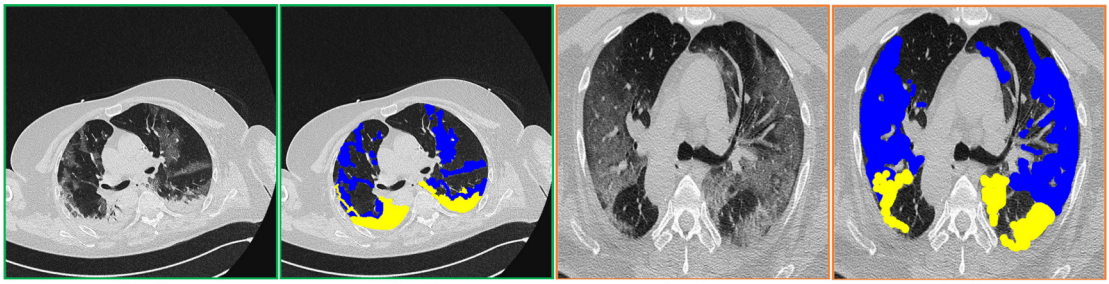

Figure 1. (A1) and (A2) represent the CT images and (B1) and (B2) correspond to the COVID-19 infected areas in CT axial section. Here the blue and yellow masks represent the Ground Glass Opacity (GGO) and the late lung consolidation segments respectively. The images were obtained from (Medseg.ai, 2020).

imaging features that are similar to other types of pneumonia, making it difficult to differentiate them. Moreover, the manual depiction of lung infection is a tedious and time-consuming job, which is often influenced by personal bias and clinical experience.

In recent years, deep learning has been gaining popularity in the field of medical imaging due to it's intelligent and efficient feature extraction ability (Kong et al., 2019; Ye et al., 2019), and has achieved great success. An earliest classic example is the application of deep learning to children's chest X-rays to detect and distinguish bacterial and viral pneumonia (Kermany DS, 2018; Rajaraman et al., 2018). Also using deep learning methods have been applied to detect various imaging features of chest CT images (Depeursinge et al., 2015; Anthimopoulos et al., 2016). Recently, researchers proposed to detect COVID-19 infections in patients by radiation imaging combined with deep learning technology. Li et al. (2020) proposed a simple Cov-Net deep learning network in combination with a deep learning algorithm, which was used to distinguish COVID-19 and Community-Acquired Pneumonia (CAP) from chest CT scans. Wang and Wong (2020) proposed Covid-Net to detect COVID-19 cases from chest X-ray images, with an accuracy rate of 93.3%. The infection probability of COVID-19 Xu et al. (2020) was calculated from CT scans by adopting a position-oriented attention model that presented accuracy close to 87%. However, the above models rarely involved the segmentation of COVID-19 infection (Chaganti et al., 2020; Shan et al., 2020). The challenges involved in segmentation include: **a)** variations in texture, size, and position of the infected areas in CT scans. For example, some infection areas are small, which easily lead to a high probability of false negatives in CT scans. **b)** The boundary of GGO is usually of low contrast and fuzzy in appearance, which makes it difficult to distinguish from the healthy regions during the segmentation process. **c)** The noise around the infected area is high, which greatly affects the segmentation accuracy and **d)** finally the cost and time consumed in obtaining high-quality pixel-level annotation of lung infection in CT scans is high. Therefore, most of the COVID-19 CT scan datasets are focused on diagnosis, and only a few of them provide segmentation labels. However, with the passage of time, the annotated datasets for the segmentation of COVID-19 pulmonary infection were released but due to a lesser amount of data, the phenomenon of overfitting could cause problems while training thus necessitating the need for more segmentation datasets and better algorithms for accurate results.

Therefore, to address the challenges stated above, we propose a new deep learning network called Attention Gate-Dense Network- Improved Dilation Convolution-UNET (ADID-UNET) for the segmentation of COVID-19 from lung infection CT scans. Experimental results on a publicly available dataset illustrate that the proposed model presents reliable segmentation results that are comparable to the ground truths annotated by experts. Also, in terms of performance, the proposed model surpasses other state-of-the-art segmentation models, both qualitatively and quantitatively.

Our contributions in this paper are as follows:

1. To address the problem that the gradient disappearance in the deep learning network pose, we employ a dense network (Huang et al., 2017) instead of a traditional convolution and max-pooling operations. The dense network extracts dense features and enhances feature propagation through the model. Moreover, the training parameters of the dense network are less, which reduces the size and the

computational cost.

2. To increase the size of the respective field and to compensate for the problems due to blurry edges, an improved dilation convolution (IDC) module is used to connect the encoder and decoder pipelines. The IDC model increases the receptive field of the predicted region providing more edge information, which enhances the edge recognition ability of the model.

3. Since the edge contrast of GGO is very low, we use the attention gate (AG) instead of simple cropping and copying. This further improves the accuracy of the model to detect the infection areas by learning the characteristics of the infected regions.

4. Due to the limited number of COVID-19 segmented datasets with segmentation labels, which is less than the minimum number of samples required for training a complex model, we employ data augmentation techniques and expand the dataset on the basis of the collected public datasets.

The rest of the paper is organized as follows: Section 2 describes the work related to the proposed model. Section 3 introduces the basic structure of ADID-UNET. Details of the dataset, experimental results and discussion are dealt with in section 4. Finally, section 5 presents the conclusion.

2 RELATED WORK

ADID-UNET model proposed in this paper is based on UNET (Ronneberger et al., 2015) architecture and therefore, we will discuss the literature related to our work which includes: deep learning and medical image segmentation, improvement of medical image segmentation algorithms, CT scan segmentation, and application of deep learning in segmentation of COVID-19 lesions from lung CT scans.

2.1 Deep learning and medical image segmentation

In recent years, deep learning algorithms have become more mature leading to various artificial intelligence (AI) systems based on deep learning algorithms being developed. Also, semantic segmentation using deep learning algorithms (Oktay et al., 2018) has developed rapidly with applications in both natural and medical images. Long et al. (2015) pioneered the use of a fully connected CNN (FCN) to present rough segmentation outputs that were of the input resolution through fractionally strided convolution process also referred as the upsampling or deconvolution. The model was tested on PASCAL VOC, NYUDv2, and SIFT datasets and, presented a Mean Intersection of Union (M-IOU) of 62.7%, 34%, 39.5%, respectively. They also reported that upsampling, part of the in-network, was fast, accurate, and provided dense segmentation predictions. Later through a series of improvements and extensions to FCN (Ronneberger et al., 2015; Badrinarayanan et al., 2017; Xu et al., 2018), a symmetrical structure composed of encoder and decoder pipelines, called UNET (Ronneberger et al., 2015), was proposed for biomedical or medical image segmentation. The encoder structure predicted the segmentation area, and then the decoder recovered the resolution and achieved accurate spatial positioning. Also, the UNET used crop and copy operations for the precise segmentation of the lesions. Further, the model achieved good segmentation performance at the International Symposium on Biomedical Imaging (ISBI) challenge (Cardona et al., 2010) with the M-IOU of 0.9203. Moreover, an improved network referred as the SegNet was proposed by Badrinarayanan et al. (2017). The model used the first 13 convolution layers of the VGG16 network (Karen and Andrew, 2014) to form an encoder to extract features and predict segmentation regions. Later by using a combination of convolution layers, unpooling and softmax activation function in the decoder, segmentation outputs of input resolution were obtained. When tested with the CamVid dataset (Brostow et al., 2009), the M-IOU index of SegNet was nearly 10% higher than that of FCN (Long et al., 2015). Xu et al. (2018) regarded segmentation as a classification problem in which each pixel was associated with a class label and designed a CNN network composed of three layers of convolution and pooling, a fully connected layer (FC) and softmax function. The model of successfully segmented three-dimensional breast ultrasound (BUS) image datasets was presented into four parts: skin, fibroglandular tissue, mass, and fatty tissue and achieved a recall rate of 88.9%, an accuracy of 90.1%, precision of 80.3% and F1 score of 0.844 . According to the aforementioned literature, FCN (Long et al., 2015) and their improved variants presented accurate segmentation results for both natural or medical images. Therefore, the UNET and variants (Almajalid et al., 2019; Negi et al., 2020), due to its advantages of fast training and high segmentation accuracy are widely used in the field of medical image segmentation.

2.2 Improvement of medical image segmentation algorithms

Medical images such as the ultrasound images are generally prone to speckle noise, uneven intensity distribution, and low contrast between the lesions and the backgrounds which affect the segmentation ability of the traditional UNET (Ronneberger et al., 2015) structure. Therefore, considerable efforts were invested in improving the architecture. Xia and Kulis (2017) proposed a fully unsupervised deep learning network called W-Net model that connects two UNETs to predict and reconstruct the segmentation results. Schlemper et al. (2019) proposed an attention UNET network, which integrated attention modules into the UNET (Ronneberger et al., 2015) model to achieve spatial positioning and subsequent segmentation. The model presented a segmentation accuracy of 15% higher than the traditional UNET architecture. Zhuang et al. (2019a) combined the goodness of the attention gate system and the dilation convolution module and proposed a hybrid architecture referred as the RDA-UNET. By introducing residual network (He et al., 2016) instead of traditional convolution layers they reported a segmentation accuracy of 97.91% towards the extraction of lesions in breast ultrasound images. Also, the GRA-UNET (Zhuang et al., 2019b) model included a group convolution module in-between the encoder and decoder pipelines to improve the segmentation of the nipple region in breast ultrasound images. Therefore, from the literature,it can be inferred that introducing additional modules like attention gate instead of traditional cropping and copying, inclusion of dilation convolution to increase the receptive fields and use of residual networks can favorably improve the accuracy of the segmentation model. However, these successful segmentation models (Schlemper et al., 2019; Zhuang et al., 2019a; Xia and Kulis, 2017) were rarely tested with CT scans, hence the next section concentrates on the segmentation of CT scans.

2.3 CT scan segmentation

CT imaging is a commonly used technology in the diagnosis of lung diseases since lesions can be segmented more intuitively from the chest CT scans. The segmented lesion aid the specialist in the diagnosis and quantification of the lung diseases(Gordaliza et al., 2018). In recent years, most of the classifier models and algorithms based on feature extraction have achieved good segmentation results in chest CT scans. Ye et al. (2009) proposed a shape-based Computer-Aided Detection (CAD) method where a 3D adaptive fuzzy threshold segmentation method combined with chain code was used to estimate infected regions in lung CT scans. In feature-based techniques, due to the low contrast between nodules and backgrounds, the boundary discrimination is unclear leading to inaccurate segmentation results. Therefore, many segmentation techniques based on deep learning algorithms have been proposed. Wang et al. (2017) developed a central focusing convolutional neural network for segmenting pulmonary nodules from heterogeneous CT scans. Jue et al. (2018) designed two deep networks (an incremental and dense multiple resolution residually connected network) to segment lung tumors from CT scans by adding multiple residual flows with different resolutions. Guofeng et al. (2018) proposed a UNET model to segment pulmonary nodules in CT scans which improved the overall segmentation output through the avoidance of overfitting. Compared with other segmentation algorithms such as graph-cut (Ye et al., 2009), their model had better segmentation results with a Dice coefficient of 0.73. Recently, Peng et al. (2020) proposed an automatic CT lung boundary segmentation method, called Pixel-based Two-Scan Connected Component Labeling-Convex Hull-Closed Principal Curve method (PSCCL-CH-CPC). The model included the following: **a)** the image preprocessing step to extract the coarse lung contour and **b)** coarse to finer segmentation algorithm based on the improved principal curve and machine learning model. The model presented good segmentation results with Dice coefficient as high as 96.9%. Agarwal et al. (2020) proposed a weakly supervised lesion segmentation method for CT scans based on an attention-based co-segmentation model (Mukherjee et al., 2018). The encoder structure composed of a variety of CNN architectures that includes VGG-16 (Karen and Andrew, 2014), Res-Net101(He et al., 2016), and an attention gate module between the encoder-decoder pipeline, while decoder composed of upsampling operation. The proposed method first generated the initial lesion areas from the Response Evaluation Criteria in Solid Tumors (RECIST) measurements and then used co-segmentation to learn more discriminative features and refine the initial areas. The paper reported a Dice coefficient of 89.8%. The above literatures suggest that deep learning techniques are effective in segmenting lesions in lung CT scans and many researchers have proposed different deep learning architectures to deal with COVID-19 CT scans. Therefore, in the next section we will further study their related works.

