

# Residual Dilated U-net For The Segmentation Of COVID-19 Infection From CT Images

Alyaa Amer  
University of Lincoln  
Lincoln, UK  
Arab Academy for Science and Technology  
Cairo, Egypt  
aamer@lincoln.ac.uk  
alyaa.amer@aast.edu

Xujiong Ye  
University of Lincoln  
Lincoln, UK  
XYe@lincoln.ac.uk

Faraz Janan  
Imperial College  
London,uk  
fjanan@ic.ac.uk

## Abstract

*Medical imaging such as computed tomography (CT) plays a critical role in the global fight against COVID-19. Computer-aided platforms have emerged to help radiologists diagnose and track disease prognosis. In this paper, we introduce an automated deep-learning segmentation model, which builds upon the current U-net model, however, leverages the strengths of long and short skip connections. We complemented the long skip connections with a cascaded dilated convolution module that learns multi-scale context information, compensates the reduction in receptive fields, and reduces the disparity between encoded and decoded features. The short connections are considered in utilizing residual blocks as the basic building blocks for our model. They ease the training process, reduce the degradation problem, and propagate the low fine details. This enables the model to perform well in capturing smaller regions of interest. Furthermore, each residual block is followed by a squeeze and excitation unit, which stimulates informative features and suppresses less important ones, thus improving the overall feature representation. After extensive experimentation with a dataset of 1705 COVID-19 axial CT images, we demonstrate that performance gains can be achieved when deep learning modules are integrated with the basic U-net model. Experimental results show that our model outperformed the basic U-net and ResDUnet model by 8.1% and 1.9% in dice similarity, respectively. Our model provided a dice similarity measure of 85.3%, with a slight increase in trainable parameters, thus demonstrating a huge potential for use in the clinical domain.*

## 1. Introduction

Coronavirus Disease (COVID-19) infected more than 111 million people and it has caused the death of more than 2 million people worldwide, as of June 25<sup>th</sup> 2021 [1]. Typically, COVID-19 is diagnosed using the reverse transcription polymerase chain reaction (RT-PCR). RT-PCR test is available readily, nevertheless, it only shows the positive nature of the disease, not the extent of the damage to the lungs [4]. Therefore, to assess the severity of the disease and monitor the infection stage, many hospitals have employed medical radiology imaging as a rapid option for discovering the disease. Among them, X-rays and CT scans. Both screening modalities confirm the presence of lung infections [5]. However, a study presented in [26] confirms that CT scans provide higher sensitivity to COVID-19 infection than RT-PCR, where RT-PCR provides a sensitivity of 71%, and CT scans provide a sensitivity of 98%. Also, CT scans are way faster than RT-PCR, which is very useful in patients' follow-up assessment and evaluation of disease progression [22].

Infected lesion segmentation from CT images is an essential step for further diagnosis, evaluation, and quantification of COVID-19. This procedure provides crucial information for doctors to diagnose and quantify lung diseases [19]. Despite the increasing popularity of CT scans in diagnosing COVID-19 patients, they still pose burden on clinicians, where an experienced radiologist need more than 20 minutes to analyze each suspected patient [14]. Common symptoms of infection observed from CT slices include ground-glass opacity (GGO) in the initial stage, pulmonary consolidation in the advanced stage. The lesions in both infections are variable in size and shape. Also, GGO boundaries are difficult to identify, due to their low contrast and blurred boundaries. Figure 1 shows the challenging fuzzy nature and irregular shape of the COVID-19 infected

lesions.

These limitations make the manual delineation tedious, time-consuming, and influenced by individual bias. Therefore, in this work we provide an automated COVID-19 lesion segmentation model, for helping physicians in their diagnosis process, taking into consideration the challenging nature of COVID-19 infected lesions.

The rest of the paper is organized as follows. In Section 2, we review related segmentation models. In Section 3, we demonstrate the methodology of the proposed model then describe the utilized modules in details. In Section 4, we describe the experimental setup followed by description of the dataset used. Section 5 demonstrates the evaluation results followed by discussion and Section 6 concludes the paper.

## 2. Related work

Machine learning algorithms have the potential to play an important role in helping radiologists analyze chest CT scans to find the epidemic patterns and reduce the time required to examine each patient. However, there is still a shortage in the labelled COVID-19 dataset, due to the emergency of the COVID-19 pandemic within a short period. This has encouraged some researchers to integrate human in-the-loop strategy. For instance, developing deep learning models that involve interactivity with radiologists in the training process or to provide initial seeds given by the radiologist to guide their segmentation model [21, 17, 23].