2.4 Application of deep learning in segmentation of COVID-19 lesions from lung CT scans

In recent months, COVID-19 has become a hot topic of concern all over the world and CT imaging is considered to be a convincing method to detect COVID-19. However, due to the limited datasets and the time and labor involved in annotations, segmentation datasets related to COVID-19 CT scans are less readily available. But, many researchers have still proposed advanced methods to deal with COVID-19 diagnosis, which also includes segmentation techniques (Fan et al., 2020; Wang et al., 2020; Yan et al., 2020; Zhou et al., 2020; Elharrouss et al., 2020; Chen et al., 2020). On the premise of insufficient datasets with segmentation labels, the Inf-Net network proposed by Fan et al. (2020), combined a semi-supervised learning model and FCN8s network (Long et al., 2015) with implicit reverse attention and explicit edge attention mechanism to improve the recognition rate of infected areas. The model successfully segmented COVID-19 infected areas from CT scans and reported a sensitivity and accuracy of 72.5% and 96.0%, respectively. Elharrouss et al. (2020) proposed an encoder-decoder-based CNN method for COVID-19 lung infection segmentation based on a multi-task deep-learning based method, which overcame the shortage of labeled datasets, and segmented lung infected regions with a high sensitivity of 71.1%. Wang et al. (2020) proposed a noise-robust COVID-19 pneumonia lesions segmentation network which included a noise-robust dice loss function along with convolution function, residual network, and Atrous Spatial Pyramid Pooling (ASPP) module. The model was referred as Cople-Net presented automatic segmentation of COVID-19 pneumonia lesions from CT scans. The method proved that the proposed new loss function was better than the existing noise-robust loss functions such as Mean absolute error (MAE) loss (Ghosh et al., 2017) and Generalized Cross-Entropy (GCE) loss (Zhang and Sabuncu, 2018) and achieved a Dice coefficient and Relative Volume Error (RVE) of 80.72% and 15.96%, respectively. Yan et al. (2020) employed an encoder-decoder deep CNN structure composed of convolution function, Feature Variation (FV) module (mainly contains convolution, pooling, and sigmoid function), Progressive Atrous Spatial Pyramid Pool (PASPP) module (including convolution, dilation convolution, and addition operation) and softmax function. The convolution function obtained features, FV block enhanced the feature representation ability and the PASPP was used between encoder and decoder pipelines compensated for the various morphologies of the infected regions. The model achieved a good segmentation performance with a Dice coefficient of 0.726 and a sensitivity of 0.751 when tested on the COVID-19 lung CT scan datasets. Zhou et al. (2020) proposed an encoder-decoder structure based UNET model for the segmentation of the COVID-19 lung CT scan. The encoder structure was used to extract features and predict rough lesion areas which composed convolution function and Res-dil block (combines residual block (He et al., 2016) and dilation convolution module). The decoder pipeline was used to restore the resolution of the segmented regions through the upsampling and the attention mechanism between the encoder-decoder framework to capture rich contextual relationships for better feature learning. The proposed method can achieve an accurate and rapid segmentation on COVID-19 lung CT scans with a Dice coefficient, sensitivity, and specificity of 69.1%, 81.1%, and 97.2%, respectively. Further, Chen et al. (2020) proposed a residual attention UNET for automated multi-class segmentation of COVID-19 lung CT scans, which used residual blocks to replace traditional convolutions and upsampling functions to learn robust features. Again, a soft attention mechanism was applied to improve the feature learning capability of the model to segment infected regions of COVID-19. The proposed model demonstrates a good performance with a segmentation accuracy of 0.89 for lesions in COVID-19 lung CT scans. Therefore, the deep learning algorithms are helpful in segmenting the infected regions from COVID-19 lung CT scans which aid the clinicians to evaluate the severity of infection (Tang et al., 2020), large-scale screening of COVID-19 cases (Shi et al., 2020) and quantification of the lung infection (Ye et al., 2020). Table 1 summarizes the deep learning-based segmentation techniques available for COVID-19 lung infections.

3 METHODS

In this section, we first introduce the proposed ADID-UNET network with detailed discussion on the core network components including dense network, improved dilation convolution, and attention gate system. To present realistic comparisons, experimental results are presented at each subsection to illustrate the performance and superiority of the model after adding core components. Further in Section 4 we have presented a summary of the % improvements achieved when compared to the traditional UNET

architecture.

Literature	Data Type	Dataset	Technique	Segmentation Results
Fan et al. (2020)	CT Scan	100 CT images	Semi supervised CNN FCN8s network	73.9%(DC) 96.0%(S_p)
Wang et al. (2020)	CT Scan	558 CT images	Residual connection CNN	80.7%(DC) 16.0%(RVE)
Yan et al. (2020)	CT Scan	21,658 CT images	Deep CNN	72.6%(DC) 75.1%(S_{en})
Zhou et al. (2020)	CT Scan	100 CT images	Attention mechanism Res-Net,dilation convolution	69.1%(DC) 81.1%(S_{en})
Elharrouss et al. (2020)	CT Scan	100 CT images	Encoder-decoder-based CNN	78.6%(Dice) 71.1%(S_{en})
Chen et al. (2020)	CT Scan	110 CT images	Encoder-decoder-basedCNN	83.0%(DC) 89.0%(ACC)
Xu et al. (2020)	CT Scan	110 CT images	CNN	86.7%(ACC) 83.9%(F1)
Shuai et al. (2020)	CT Scan	670 CT images	CNN	73.1%(ACC) 67.0%(S_p)

Table 1. The summary of various deep learning algorithms for COVID-19 lung CT scans and the segmentation results.RVE, ACC, DC, S_{en}, S_p and F1 represent relative volume error, accuracy, Dice coefficient, sensitivity, specificity and F1 score, respectively.

3.1 ADID-UNET Architecture

ADID-UNET is based on UNET (Ronneberger et al., 2015) architecture with the following improvements: **(a)** The dense network proposed by Huang et al. (2017) is used in addition to the convolution modules of encoder and decoder structures, **(b)** an improved dilation convolution (IDC) is introduced between the frameworks, and **(c)** the attention gate (AG) system is used instead of the simple cropping and copying operations. The structure of ADID-UNET is shown in Figure 2. Here f_{en}, f_{upn}, f_{idc} describe the features at the $n-th$ layer of the encoder, decoder, and IDC modules, respectively.

When COVID-19 CT scans are presented to the encoder, the first four layers (each layer has convolutions, rectification, and max pooling functions) extract features (f_1 to f_4) that are passed to dense networks. Here dense networks are used instead of convolution and max-pooling layers to further enhance the features (f_5 to f_6) and in Section 3.2, we elaborate the need for the dense network and present experimental results to prove its significance. Next, an improved dilation convolution module referred as the IDC model, is used between the encoder-decoder structure to increase the receptive field and gather detailed edge information that assists in extracting the characteristic. The module accepts the feature f_6 from the dense networks and after improvement, present f_{idc} them as inputs to the decoder structure. To ensure consistency in the architecture and to avoid losing information, the decoder mirrors the encoder with two dense networks that replace the first two upsampling operations. Further for the better use of the context information between the encoder-decoder pipeline, the AG model is used instead of cropping and copying operations, which aggregates the corresponding layer-wise encoder features with the decoder and presents it to the subsequent upsampling layers. Likewise, the decoder framework presents upsampled features f_{up1} to f_{up6} and final feature map (f_{up6}) is presented to the sigmoid activation function to predict and segment the COVID-19 lung infected regions. The following section explains the components of ADID-UNET in detail.

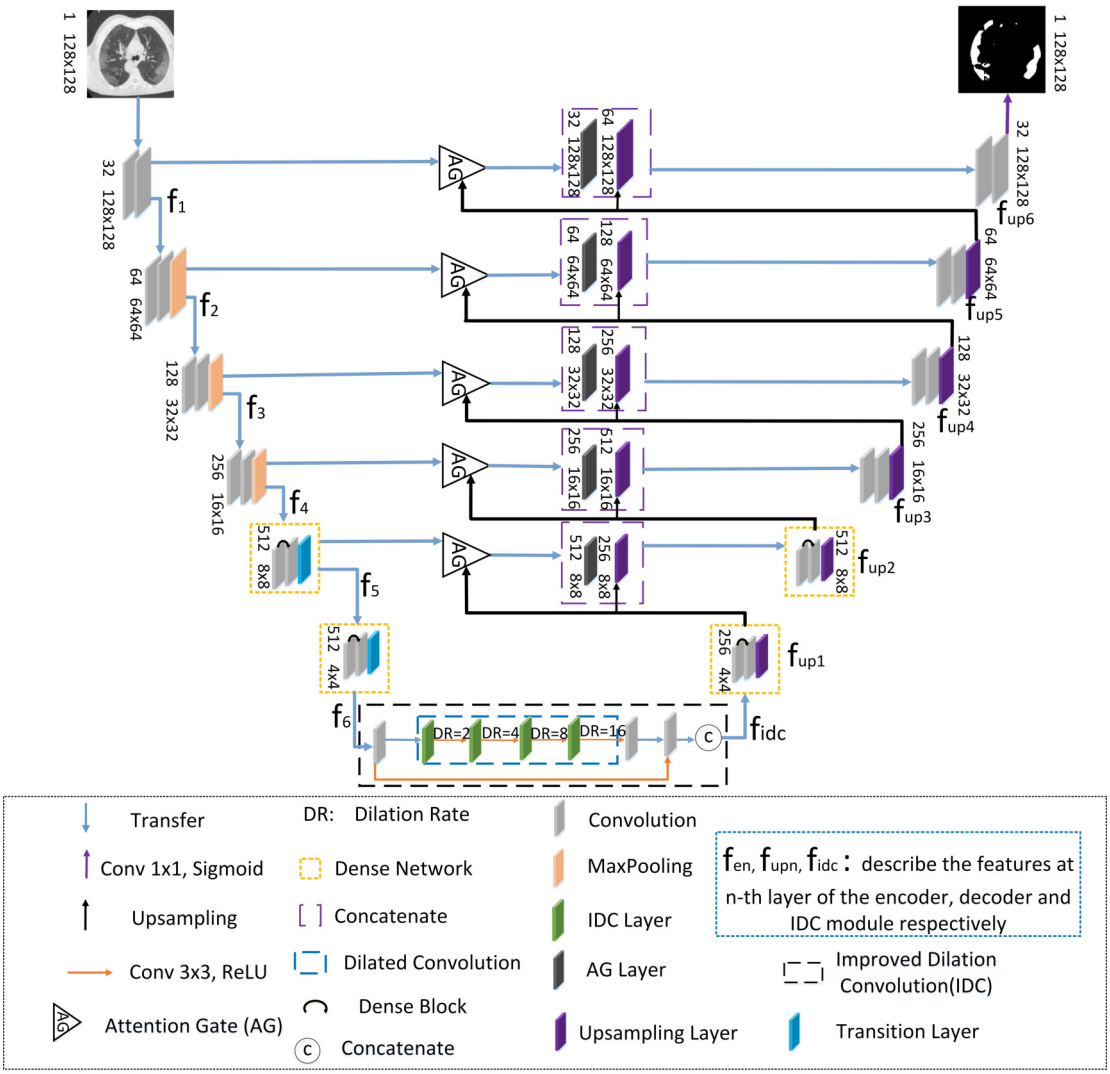

Figure 2. The structure of the ADID-UNET network.

3.2 Dense Network

It was presumed that with the increase of network layers, the learning ability of the network will gradually improve, but during the training, for deep networks, the gradient information that is helpful for the generalization may disappear or expand excessively. In literature, the problem is referred as vanishing or explosion of the gradient. As the network begins to converge, due to the disappearance of the gradient the network saturates, resulting in a sharp decline in network performance. Therefore, Zhuang et al. (2019a) introduced residual units proposed by He et al. (2016) into UNET structure to avoid performance degradation during training. The residual learning correction scheme to avoid performance degradation is described in (1):

$$y = G(x, \{F_i\}) + x \tag{1}$$

Here x and y are the input and output vectors of the residual block, F_i is the weight of the corresponding layer. The function $G(x, \{F_i\})$ is a residue when added to x, avoids vanishing gradient problems, and enables efficient learning.

From (1) the summation of $G(x, \{F_i\})$ and x in Res-Net (He et al., 2016) avoids the vanishing gradient problems but forwarding the gradient information alone to the proceeding layers may hinder the information flow in the network and the recent work by (Huang et al., 2016) illustarted that of Res-Nets discard features randomly during training. Moreover, Res-Nets include large number of parameters,

which increases the training time. To solve this problem, Huang et al. (2017) proposed a dense network (as shown in Figure 3), which directly connects all layers, and thus skillfully obtains all features of the previous layer without convolution. The dense network is mainly composed of convolution layers, pooling

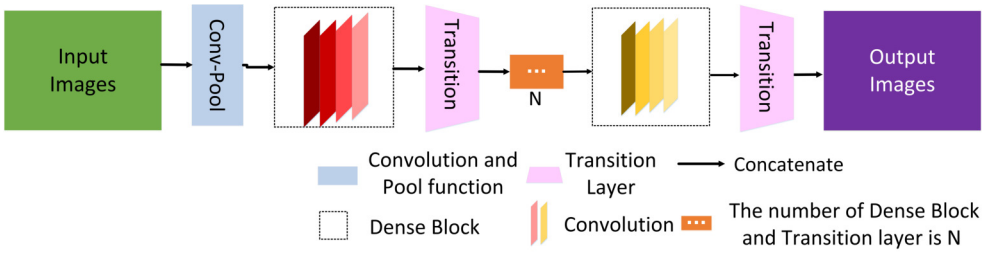

Figure 3. Schematic diagram of the dense network.

function, multiple dense blocks, and transition layers. Let us consider a network with L layers, and each layer implements a nonlinear transformation H_i. Let x_0 represent the input image, i represents layer i, x_{i-1} is the output of layer $i-1$. H_i can be a composite operation, such as batch normalization (BN), rectified linear function (RELU), pooling, or convolution functions. Generally, the output of traditional network in layer i is as follows:

$$x_i = H_i \times (x_{i-1}) \tag{2}$$

For the residual network, only the identity function from the upper layer is added:

$$x_i = H_i \times (x_{i-1}) + x_{i-1} \tag{3}$$

For a dense network, the feature mapping $x_0, x_1, \cdots, x_{i-1}$ of all layers before layer i is directly connected, which is represented by equation (4):

$$x_i = H_i \times ([x_0, x_1, \cdots, x_{i-1}]) \tag{4}$$

where $[x_0, x_1, \cdots, x_{i-1}]$ denotes the cascade of characteristic graphs and $\times$ represents the multiplication operation. Figure 4 shows the forward connection mechanism of the dense network where the output of layers is connected directly to all previous layers.

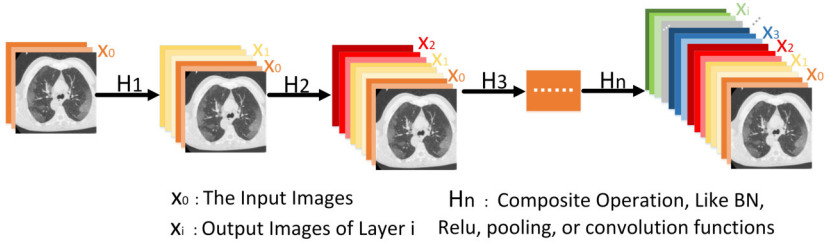

Figure 4. It is more intuitive to understand the forward connection mode of a dense network. The output x_i includes the input from $x_0, x_1, \cdots, x_{i-1}$.

Generally, a dense network is composed of several dense blocks and transition layers. Here we only use two dense blocks and transition layers to form simple dense networks. Using equation (5) to express the dense block:

$$\gamma = \alpha([x_0, x_1, \cdots, x_{i-1}], \beta_i) \tag{5}$$

where $[x_0, x_1, \cdots, x_{i-1}]$ denotes the cascade of characteristic graphs, β_i is the weight of the corresponding layer. In the ADID-UNET model proposed in this paper, the feature f_4 (refer to Figure 2) is fed to the transition layer, which is mainly composed of BN, RELU, and average pooling operation. Later the feature is batch standardized and rectified before convolving with a 1×1 kernel function. Again, the filtered outputs go through the same operation and are convoluted with 3×3 kernel, before concatenating with the input feature f_4. The detailed structure of the two dense blocks and transition layers used in the encoder

structure is shown in Figure 5 (A). Here w, h correspond to the width and height of the input, respectively, and b represents the number of channels. Besides, s represents the step size of the pooling operation, n represents the number of filtering operations performed by each layer. In our model,n takes values 32, 64, 128, 256, and 512. It should be noted that the output of the first dense layer is the aggregated result of 4 convolution operations $(4 \times n)$, which is employed to emphasize the features learning by reducing the loss of features. In the decoding structure, to restore the resolution of the predicted segmentation, a traditional upsampling layer of the UNET (Ronneberger et al., 2015) is used instead of the transition layer. The detailed structure is shown in Figure 5 (B).

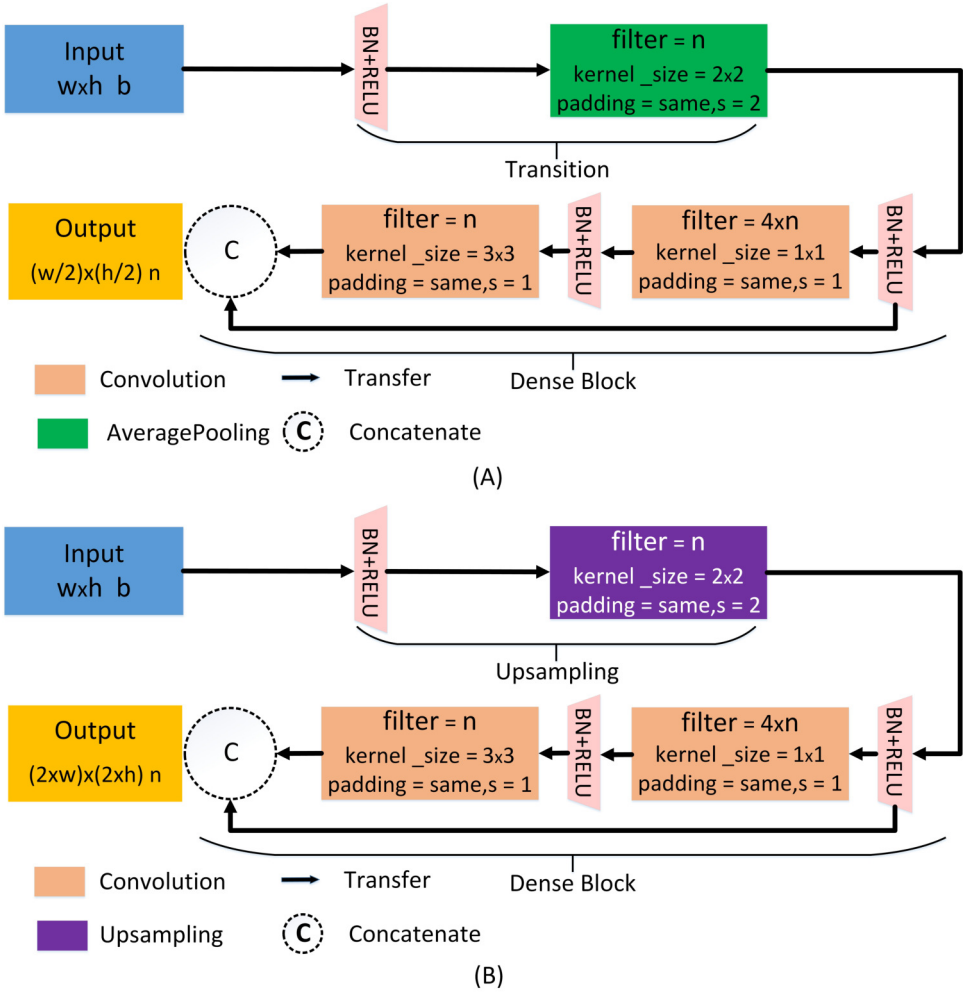

Figure 5. Dense network for encoder and decoder pipelines. (A) The dense network of the encoder pipeline. (B) The dense network of the decoder pipeline. Here, w, h, b, s, and n correspond to the width and height of the input, the number of channels, the step size of the pooling operation, and the number of filtering operations performed by each layer, respectively. For layers 6, 5, 4, 3, 2, 1, the values of n are 512, 512, 256, 128, 64, and 32, respectively.

For the proposed network, we use only two dense networks mainly **a)** to reduce the computation costs and **b)** experiments with different layers of dense networks suggest that the use of two dense networks was sufficient since the segmentation results were accurate and comparable to the ground truth. Figure 6 and Table 2 illustrate the qualitative and quantitative comparisons with different numbers of dense network in the encoder-decoder framework.

From the analysis of results in Figure 6 and Table 2, it is found that the effect of using two dense networks in the model is obvious and can present accurate segments of the infected areas that can be inferred directly from the qualitative and quantitative metrics.

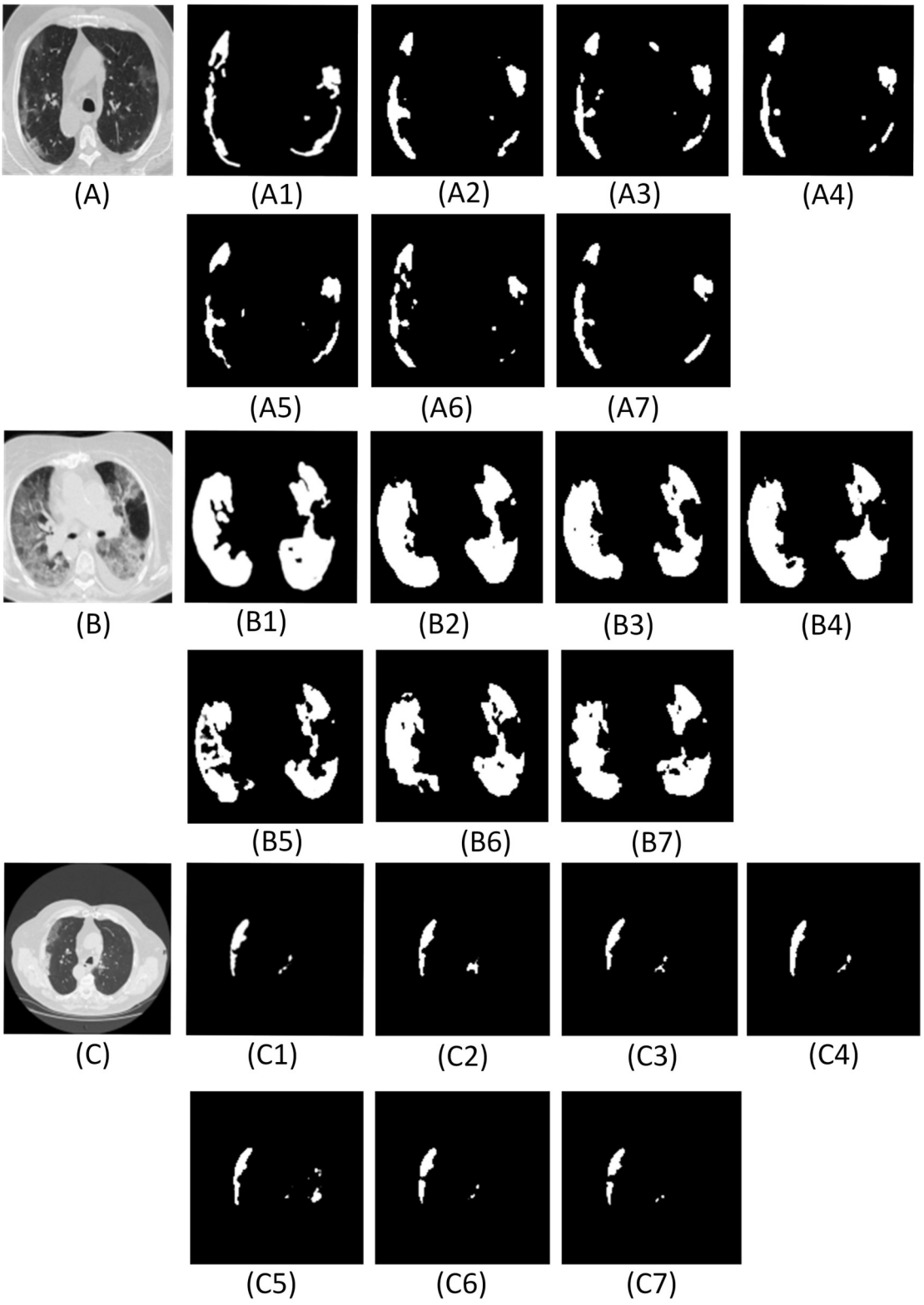

Figure 6. Results of adding different numbers of dense networks. Here (A)—(C) are the CT scans, (A1), (B1), (C1) are the ground truth, (A2)—(A7), (B2)—(B7), (C2)—(C7) are the segment results from UNET, Num_1, Num_2, Num_3, Num_4 and Res-Net, respectively. $Num_1 \sim Num_4$ denotes the number of dense networks used in the encoder and decoder pipelines. Res-Net refers to the network where convolution operations are replaced by the residual network (He et al., 2016). UNET (Ronneberger et al., 2015) denotes the traditional architecture without dense networks.