Keshani et. al [18] proposed a segmentation model based on support vector machine classifier to detect lung lesions from CT images. Moreover, Shen et.al [24] proposed an automated segmentation model based on bidirectional chain code to improve lesion segmentation performance. However, those two models fail to significantly improve lesion segmentation due to the intensity homogeneity between the lung and infected lesion.

Owing to the profound significance of deep learning, many models have emerged to provide more powerful feature representations to improve infection diagnosis efficiency in CT images and facilitate subsequent quantification. For example, Wu et. al [25] developed a joint classification and segmentation model using their collected dataset, providing a sensitivity of 96% and specificity of 91.5%. After fine-tuning the model using vgg as the encoder part for low-level feature extraction, they achieved a segmentation accuracy of 76%. The authors claim that combining several pieces of information from various tasks helped their model improve and generalize. Following the same methodology, Jin et. al [17] provided an integrated model for segmentation and classification trained on a small dataset of 732 cases. The model was able to attain a segmentation accuracy of 75% on a dataset of 877 positive cases.

Finally, Elharrouss et al. [10] proposed an encoder-

decoder network that first extracts the lung region as a region of interest and then input the extracted lung region concatenated with the original CT image to another encoder-decoder model. The model was compared with U-net model, achieving a dice similarity measure of 78%.

Whilst having several advantages, we observe that the methods mentioned above have few limitations. First, they either rely on hand-crafted features, which results in a less generalized model, or shallow design networks that would struggle to extract multi-context discriminative features. Second, they use a joint model for both classification and segmentation tasks, where they mainly focused on using features learned from the segmentation model to enhance the classification performance. Finally, they designed a complicated model with a two-stage pipeline, one dedicated to construct a region of interest and follow it with another model for segmenting the infected region, which is computationally expensive compared to their achieved results.

Some of the challenges in segmentation of COVID-19 lesions, that we tackle in this paper are: 1) The intensity homogeneity and low contrast of CT images with a fuzzy boundary of the infected region. 2) The size and shape variability of the infected lesions. 3) The imbalanced data problem, where the infected region in some images would occupy less than 5% of the image, and in another image, it would occupy around 30% of the image. This imbalance would bias the model toward detecting more of the background pixels, resulting in more false-negatives.

The above challenges have motivated us to build a deep learning segmentation model efficient enough to overcome the fuzzy nature of lesions and consider their shape and size variability. In a nutshell, our contributions can be summarized as follow:

1. We analyze the U-net segmentation model and build upon it to provide an enhanced robust method that offers better performance with a slight increase in trainable parameter.
2. To ease the training process and help to propagate the low fine details, we replace the basic U-net blocks with residual blocks. The short skip connections in each residual block converge the model faster and help to capture small lesions that were mistakenly unidentified by the basic U-net model.
3. We address the problem of shape and size variability in lesions by integrating a cascaded dilated convolution module. This module expands the receptive field to capture multi-scale context information with rich spatial information. Also, it overcomes the drawback of successive sub-sampling operations, which degrades the spatial information.
4. We integrate dilated convolution along the long skip

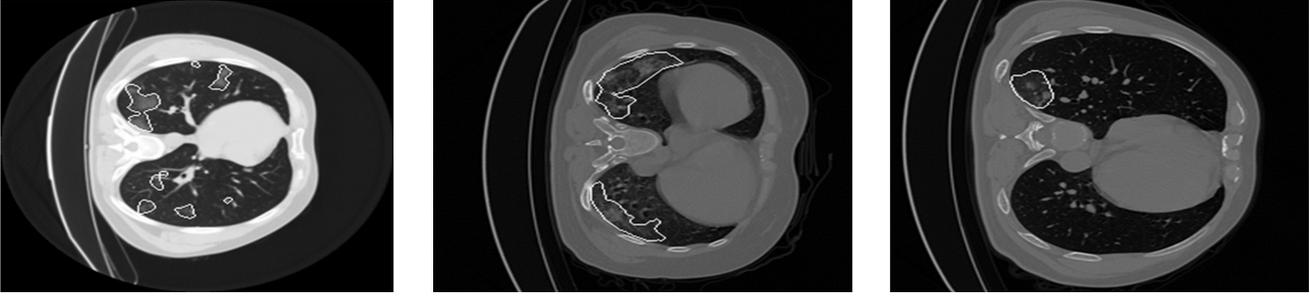


Figure 1. A sample of CT images with COVID-19 infection segmented by experts (illustrated in white). Showing the blurred boundaries and the variability in shape and size of the infection.

connections to reduce the semantic gap between features learned at encoding path and those learned at decoding path, which improves the model prediction.