Moreover, with high accuracy and a good Dice coefficient, the choice of two dense networks is the best choice in the encoder decoder pipeline. Also, using two dense networks in place of traditional convolutions or residual networks enable global feature propagation, encourage feature reuse, and also solve the gradient disappearance problems associated with deep networks thereby significantly improving the segmentation outcomes.

Number of Dense Network	ACC	DC	S_{en}	S_p	P_c	AUC	F1	S_m	E_α	MAE
Num_1	0.9696	0.7971	0.8011	0.9958	0.8290	0.9513	0.8129	0.8411	0.9315	0.0088
Num_2	0.9700	0.8011	0.8096	0.9966	0.8596	0.9492	0.8184	0.8528	0.9394	0.0083
Num_3	0.9686	0.7569	0.7546	0.9957	0.8200	0.9334	0.7806	0.8349	0.9379	0.0104
Num_4	0.9699	0.7869	0.7579	0.9961	0.8485	0.9495	0.8241	0.8341	0.9348	0.0090
UNET	0.9696	0.7998	0.8052	0.9957	0.8247	0.9347	0.8154	0.8400	0.9390	0.0088
Res-Net	0.9698	0.8002	0.7978	0.9962	0.8344	0.9504	0.8180	0.8415	0.9352	0.0094

Table 2. Quantitative comparisons with respect to ground truth for different dense layers included the UNET (Ronneberger et al., 2015) and Res-Net (He et al., 2016) architecture. ACC, DC, S_{en}, S_p, P_c, AUC, F1, S_m, E_α and MAE represent accuracy, Dice coefficient, sensitivity, specificity, precision, the area under the curve, F1 score, structural metric, enhancement alignment meter and mean absolute error, respectively.

3.3 Improved dilation convolution

Since the encoder pipeline of the UNET structure is analogous to the traditional CNN architecture, the pooling operations involved at each layer propagate either the maximum or the average characteristics of the extracted features, hence connecting the encoder outputs directly to decoder, thus limiting the segmentation accuracy of the network. The RDA-UNET proposed by Zhuang et al. (2019a) utilized a dilation convolution (DC) module between the encoder-decoder pipeline to increase the receptive field and further learn the boundary information accurately. Also, the DC module is often used in many variant UNETs (Chen et al., 2019; Yu and Koltun, 2015) to improve the receptive field, hence, we use the DC module and introduce additional novelty in the DC module.

Equation (6) describes the DC operation between the input image $f(x,y)$ and the kernel $g(i,j)$.

$$p(x,y) = \alpha\left\{\sum_{i,j} f(x+i\times r, y+j\times r)\times g(i,j)+k\right\} \tag{6}$$

where α is the RELU function, k is a bias unit, (i,j) and (x,y) denote the coordinates of the kernel and those of the input images respectively, and r is the dilation rate that controls the size of receptive fields. The size of the receptive field obtained can be expressed as follows:

$$N = ((k\_f_{size}+1)\times(r-1)+k\_f_{size}) \tag{7}$$

where k_f_{size} is the convolution kernel size, r is the convolution rate of the dilation and N is the size of the receptive field. As shown in Figure 7.

Based on our experimental analysis we understand that DC module has a pronounced effect in extracting information for larger objects or lesions and considering that most of the early ground-glass opacity (GGO) or late lung consolidation lesions have smaller areas, we present an improved dilation convolution (IDC) module between the encoder-decoder framework to accurately segment smaller regions.

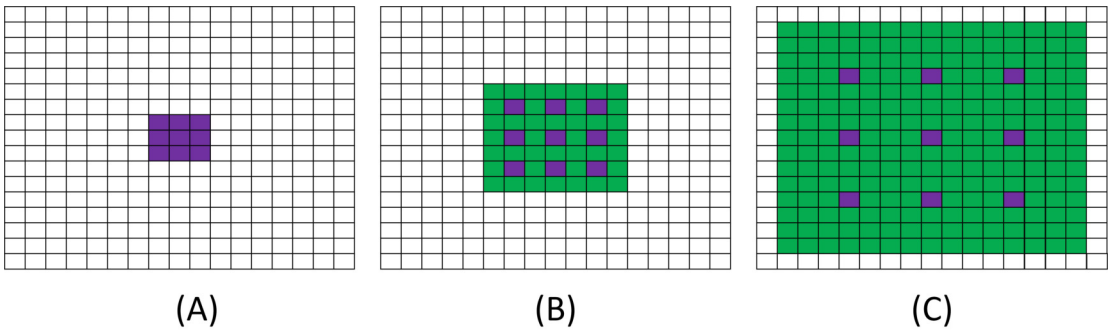

Figure 7. Schematic diagram of dilation convolution. (A) It shows the visual field effect of the classical 3×3 convolution kernel, covering 3×3 field of view each time (purple part of figure (A)); (B) corresponds to 3×3 with $r = 2$. Though the size of the convolution kernel is still 3×3, but the receptive field of convolution kernel is increased to 7×7 (purple and green parts in figure (B)); (C) corresponds to 3×3 with r =3 and a receptive field of 15×15.

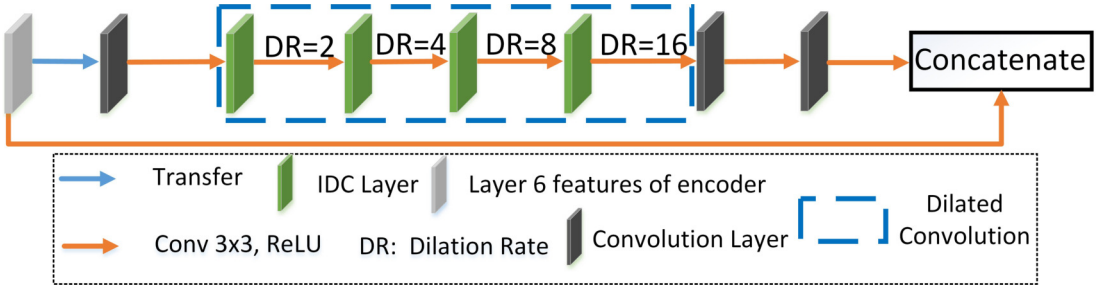

Figure 8. Improved dilation Convolution (IDC).

Method	Total Parameters	Trainable Parameters	Non-trainable Parameters	Train Time epoch/(s)	Test Time (s)
DD-UNET	56,223,034	56,190,272	32,762	145	8
DID-UNET	52,162,362	52,132,416	29,946	135	3

Table 3. Comparison results of the number of parameters of the UNET model with dense network incorporating either improved dilation convolution module (IDC) or dilation convolution module. Here DID-UNET refers to the inclusion of dense networks and IDC module and DD-UNET denotes dense networks and traditional dilation convolution added to the UNET structure.

Figure 8 illustrates the IDC module that consists of several convolution functions with different dilation rates and rectified linear functions (RELU). Our improvements are as follows: **a)** combining single strided convolution operations and dilated convolutions with dilation rate such as 2, 4, 8, and 16, respectively. The above combination helps in the extraction of features from both smaller and larger receptive fields thus assisting in the isolation of the small infected COVID-19 regions seen in lung CT scans and **b)** referring to the idea of the dense network (Huang et al., 2017), we concatenate the input of the IDC module to its output and use the information of input features to further enhance feature learning. The input of IDC module is the rough segmentation regions obtained by encoder structure. The combination of the original segmentation region features and the accurate features extracted by IDC module not only avoids the loss of useful information, but also provides accurate input for the decoding pipeline, which is conducive to improve the segmentation accuracy of the model. As the inputs advance (left to right in Figure 8), they get convolved with a 3×3 kernel of convolution layers and the dilation rate of IDC is 2, 4, 8, and 16, respectively. From the comparative experiments with the traditional DC model (the dilation rate is the same for both the models), we find that the computational cost and computation time required for the IDC module is less than that of the DC module, as shown in Table 3.

From Figure 9 and Table 4, it is found that the use of layers with convolution and smaller dilation rates at the end along with others ensures the cumulative extraction of features from both smaller and larger receptive fields thus assisting in the isolation of the small infected COVID-19 regions seen in lung CT scans. Also, the performance scores specifically the Dice coefficient is higher (about 3%) for DID-UNET compared to DD-UNET. In summary, the IDC model connected between the encoder-decoder structure, reduces loss of the original features but additionally expands the field of the segmented areas thereby improving the overall segmentation effect.

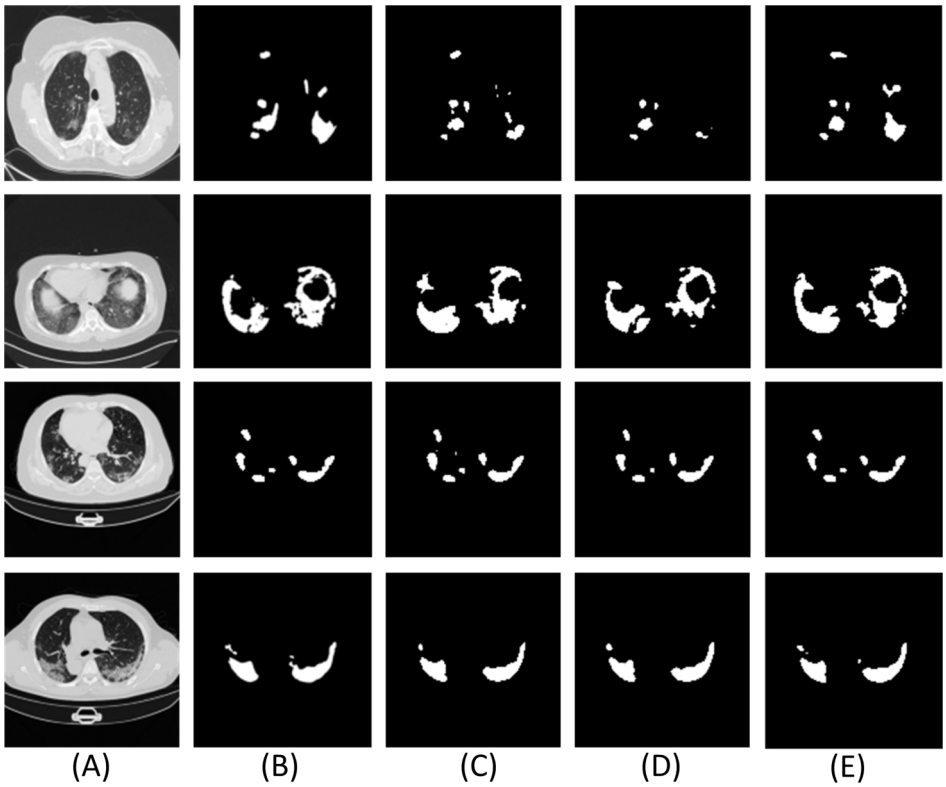

Figure 9. Experimental results of the improved dilation convolution and the traditional dilation convolution. Here (A) is the CT scan, (B) is the ground truth, (C)—(E) are the segment results from UNET, DD UNET and DID UNET, respectively.

Method	ACC	DC	S_{en}	S_p	P_c	AUC	F1	S_m	E_α	MAE
UNET	0.9696	0.7998	0.8052	0.9957	0.8247	0.9347	0.8154	0.8400	0.9390	0.0088
DD-UNET	0.9697	0.7757	0.7402	0.9971	0.8622	0.9214	0.7923	0.8401	0.9312	0.0094
DID-UNET	0.9700	0.8023	0.7987	0.9964	0.8425	0.9549	0.8241	0.8447	0.9374	0.0084

Table 4. The results of comparison indees of the improved dilation convolution experiment and the traditional dilation convolution experiment.

3.4 Attention Gate

Although the improved dilation convolution improves the feature learning ability of the network, due to the loss of spatial information in the feature mapping at the end of the encoder structure, the network has difficulties in reducing false prediction for **(a)** small COVID-19 infected regions and **(b)** areas with blurry edges with poor contrast between the lesion and background. To solve this problem, we introduce the

attention gate (AG) model shown in Figure 10 mechanism into our model instead of simple cropping and copying. AG model computes the attention coefficient $\sigma \in [0,1]$, based on equation (8):

$$\sigma = \varepsilon_2 \{p_k [p_i (\varepsilon_1 (p_n \times n + p_m \times m + b_{m,n})) + b_{\text{int}}] + b_k\} \tag{8}$$

$$\varepsilon_2 (x) = \frac{1}{1 + \exp(-x)} \tag{9}$$

where n and m represent the feature mapping of the AG module input from the decoder and encoder pipelines, respectively. And p_m, p_n, p_i, p_k are the convolution kernels of size 1×1. $b_{m,n}$, b_{int}, b_k represent the offset unit. ε_1 and ε_2 denote the RELU and sigmoid activation function respectively. Here ε_2 limits the range between 0 and 1.