### 3. Methodology

In this section we describe the proposed segmentation model, which carries the strengths of both residual and U-net architecture as depicted in figure 2. Our model replaces the building blocks utilized in the basic U-net model with repetitive residual blocks along the encoding and decoding path.

Each residual block is constituted of convolution, batch normalization, rectified linear unit (ReLU) activation function. We use 16 feature maps in the first layer and increment it with a factor of 2 to reach 512 feature maps at the bottom layer. A  $3 \times 3$  filter size is used in all convolution operations. Batch normalization is applied to reduce the internal co-variant shift and regularize the model, and ReLU is applied to introduce non-linearity into the model. To avoid over-fitting, one dropout operation with a rate of 0.5 is used after the first ReLU activation function in each residual block. The output is then concatenated with the input (identity short-cut), then passed through a squeeze and excitation (SE) unit to remove redundant features and excite more informative ones. After that, a max-pooling  $2 \times 2$  with a stride of 2 is applied to downsample the spatial dimension from one layer to another.

Prior to max-pooling operation, we pass the output of the SE unit to what we call a dilated convolution module **DC module**. In the DC module, the input is passed through a series of dilated convolutions with a filter size the same as the input filter size. For the first two layers, with the large semantic gap between encoder and decoder features, we perform dilated convolution with an increasing dilation rates of (1,2,3). This incremented dilation rate captures the large context information, and is followed by decremented dilation rates of (3,2,1) to capture local features dissipated by increasing the dilation rate. For the last two layers where the semantic gap decreases between encoder and decoder

features, we perform less processing to the encoder features, applying an increasing dilation rate of (1,2), followed by a decremental rate of (2,1). The output of each DC module is concatenated with the corresponding high-level feature at the decoding path.

At the decoding path, feature maps are up-sampled with bilinear interpolation, concatenated with the corresponding output of the DC module, then passed through residual blocks followed by Squeeze and excitation unit. The final layer is a  $1 \times 1$  convolutional layer followed by sigmoid function activation function to produce the final segmented lesion. The details of each utilized module is further investigated in the succeeding subsections.

#### 3.1. Residual blocks

It is confirmed that deeper networks provide better performance. However, training a deep network is difficult, especially for a limited number of training data, which can be defeated by employing a pre-trained network with fine tuning, or increasing the network depth [12]. However, increasing the network depth causes the problem of vanishing gradient when back propagating the signal across many layers [9].

Therefore, to avoid going deeper and solve the problem of gradient explosion and network degradation, we chose to replace the basic U-net blocks with residual blocks [12] at the encoding and decoding path. Residual blocks further contribute to feature propagation by sharing the same idea of concatenating features as the skip connections in the U-net. They implement 'identity short-cut mapping' of input and output, allowing the propagation of low fine details throughout the network, which improves the network performance without going deeper [12].

The utilized residual block is explained in figure 3(b) and compared with the basic U-net blocks illustrated in figure 3(a). Each residual unit is defined as:

$$y_i = h(x_i) + F(x_i, W_i) \quad (1)$$

$$x_{i+1} = f(y_i) \quad (2)$$

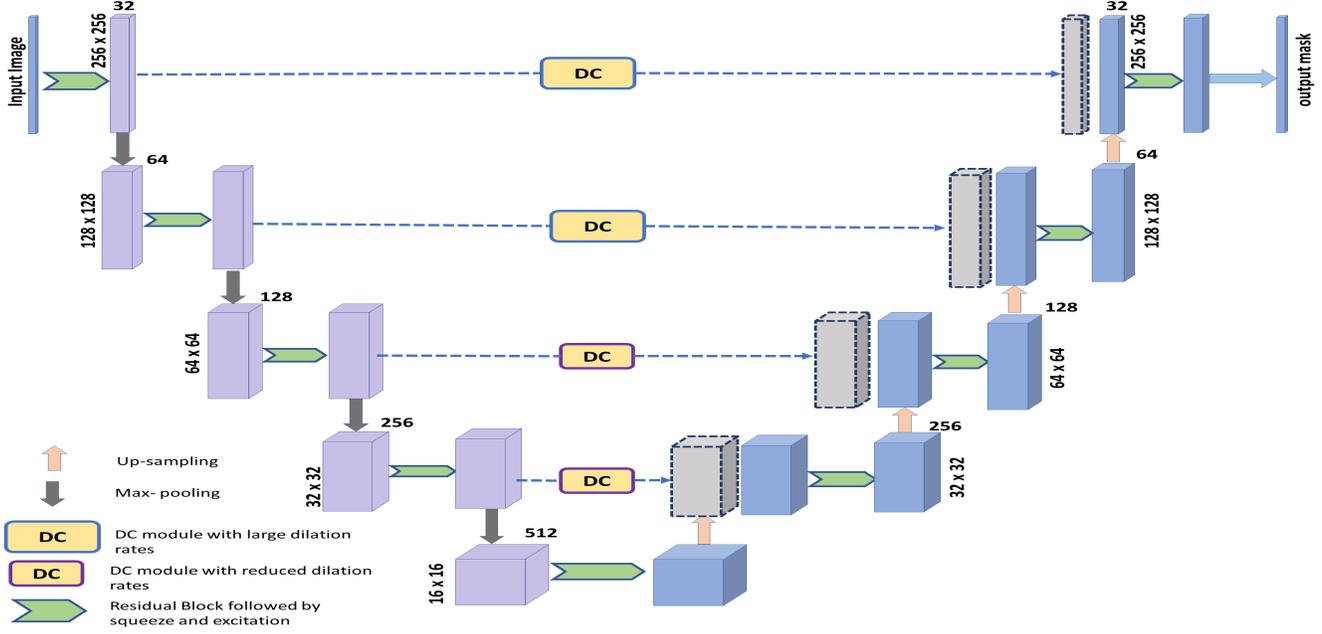


Figure 2. Overview of residual dilated U-net.

where  $x_i$  and  $x_{i+1}$  are the input and output of the  $i$ -th residual unit,  $F(\cdot)$  is the residual function,  $f(y_i)$  is the activation function and  $h(x_i)$  is an identity mapping function.

### 3.2. Squeeze and excitation unit

According to [13], SE enhances feature representation power by exciting more informative features and discarding less informative ones. Therefore, we followed each residual block with a squeeze and excitation unit (illustrated in figure 3(c)). The mechanism of SE unit allows the network at earlier layers to excite informative features, and strengthens the shared low-level representation. At deeper layers, it responds differently to inputs in a highly class-specific manner. Each unit performs two fundamental operations on the input feature map; squeeze for global information embedding and excitation for feature recalibration. At first, feature maps are aggregated across their spatial dimension  $H \times W$ , producing a channel descriptor. This operation is done through global average pooling to generate channel-wise statistics, which embed global distribution of channel-wise feature responses ( $C$ ). This information from the global receptive field of the network is used by all its layers. Afterwards, the aggregated channels go through an excitation process which is a simple self-gating mechanism using sigmoid activating function.

After global average pooling, there is a bottleneck with two fully connected (FC) layers at a reduction ratio ( $r$ ). This reduction ratio was set to 8 in our experiments which provides the lowest overall error with Resnet [13]. The output of the SE unit is obtained by rescaling the transformed fea-

ture maps with the activations, which act as weights adapted to the input features.