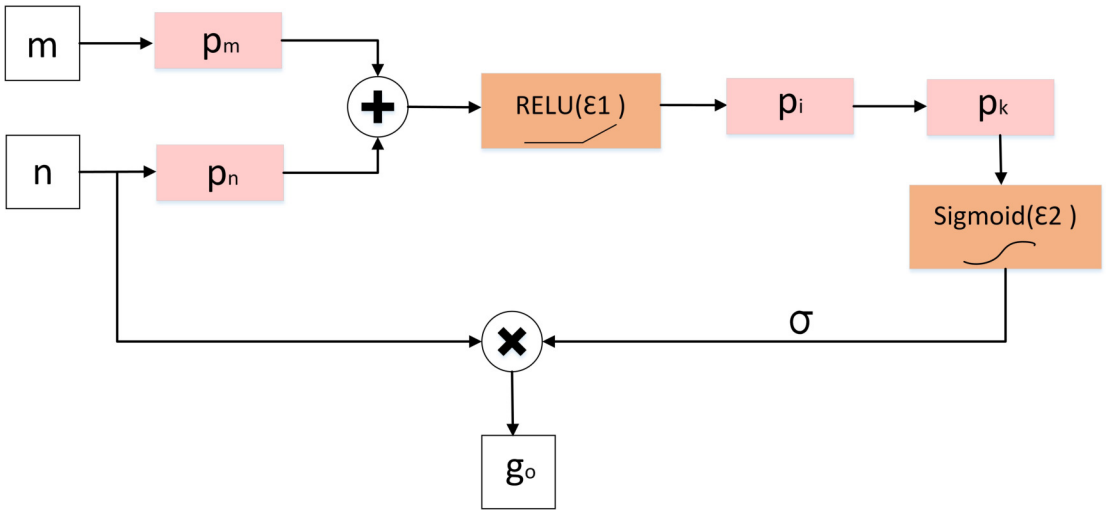

Figure 10. Diagram of Attention Gate(AG).

Finally, the attention coefficient σ is multiplied by the input feature map f_i to present the output g_o as shown in equation(10):

$$g_o = \sigma \times f_i \tag{10}$$

From Figure 11 and Table 5, results showed that the inclusion AG module improved the performance of the network (ADID-UNET), with segmentation accuracy of almost 97%. Therefore, by introducing the AG model, the network makes full use of the output feature information of encoder and decoder, which greatly reduces the probability of false prediction of small targets, and effectively improves the sensitivity and accuracy of the model.

4 EXPERIMENCE RESULTS

4.1 COVID-19 segmentation dataset collection and processing

Organizing a COVID-19 segmentation dataset is time-consuming and hence there are not many CT scan segmentation datasets. At present, there was only one standard dataset namely the COVID-19 segmentation dataset (Medseg.ai, 2020) , which was composed of 100 axial CT scans from different COVID-19 patients. All CT scans were segmented by radiologists associated with the Italian Association of medicine and interventional radiology. Since the database was updated regularly, on April 13, 2020, another segmented CT scans dataset with segment labels from Radiopaedia was added. The whole datasets that contained both positive and negative slices (373 out of the total of 829 slices have been evaluated by a radiologist as positive and segmented), were selected for training and testing the proposed model.

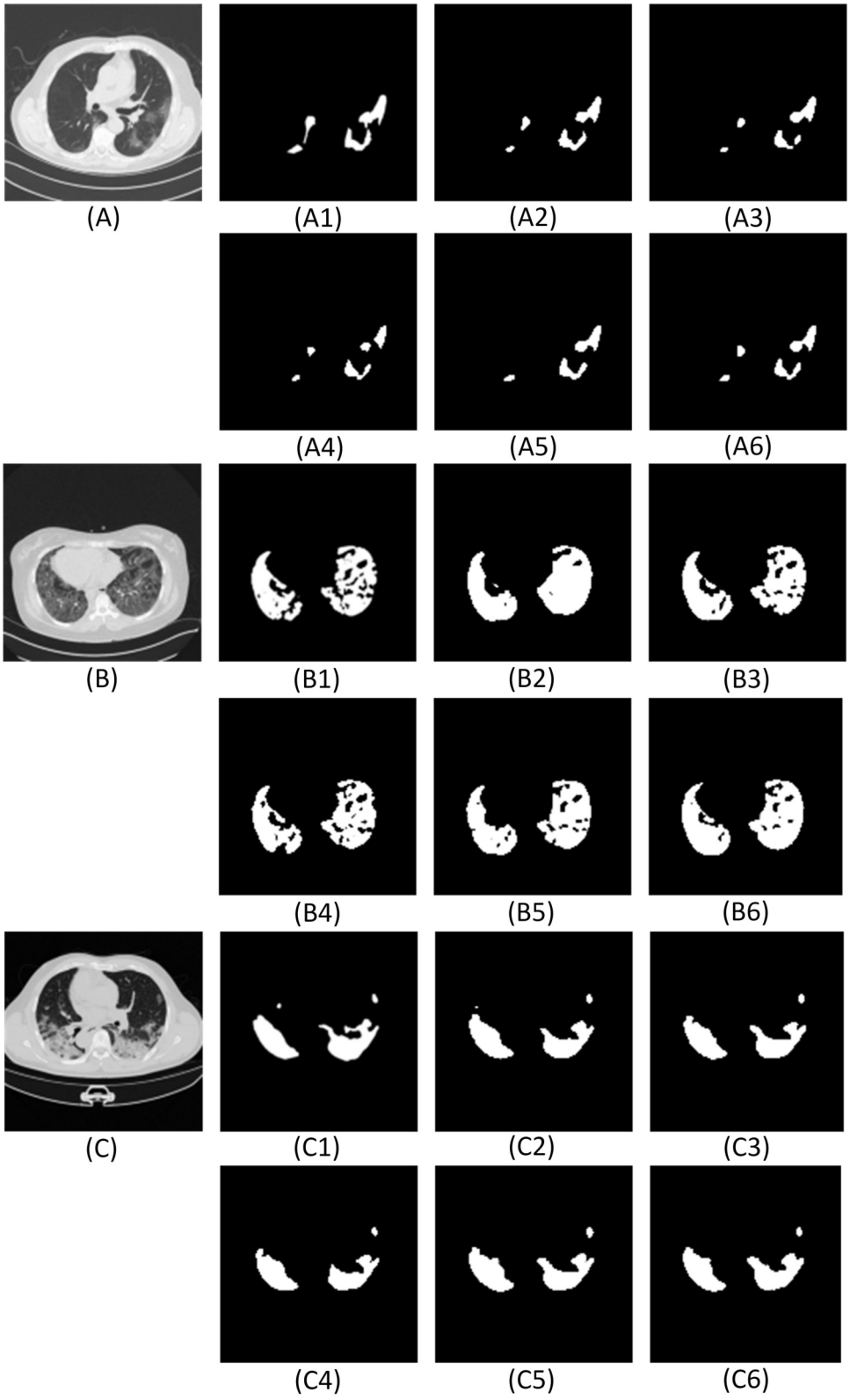

Figure 11. The qualitative results of comparative experiments with or without Attention Gate in the network. Here (A—C) illustrates the test images. (A1), (B1), (C1) are the ground truth. (A2)—(A6), (B2)— (B6), (C2)—(C6) are the segmentation results from UNET, AG-UNET, DA-UNET, IDA-UNET and ADID-UNET, respectively. AG-UNET– the addition of AG module to UNET, DA-UNET–adding two dense networks and AG module to the network without including the IDC module. IDA-UNET refers to adding IDC and AG modules to the UNET without adding dense networks, and ADID-UNET indicates that dense networks, IDC and AG module are added to the network.

Method	ACC	DC	S_{en}	S_p	P_c	AUC	F1	S_m	E_α	MAE
UNET	0.9696	0.7998	0.8052	0.9957	0.8247	0.9347	0.8154	0.8400	0.9390	0.0088
AG-UNET	0.9697	0.8020	0.8106	0.9962	0.8347	0.9571	0.8116	0.8511	0.9345	0.0087
DA-UNET	0.9698	0.7754	0.7400	0.9959	0.8470	0.9274	0.7930	0.8334	0.9104	0.0091
IDA-UNET	0.9698	0.7961	0.7834	0.9964	0.8469	0.9450	0.8126	0.8513	0.9437	0.0085
ADID-UNET	0.9701	0.8031	0.7973	0.9966	0.8476	0.9551	0.8200	0.8509	0.9449	0.0082

Table 5. The quantitative results of the comparison with or without the AG model experiment. AG-UNET– the addition of AG module to UNET, DA-UNET–adding two dense networks and AG module to the network without including the IDC module. IDA-UNET refers to adding IDC and AG modules to the UNET without adding dense networks, and ADID-UNET indicates that dense network, IDC and AG module are added to the network.

The dataset consists of 1838 images with annotated ground truth was randomly divided into 1318 training samples, 320 validation samples, and 200 test samples. Since the number of training images is less, we expand the training dataset where we first merge the COVID-19 lung CT scans with the ground scene and then perform six affine transformations as mentioned in Krizhevsky et al. (2012). Later the transformed image is separated from the new background truth value and added to the training dataset as additional training images. Therefore, the 1318 images of the training dataset are expanded, and 9226 images are obtained for training. Figure 12 illustrates the data expansion process.

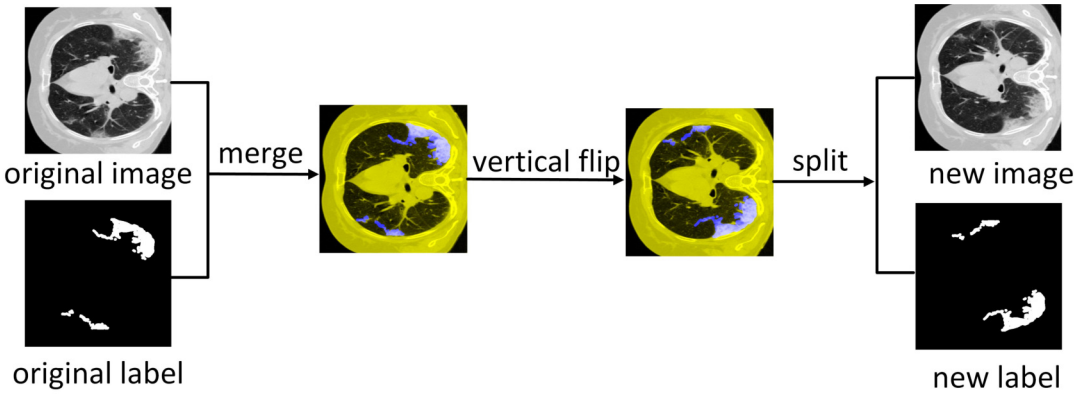

Figure 12. Data augmentation. Illustration of vertical flipping process showing the expansion of the training dataset.

4.2 Segmentation evaluation index

The commonly used evaluation indicators for segmentation such as accuracy (ACC), precision (P_c), Dice coefficient (DC), the area under the curve (AUC), sensitivity (S_{en}), specificity (S_p) and F1 score (F1) were used to evaluate the performance of the model. These performance indicators are calculated as follows: (1) For computing accuracy, precision, sensitivity, specificity, and F1 score we generate the confusion matrix where the definitions of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are shown in Table 6.

Category	Actual Lesion	Actual Non-Lesion
Predicted Lesion	True Position(TP)	False Position(FP)
Predicted Non-Lesion	False Negative(FN)	True Negative(TN)

Table 6. Definition of TP, FP, FN, TN.