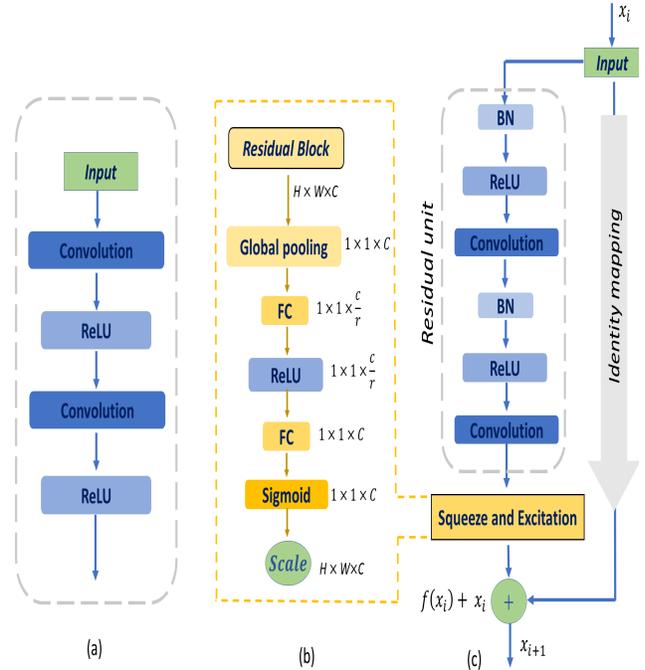


Figure 3. Building blocks. (a) U-net basic building unit. (b) Residual unit. (c) Squeeze and excitation unit.

### 3.3. Dilated convolution module

U-net skip connections help propagate the spatial information from the low-level layers to high-level ones. Despite maintaining the spatial details, the skip connections concatenate features learned at low level layers with minimal processing, with more processed features at higher layers. This process provides inconsistency throughout the model learning, due to the semantic gap between concatenated features [15]. Also, the sub-sampling process gradually loses the resolution of features as we go deeper in the network. Subsequently, this subsampling process misses the details of small objects that are hard to recover even with skip connections.

Therefore, we used a series of dilated convolution and concatenated them with the corresponding high level features along skip connections, to decrease the semantic gap between encoder and decoder features, capture multi-scale information and compensate the resolution degradation along the encoding path. Dilated convolution works by expanding the field of view of the filters, which help capture multi-scale information without losing the spatial resolution. In dilated convolution, the alignment of kernel weights is expanded by a dilation rate. Increasing this rate places the weights away at a given interval and increases the kernel size.

Dilated convolution is applied over a two-dimensional feature map  $x$ , where for each location  $i$  on the output  $y$ , a filter  $w$  is applied as defined in [7]:

$$y[i] = \sum_k x[i + d.k]w[k] \quad (3)$$

where the dilation rate  $d$ , is equivalent to the stride with which the input signal is sampled. This process resembles convolving the input  $x$  with the up-sampled filters produced by inserting  $(d-1)$  zeros between two consecutive filter values along each spatial dimension.

Thereby, using dilated convolution rates, we can adjust the filter's field of view, to capture multi-context information. Figure 4 shows that a dilated  $3 \times 3$  convolution with rate = 2 resembles a  $5 \times 5$  standard convolution, which means that the receptive field of the output is the same for both kernels. However, increasing this dilation rate is essential for resolution and context, and is damaging to very small objects [27]. Also, aggressively increasing the dilation rate fails to capture small objects' local features [11].

Therefore, we followed the approach presented in [11], where we gradually increase the dilation rate and follow that with a gradual decrease in dilation rate. We call this operation 'DC module', illustrated in figure 4. The gradual increase in dilation rates helps capture large-scale contexts. Following it with a gradual decrease in the dilation rate helps propagate local features scattered by increasing the dilation rate. At the last two layers, the dilation rate is

reduced, since the inconsistency between the concatenated features decreases gradually. This is due to the fact that, not only the features from the encoding path are going through more processing, but they are concatenated with processed decoder features.

To provide multi-resolution analysis, we incorporate consecutive  $3 \times 3$  convolution along the DC module, since this arrangement of convolution resembles a  $7 \times 7$  convolution. According to [28], this convolution arrangement helps the network mitigate the features learned from the image at a different scale.

## 4. Experimental setup

We implemented the model in Python environment with Keras and Tensorflow. For optimization, Adam optimizer was used at 350 epochs. The learning rate is initially set to  $10^{-3}$  and is reduced by a factor of 10 every 20 epochs. For each training run, the model with the best validation loss was stored and evaluated. Dice loss is used as the network objective function. All computations were carried out using Nvidia GeForce with RTX 2070 GPU.

### 4.1. Dataset

In this study we used a public available COVID-19 CT segmentation dataset introduced in [2], which includes 1705 COVID-19 axial CT images. The images' dimension varies from  $630 \times 630$  to  $512 \times 512$ . All images were intensity normalized between 0 and 1. Inspired by the basic U-net model, we resized all images to  $256 \times 256$  to reduce the computational complexity. We first split the dataset into training, validation and testing sets. Then augment each set to provide robustness and reduce over-fitting of our model. Data augmentation is applied by using the image augmentation library 'Augmenter' [6], where we applied different variations such as random horizontal and vertical flip, and random zoom. In total, the model was trained on 3410 images (after augmentation), with 20% used for validation and 20% for testing.