1) **Accuracy** (ACC): A ratio of the number of correctly predicted pixels to the total number of pixels in the image.

$$Accuracy(ACC) = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

2) **Precision** (P_c): A ratio of the number of correctly predicted lesion pixels to the total number of predicted lesion pixels.

$$\text{Precision}(P_c) = \frac{TP}{TP+FP} \tag{12}$$

3) **Sensitivity** (S_{en}): A ratio of the number of correctly predicted lesion pixels to the total number of actual lesion pixels.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{13}$$

4) **F1 score** (F1): A measure of balanced accuracy obtained from a combination of precision and sensitivity results.

$$\text{F1score}(F1) = 2 \times \frac{P_c \times S_{en}}{P_c + S_{en}} \tag{14}$$

5) **Specificity** (S_p): A ratio of the number of correctly predicted non-lesion pixels to the total number of actual non-lesion pixels.

$$\text{Specificity}(S_p) = \frac{TN}{TN+FP} \tag{15}$$

6) **Dice coefficient** (DC): Represents the similarity between the model segment output (Y) and the ground truth (X). The higher the similarity between the lesion and the ground truth, the larger the Dice coefficient and the better the segmentation effect. Dice coefficient is calculated as follows:

$$\text{Dice Coefficient}(DC) = \frac{2 \times (X \cap Y)}{X+Y} \tag{16}$$

Also, we use a Dice coefficient (Dice, 1945) loss (dice_loss) as the training loss of the model, the calculation is as follows:

$$\text{Train Loss} = \text{Dice Coefficient Loss} = 1.0 - \frac{2 \times (X \cap Y)}{X+Y} \tag{17}$$

7) **The area under the curve** (AUC): AUC is the area under the receiver operating characteristic (ROC) curve. It represents the degree or the measure of separability and indicates the capability of the model in distinguishing the classes. Higher the AUC better is the segmentation output and hence the model.

In addition to the above widely used indicators, we also introduce the Structural metric (S_m) (Fan et al., 2017), Enhanced alignment metric (E_α) (Fan et al., 2018) and Mean Absolute Error (MAE) (Fan et al., 2020; Elharrouss et al., 2020) to measure the segmentation similarity with respect to the ground truth.

8) **Structural metric** (S_m): Measures the structural similarity between the prediction map and ground truth segmented mask, it is more in line with the human visual system than Dice coefficient.

$$S_m = (1-\beta) \times S_{os}(S_{op}, S_{gt}) + \beta \times S_{or}(S_{op}, S_{gt}) \tag{18}$$

where S_{os} stands for target perception similarity, S_{or} stands for regional perceptual similarity, $\beta = 0.5$ is a balance factor between S_{os} and S_{or}. And S_{op} stands for the final prediction result and S_{gt} represents the ground truth.

9) **Enhance alignment metric** (E_α): Evaluates the local and global similarity between two binary maps computed based on equation (19):

$$E_\alpha = \frac{1}{w \times h} \sum_i^w \sum_j^h \alpha \times (S_{op}(i,j), S_{gt}(i,j)) \tag{19}$$

where w and h are the width and height of ground truth S_{gt}, (i,j) denotes the coordinates of each pixel in S_{gt}. α represents the enhanced alignment matrix:

$$\alpha = \frac{2 \times \left(S_{gt} - \sqrt{S_{gt}}\right) \times \left(S_{op} - \sqrt{S_{op}}\right)}{\left(S_{gt} - \sqrt{S_{gt}}\right)^2 + \left(S_{op} - \sqrt{S_{op}}\right)^2} \tag{20}$$

10) **Mean Absolute Error** (MAE): Measures the pixel-wise difference between S_{op} and S_{gt}, defined as:

$$MAE = \frac{1}{w \times h} \sum_i^w \sum_j^h \left| S_{op}(i,j) - S_{gt}(i,j) \right| \tag{21}$$

4.3 Experimental details

The ADID-UNET proposed in this paper is implemented in Keras framework and is trained and tested by using the workstation with NVIDIA GPU P5000. During the training process, we set the learning rate as $l_r = 1 \times 10^{-3}$, and Adam optimizer was selected as the optimization technique. The 9226 training samples, 320 verification samples, and 200 test samples were resized to 128 $\times$ 128 and trained with a batch size of 32 for 300 epochs. Figure 13-14 shows the performance curves obtained for the proposed ADID-UNET during training, validation, and testing.

4.4 Segmentation results and discussion

(1) **Qualitative results**: To show the performance of the ADID-UNET model, we used 200 pairs of COVID-19 lung infection CT scans as test data, and the segmentation results are shown in Figure 15. From the analysis of Figure 15, it was found that the ADID-UNET model can accurately segment the COVID-19 lung infection areas from the CT scans, especially the smaller infected areas, and the segmentation result is very close to the ground truth. This illustrates the effectiveness of the proposed method for the segmentation of COVID-19 lung infection regions from CT scans.

Further, we also compare the proposed model with other state-of-art segmentation models. From the results (Figures 16 and 17 and Table 7), we can infer that the ADID-UNET model presents segmentation outputs closer to the ground truth. In contrast, the FCN8s network (Long et al., 2015) presents more under and over segmented regions. Further RAD-UNET (Zhuang et al., 2019a) presents comparable segmentation results but its effect is less pronounced for smaller segments. Analyzing the segmentation visual results from Figures 16 and 17, we can clearly find that the ADID-UNET model proposed in this paper can accurately segment the COVID-19 lung infection regions than other state-of-the-art model with results close to the ground truth, which proves the efficacy of the proposed ADID-UNET model.

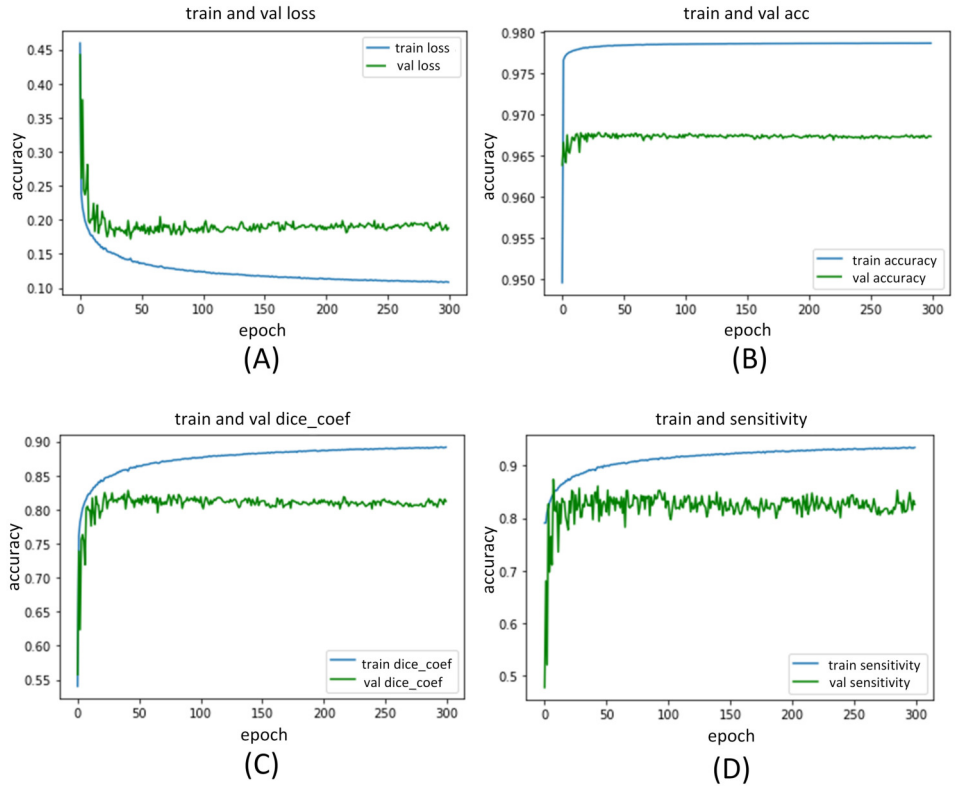

Figure 13. ADID-UNET training and validation performance index curve. (A-D) are the loss of training and validation, accuracy, Dice coefficient, and sensitivity performance curves, respectively.

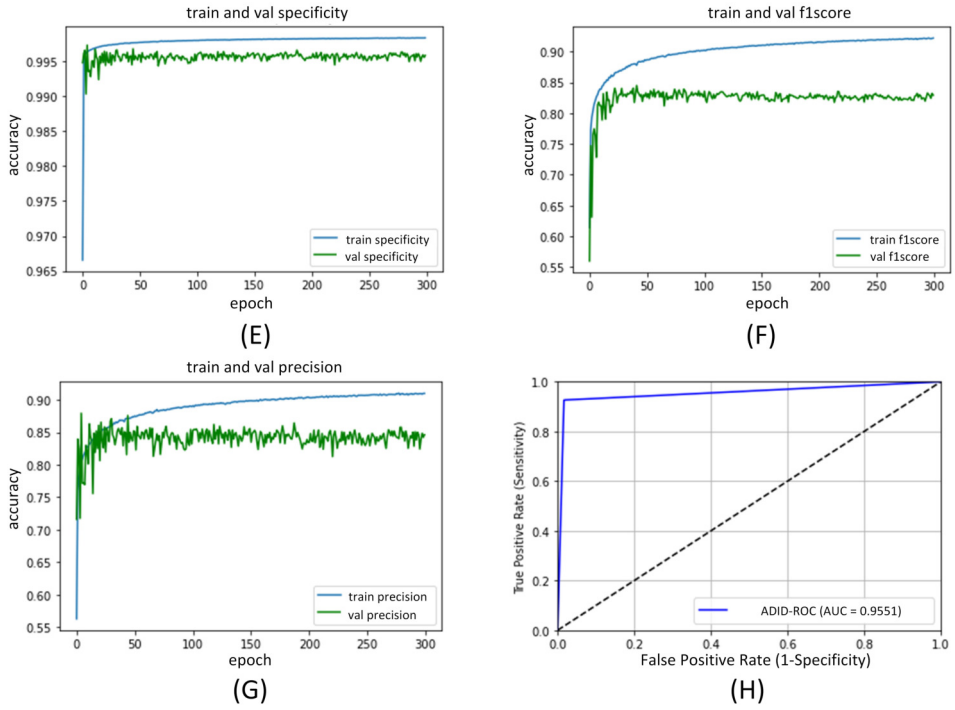

Figure 14. ADID-UNET training and validation performance index curve. (E-H) is the training and validation of the specificity, F1 score, precision, and AUC performance curves, respectively.

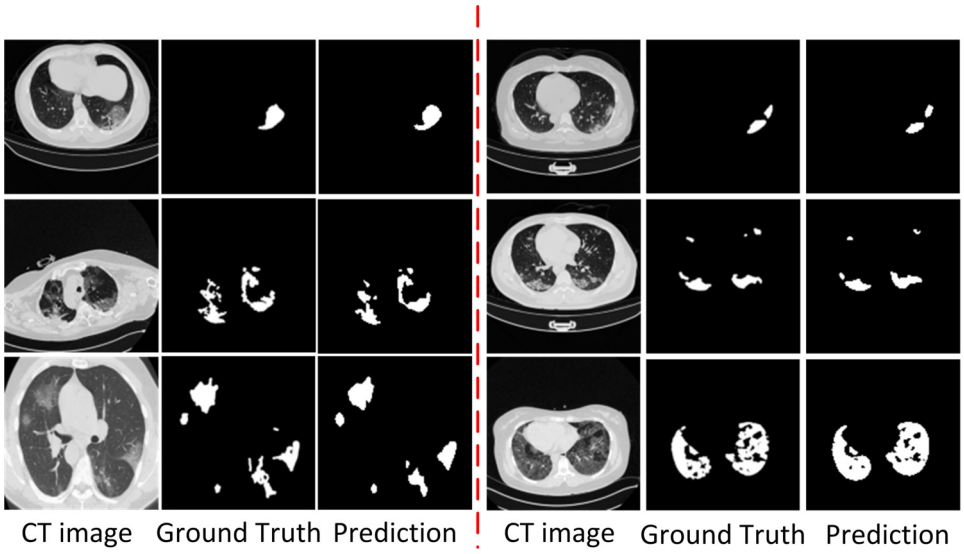

Figure 15. Visual comparison of the segmentation results of COVID-19 lung infection obtained from the proposed ADID-UNET.

(2) **Quantitative results**: Table 7, presents the performance scores for various indicators mentioned in Section 4. Here, for ADID-UNET the scores such as the Dice coefficient, precision, F1score, specificity and AUC are 80.31%, 84.76%, 82.00%, 99.66% and 95.51%, respectively. Further, most of the performance indexes are above 0.8 with the highest segmentation accuracy of 97.01%. The above results clearly indicates that the proposed model presents segmentation outputs closer to ground truth annotations.

(3) **Discussion**: The proposed model presents an improved verion of the UNET model obtained by the inclusion of modules such as the dense network, IDC and the attention gates to the existing UNET (Ronneberger et al., 2015) structure. The effectiveness of these additions were experimentally verified

in Section 3. Further, to summarize the effectiveness of the addition of each module to the UNET architecture, Table 8 tabulates the improvement at each stage of the addition. From Table 8, it is found that adding additional components to the UNET (Ronneberger et al., 2015) structure can obviously improve the overall segmentation accuracy of the network. For example, with the inclusion of the dense networks (D-UNET), the metrics such as Dice coefficien (DC) and AUC reached 79.98% and 93.47%, respectively.