## 5. Evaluation results and discussion

To assess our proposed model's performance, we compare the lesion segmentation results with the experts' ground truth delineation. Also, our results are compared with the basic U-net and the results obtained from ResDUnet model, presented in [3]. ResDUnet model is also based on U-net with residual blocks integrated with squeeze and excitation unit, and utilizes dilated convolution with decremental dilation rates along the encoding path. For comparing the results, we used three quantification measures, the Dice similarity coefficient (DSC)[8], F-score (F)[20], and Jaccard index (JC)[16]. The DSC measure is

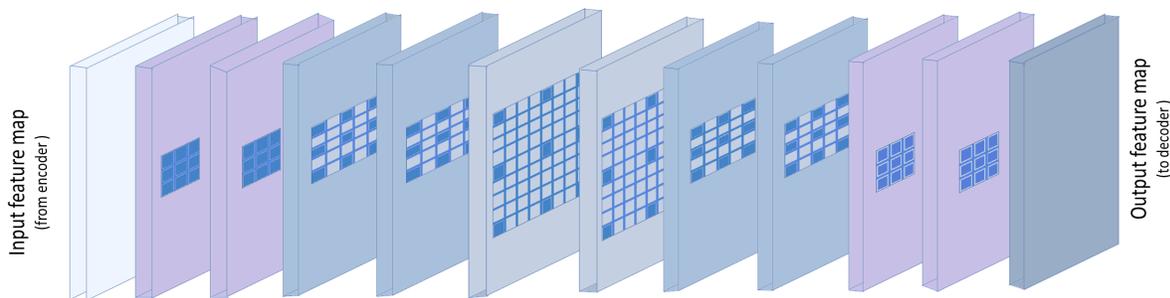


Figure 4. Overview of the DC module used in skip connections. Each layer is convolved with a kernel size of  $3 \times 3$ . Filter kernel is depicted in each layer as a square with grid pattern. The darker cells represent valid weights, while brighter ones represent invalid region. As shown, kernels with a dilation factor of 2 and  $3 \times 3$  convolution, resembles a  $5 \times 5$  convolution, thus showing how dilated convolution increases the receptive field.

the spatial overlap index and reproducibility validation index, defined as:

$$DSC = \frac{2TP}{2TP + FN + FP} \quad (4)$$

F-score is the test's accuracy, calculated from the precision and recall, defined as:

$$F = \frac{TP}{TP + \frac{1}{2}(FN + FP)} \quad (5)$$

JC is the intersection over union, and is defined as:

$$JC = \frac{TP}{TP + FN + FP} \quad (6)$$

where, TP is true positive, TN is true negative, FP is false positive and FN is false negative values. The higher the value of the above indices the better the performance.

The proposed model's segmentation performance in comparison with U-net and ResDUNet is illustrated in Table 1. It can be seen that our model outperforms U-net and ResDUNet by 8.1% and 1.9% in the dice similarity measure, respectively. We argue that the low performance seen in U-net is due to the process of sub-sampling layers throughout the network, where feature resolution is gradually lost. Both ResDUNet and our model outperform U-net due to the integration of residual blocks followed by squeeze and excitation unit. The short skip connections in residual blocks help propagate the low fine details throughout the network. The squeeze and excitation unit's addition has further contributed to exciting informative features and suppressing less informative ones. However, in ResDUNet, cascaded dilated convolution usage in a decremental way (along the encoding path) has focused on capturing small context information, thus discarding larger ones, which explains the lower accuracy compared to our model.

Figure 5, shows a qualitative comparison between the predicted contours from our model versus U-net and ResDUNet, where the expert manual segmentation is illustrated

Method	DSC	JC	F-score
U-net	77.2	75.4	79.4
ResDUNet	83.4	81.8	87.6
Our model	<b>85.3</b>	<b>83.9</b>	<b>90.6</b>

Table 1. Statistical evaluation for our model in comparison with other segmentation models.