Method	ACC	DC	S_{en}	S_p	P_c	AUC	F1	S_m	E_α	MAE
FCN8s	0.9666	0.6697	0.6692	0.9923	0.6860	0.9485	0.6724	0.7539	0.9134	0.0157
UNET	0.9696	0.7998	0.8052	0.9957	0.8247	0.9347	0.8154	0.8400	0.9390	0.0088
Segnet	0.9684	0.7408	0.7608	0.9937	0.7549	0.9492	0.7558	0.8080	0.9374	0.0125
Squeeze UNET	0.9689	0.7681	0.7827	0.9946	0.7776	0.9446	0.7785	0.8227	0.9326	0.0107
Residual UNET	0.9697	0.7924	0.7905	0.9961	0.8248	0.9444	0.8055	0.8397	0.9324	0.0094
RAD UNET	0.9699	0.7895	0.7625	0.9970	0.8601	0.9419	0.8062	0.8475	0.9328	0.0096
Fan et al. (2020)	- - -	0.7390	0.7250	0.9600	- - -	- - -	- - -	0.8000	0.8940	0.0640
Elharrouss et al. (2020)	- - -	0.7860	0.7110	0.9930	0.8560	- - -	0.7940	- - -	- - -	0.0760
Yan et al. (2020)	- - -	0.7260	0.7510	- - -	0.7260	- - -	- - -	- - -	- - -	- - -
Zhou et al. (2020)	- - -	0.6910	0.8110	0.9720	- - -	- - -	- - -	- - -	- - -	- - -
Chen et al. (2020)	0.8900	- - -	- - -	0.9930	0.9500	- - -	- - -	- - -	- - -	- - -
ADID-UNET	0.9701	0.8031	0.7973	0.9966	0.8476	0.9551	0.8200	0.8509	0.9449	0.0082

Table 7. Quantitative results of infected areas in the COVID-19 dataset. - - - means no relevant data in the original literature.

Method	ACC	DC	S_{en}	S_p	P_c	AUC	F1	S_m	E_α	MAE
UNET	0.9696	0.7998	0.8052	0.9957	0.8247	0.9347	0.8154	0.8400	0.9390	0.0088
D-UNET	0.9700	0.8011	0.8096	0.9966	0.8596	0.9492	0.8184	0.8528	0.9394	0.0083
DID-UNET	0.9700	0.8023	0.7987	0.9964	0.8425	0.9549	0.8241	0.8447	0.9374	0.0084
ADID-UNET	0.9701	0.8031	0.7973	0.9966	0.8476	0.9551	0.8200	0.8509	0.9449	0.0082
Improvement of D-UNET	↑0.04%	↑0.13%	↑0.44%	↑0.09%	↑3.49%	↑1.45%	↑0.30%	↑1.28%	↑0.04%	↓0.05%
Improvement of DID-UNET	↑0.04%	↑0.25%	↓0.65%	↑0.07%	↑1.78%	↑2.02%	↑0.87%	↑0.47%	↓0.16%	↓0.04%
Improvement of ADID-UNET	↑0.05%	↑0.33%	↓0.79%	↑0.09%	↑2.29%	↑2.04%	↑0.46%	↑1.09%	↑0.59%	↓0.06%

Table 8. The quantitative results showing percentages improvements of the model after adding additional components to UNET (Ronneberger et al., 2015) structure. D-UNET denotes dense networks with UNET structure, DID-UNET represents dense networks and improved dilation convolution to the structure of UNET, and ADID-UNET refers to proposed model with dense networks improved dilation convolution and attention gate modules to the UNET structure. ↑ indicates that the performance index is higher than that of UNET structure, ↓ indicates that the performance index is lower than that of UNET structure.

Further, the inclusion of the IDC improved the scores further (DID-UNET). Finally, the proposed model with dense network, IDC and the AG modules (namely ADID-UNET) presented the best performance scores and provided an improvement of 0.05%, 0.33%, 2.29%, 2.04% and 1.09% for metrics such as accuracy, DC, precision, AUC and structural metric respectively when compared to traditional UNET architecture.

Furthermore, from Figures 16 and 17, it is obvious that ADID-UNET performs better than other well-known segmentation models in terms of visualization. Specifically, ADID-UNET can segment relatively smaller infected regions which is of great significance for clinical accurate diagnosis of COVID-19 infection location. The use of **(a)** dense networks instead of traditional convolution and max-pooling function, **(b)** inclusion of improved dilation convolution module between the encoder-decoder pipeline and **(c)** the presence of attention gate network in the skip connections have presented accurate segmentation outputs for various types of COVID-19 infections (GGO and pulmonary consolidation). However, ADID-UNET still has room for improvement in terms of Dice coefficient and sensitivity and also computational costs which can be researched in future.

5 CONCLUSION

The paper proposes a new variant of UNET (Ronneberger et al., 2015) architecture to accurately segment the COVID-19 lung infections in CT scans. The model, ADID-UNET includes dense networks, improved dilation convolution, and attention gate, which has strong feature extraction and segment capabilities. The experimental results show that ADID-UNET is effective in segmenting small infection regions, with performance metrics such as accuracy, precision and F1 score of 97.01%, 84.76%, and 82.00%, respectively. The segmentation results of the ADID-UNET network can aid the clinicians in faster screening, quantification of the lesion areas and provide an overall improvement in the diagnosis of COVID-19 lung infection.

6 ACKNOWLEDGMENTS

We are very grateful to the Italian Society of medicine and interventional radiology, Radiopedia, and (Ma et al., 2020) for providing the COVID-19 CT scan segmentation database.

7 APPENDIX

We describe the abbreviations of this paper in detail, as shown in Table 9:

Abbreviation	Explanation
D-UNET	Inclusion of Dense networks to the UNET structure
AG-UNET	Inclusion of Attention gate module to the UNET structure
DA-UNET	Inclusion of both dense networks and attention gate module to the UNET structure
IDA-UNET	Inclusion of Improved dilation convolution and Attention Gate module to the UNET structure
DID-UNET	Inclusion of dense networks and improved dilation convolution to the UNET structure
ADID-UNET	Inclusion of dense networks, Improved dilation convolution and Attention Gate modules to the UNET structure

Table 9. An explanation of the acronyms that appears in this article.

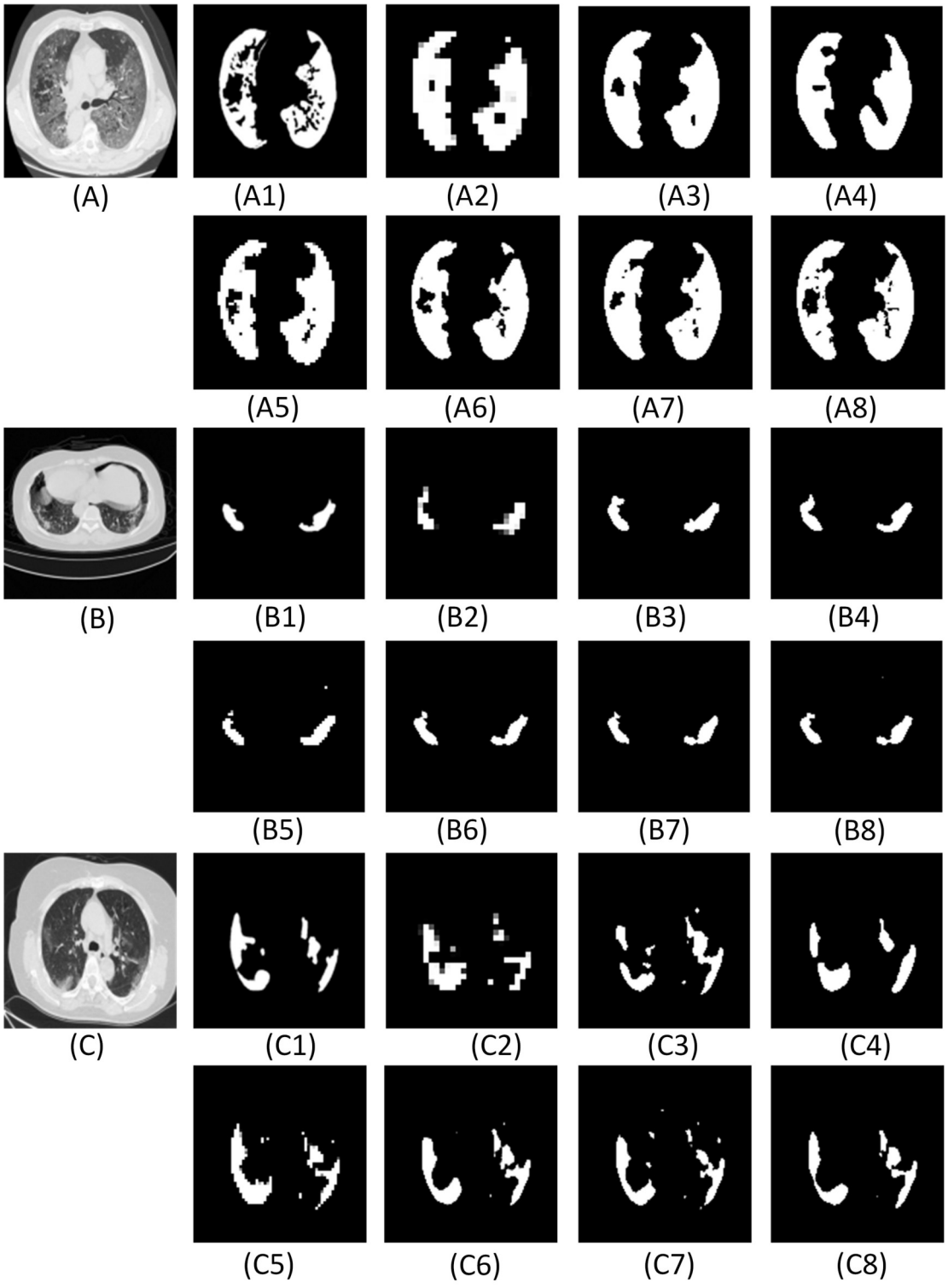

Figure 16. The visual comparison of the segmentation results of COVID-19 lung infection compared with other advanced models. Fig. 16 (A—C) illustrate the test images obtained from (Medseg.ai, 2020). Where (A1), (B1), (C1) are the ground truth, (A2)—(A8), (B2)— (B8), (C2)—(C8) are the segmentation results from FCN8s (Long et al., 2015), UNET (Ronneberger et al., 2015), SegNet(Badrinarayanan et al., 2017), Squeeze UNET (Iandola et al., 2016), Residual UNET (Alom et al., 2018), RAD UNET (Zhuang et al., 2019a), ADID-UNET, respectively.

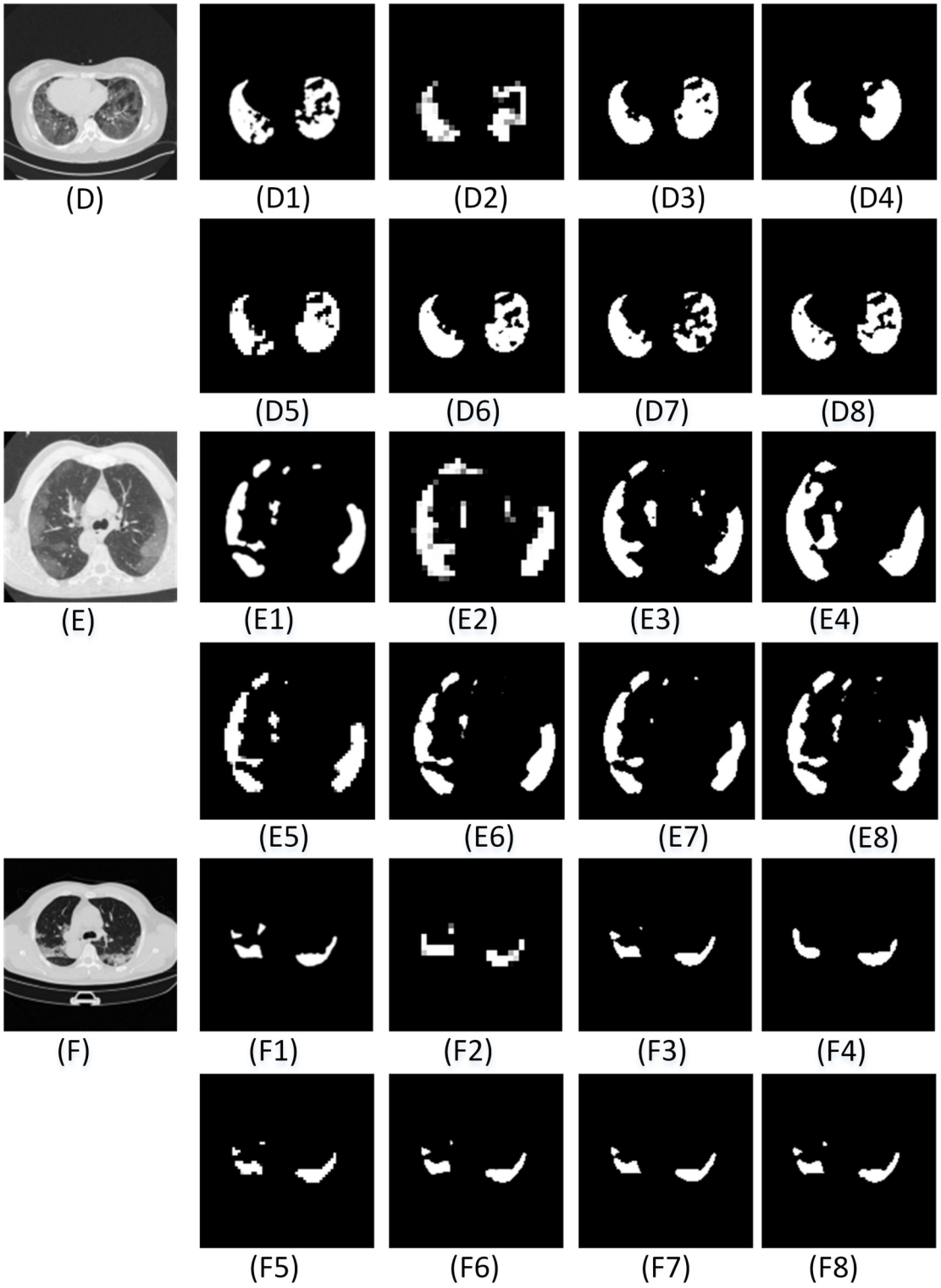

Figure 17. The visual comparison of the segmentation results of COVID-19 lung infection compared with other advanced models. Fig. 17 (D—F) illustrate the test images obtained from (Medseg.ai, 2020). Where (D1), (E1), (F1) are the ground truth, (D2)—(D8), (E2)— (E8), (F2)—(F8) are the segmentation results from FCN8s (Long et al., 2015), UNET (Ronneberger et al., 2015), Segnet(Badrinarayanan et al., 2017), Squeeze UNET (Iandola et al., 2016), Residual UNET (Alom et al., 2018), RAD UNET (Zhuang et al., 2019a), ADID-UNET, respectively.

REFERENCES

Agarwal, V., Tang, Y., Xiao, J., and Summers, R. M. (2020). Weakly-supervised lesion segmentation on CT scans using co-segmentation. In *Medical Imaging 2020: Computer-Aided Diagnosis*, volume 11314, pages 356 – 361.

Ai, T., Yang, Z., Hou, H., Zhan, C., and Xia, L. (2020). Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. *Radiology*, page 200642.

Almajalid, R., Shan, J., Du, Y., and Zhang, M. (2019). Development of a deep-learning-based method for breast ultrasound image segmentation. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*.

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on UNET (R2U-Net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*.

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., and Mougiakakou, S. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Transactions on Medical Imaging*, 35(5):1207–1216.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1.

Brostow, G. J., Fauqueur, J., and Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2):88–97.

Cardona, A., Saalfeld, S., and Preibisch, Stephan, e. a. (2010). An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. *PLoS Biology*, 8(10).

Chaganti, S., Balachandran, A., Chabin, G., Cohen, S., Flohr, T., Georgescu, B., Grenier, P., Grbic, S., Liu, S., and Mellot, F. a. (2020). Quantification of tomographic patterns associated with COVID-19 from chest CT. *arXiv*.

Chen, X., Yao, L., and Zhang, Y. (2020). Residual attention UNET for automated multi-class segmentation of COVID-19 chest CT images. *ArXiv*, abs/2004.05645.

Chen, Y., Guo, Q., Liang, X., Wang, J., and Qian, Y. (2019). Environmental sound classification with dilated convolutions. *Applied Acoustics*, 148:123–132.

Chung, Michael, Bernheim, Adam, Mei, X., Zhang, N., Huang, M., Zeng, X., Zeng, J., Xu, W., Yang, Y., and Fayad (2020). CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology*, 295(1):202–207.

Depeursinge, A., Chin, A. S., Leung, A. N., Terrone, D., Bristow, M., Rosen, G., and Rubin, D. L. (2015). Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution computed tomography. *Investigative Radiology*, 50(4):261–267.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3).

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

Fan, D.-P., Cheng, M.-M., Liu, Y., Li, T., and Borji, A. (2017). Structure-measure: A new way to evaluate foreground maps. In *Proceedings of the IEEE international conference on computer vision*, pages 4548–4557.

Fan, D.-P., Gong, C., Cao, Y., Ren, B., Cheng, M.-M., and Borji, A. (2018). Enhanced-alignment measure for binary foreground map evaluation. *arXiv preprint arXiv:1805.10421*.

Fan, D. P., Zhou, T., Ji, G. P., Zhou, Y., and Shao, L. (2020). Inf-Net: Automatic COVID-19 lung infection segmentation from CT scans. *IEEE Transactions on Medical Imaging*, PP(99):1–1.

Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., and Ji, W. (2020). Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *Radiology*, page 200432.

Ghosh, A., Kumar, H., and Sastry, P. S. (2017). Robust loss functions under label noise for deep neural networks. *arXiv preprint arXiv:1712.09482*.

Gordaliza, P. M., Arrate, M. B., Mónica, A., Manuel, D., Sally, S., and José, V. J. (2018). Unsupervised CT lung image segmentation of a mycobacterium tuberculosis infection model. *Scientific Reports*, 8(1):9802–.

Guofeng, T., Yong, L., Huairong, C., Qingchun, Z., and Huiying, J. (2018). Improved UNET network

for pulmonary nodules segmentation. *Optik - International Journal for Light and Electron Optics*, 174:460–469.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Huang, G., Liu, Z., Maaten, L. V. D., and Weinberger, K. Q. (2017). Densely connected convolutional networks. *CVPR*.

Huang, G., Sun, Y., Liu, Z., Sedra, D., and Weinberger, K. (2016). Deep networks with stochastic depth. *European conference on computer vision*.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). Squeezenet: AlexNet-level accuracy with 50x fewer parameters and ¡0.5mb model size. *arXiv preprint arXiv:1602.07360*.

Jue, J., Yu-Chi, H., Chia-Ju, L., Darragh, H., Hellmann, M. D., Deasy, J. O., Gig, M., and Harini, V. (2018). Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images. *IEEE Transactions on Medical Imaging*, PP:1–1.

Karen, S. and Andrew, Z. (2014). Very deep convolutional networks for large-scale image recognition. *Computer ence*.

Kermany DS, Goldbaum M, C. W. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122 – 1131.e9.

Kong, B., Wang, X., Bai, J., Lu, Y., and Yin, Y. (2019). Learning tree-structured representation for 3D coronary artery segmentation. *Computerized Medical Imaging and Graphics*, 80:101688.

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25.

Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., Cao, K., Liu, D., Wang, G., Xu, Q., Fang, X., Zhang, S., Xia, J., and Xia, J. (2020). Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy. *Radiology*, 296(2).

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651.

Ma, J., Ge, C., Wang, Y., An, X., Gao, J., Yu, Z., and et al., J. H. (2020). COVID-19 CT segmentation dataset. *radiopaedia.org*.

Medseg.ai (2020). Segmentation dataset,COVID-19 CT. *Dataset:https://medicalsegmentation.com/covid19/*.

Mukherjee, P., Lall, B., and Lattupally, S. (2018). Object co-segmentation using deep siamese network. *arXiv preprint arXiv:1803.02555*.

Negi, A., Raj, A. N. J., Nersisson, R., Zhuang, Z., and Murugappan, M. (2020). RDA-UNET-WGAN: An accurate breast ultrasound lesion segmentation using wasserstein generative adversarial networks. *Arabian Journal for Science and Engineering*.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., Mcdonagh, S., Hammerla, N. Y., and Kainz, B. a. (2018). Attention UNET: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Peng, T., Xu, T. C., Wang, Y., Zhou, H., and Wang, J. (2020). Hybrid automatic lung segmentation on chest ct scans. *IEEE Access*, PP(99):1–1.

Rajaraman, S., Candemir, S., Kim, I., Thoma, G., and Antani, S. (2018). Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Applied Sciences*, 8(10).

Ronneberger, Philipp, O. F., Brox, and Thomas. (2015). UNET: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*.

Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., and Rueckert, D. (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*.

Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., Xue, Z., and Shen., D. (2020). Lung infection quantification of COVID-19 in CT images with deep learning. *arXiv*.

Shi, F., Xia, L., Shan, F., Wu, D., Wei, Y., Yuan, H., Jiang, H., Gao, Y., Sui, H., and Shen, D. (2020). Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. *arXiv*.

Shuai, W., Bo, K., Jinlu, M., Xianjun, Z., Mingming, X., Jia, G., Mengjiao, C., Jingyi, Y., Yaodong, L., Xiangfei, M., and Bo, X. (2020). A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *MedRxiv*.

Tang, Z., Zhao, W., Xie, X., Zhong, Z., Shi, F., Liu, J., and Shen, D. (2020). Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest ct images. *arXiv*.

Wang, C., Horby, P. W., Hayden, F. G., and Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223):470 – 473.

Wang, G., Liu, X., Li, C., Xu, Z., Ruan, J., Zhu, H., Meng, T., Li, K., Huang, N., and Zhang, S. (2020). A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. *IEEE Transactions on Medical Imaging*, 39(8):2653–2663.

Wang, L. and Wong, A. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-Ray images. *arXiv preprint arXiv:2003.09871*.

Wang, S., Zhou, M., Liu, Z., Liu, Z., Gu, D., Zang, Y., Dong, D., Gevaert, O., and Tian, J. (2017). Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Medical image analysis*, 40:172–183.

Xia, X. and Kulis, B. (2017). W-Net: A deep model for fully unsupervised image segmentation. *arXiv preprint arXiv:1711.08506*.

Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yú, L., Chen, Y., wei Su, J., jing Lang, G., Li, Y.-T., Zhao, H., Xu, K., Ruan, L., and Wu, W. (2020). Deep learning system to screen coronavirus disease 2019 pneumonia. *ArXiv*, abs/2002.09334.

Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., and L., C. P. (2018). Medical breast ultrasound image segmentation by machine learning. *Ultrasonics*, 91:1–9.

Yan, Q., Wang, B., Gong, D., Luo, C., Zhao, W., Shen, J., Shi, Q., Jin, S., Zhang, L., and You, Z. (2020). COVID-19 chest CT image segmentation – a deep convolutional neural network solution. *arXiv preprint arXiv:2004.10987*.

Ye, H., Gao, F., and Yin, Y. (2019). Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *European radiology*.

Ye, X., Beddoe, G., and Slabaugh, G. (2009). Graph cut-based automatic segmentation of lung nodules using shape, intensity, and spatial features. *Workshop on Pulmonary Image Analysis, MICCAI*.

Ye, X., Lin, X., Dehmeshki, J., Slabaugh, G., and Beddoe, G. (2009). Shape-Based Computer-Aided detection of lung nodules in thoracic CT images. *IEEE Transactions on Biomedical Engineering*, 56(7):1810–1820.

Ye, Z., Zhang, Y., Wang, Y., Huang, Z., and Song, B. (2020). Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. *European Radiology*, 30(8):4381–4389.

Yu, F. and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Zhang, Z. and Sabuncu, M. R. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*.

Zhou, T., Canu, S., and Ruan, S. (2020). An automatic COVID-19 CT segmentation based on UNET with attention mechanism. *arXiv preprint arXiv:2004.06673*.

Zhuang, Z., Li, N., Raj, A. N. J., Mahesh, V. G. V., and Qiu, S. (2019a). An RDAU-NET model for lesion segmentation in breast ultrasound images. *PLoS ONE*, 14(8):e0221535.

Zhuang, Z., Noel, A., Raj, J., Jain, A., and Murugappan, M. (2019b). Nipple segmentation and localization using modified UNET on breast ultrasound images. *Journal of Medical Imaging and Health Informatics*, 9(9):1827–1837.