Method	Small lesions	Large lesions
ResDUNet	84.9	83.9
our model	<b>84.8</b>	<b>86.9</b>

Table 2. Comparing the performance of both models in capturing large and small lesions.

in white. There, we can see two different lesions with variability in shape and size. For large infected lesions, our model outperformed ResDUNet, since we increased the dilation rate along skip connections to consider large scale context, and further operated reversely to propagate small features scattered by increasing the dilation rate. Both, our model and ResDUNet provide similar results for small lesions, since both of them are highly efficient in capturing small context information, through the integration of residual blocks which propagate the low fine details through the utilization of short skip connections. To illustrate more, Table 2 shows the dice similarity measure calculated for large and small lesions in both our model and ResDUNet. large lesions are those of an area more than 5mm and small lesions are those of an area less than or equal to 5mm. From the table we can see that both models provide comparable results in capturing the small lesions. However, our model has outperformed in capturing large lesions. The two types of lesions demonstrate our model's robustness to segment lesions with variable size and shape.

Table 3 shows the performance gains achieved when each module is integrated into our model. It is observed that adding residual units has increased the model performance.

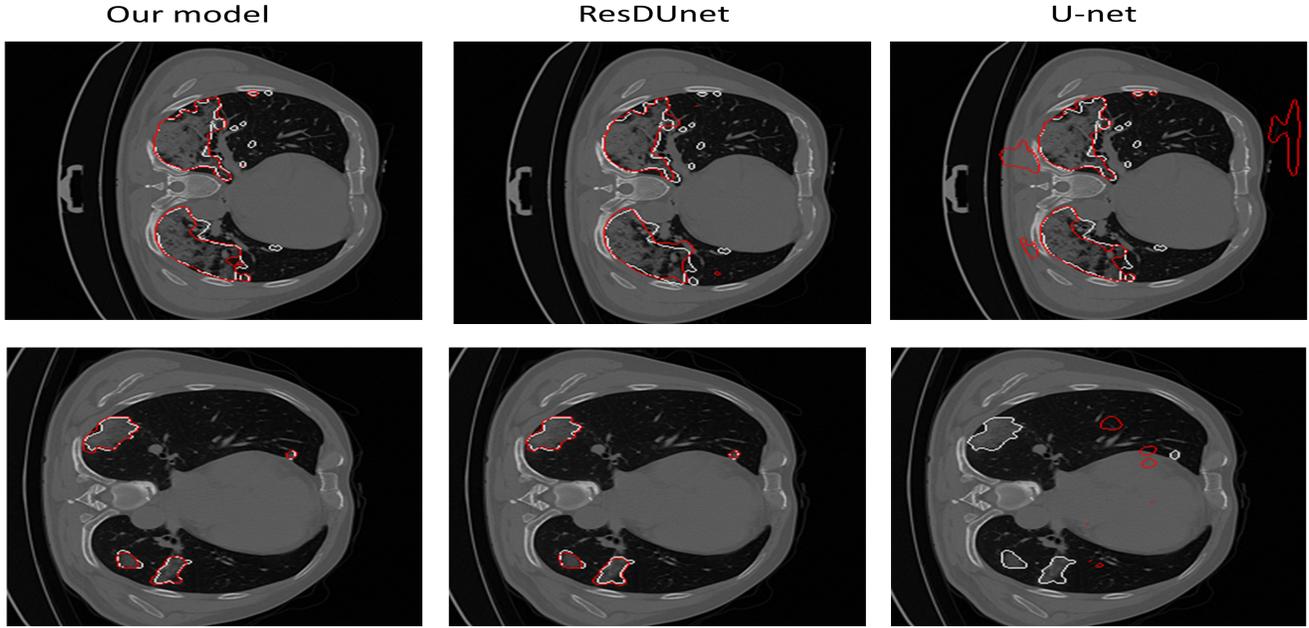


Figure 5. A visual comparison for the segmentation of COVID-19 infection from three models. The first row contains large infected region. The second row contains small infected region. Predicted region is demonstrated in red and experts annotation in white. Our model outperformed U-net and ResDUNet, in large and small infected regions.

Method	DSC	JC	F-score
Basic U-net	77.2	75.4	79.4
with RD	79.3	77.1	83.5
with RD and SE	80.5	78.6	86.2
with DC module	<b>85.3</b>	<b>83.9</b>	<b>90.6</b>

Table 3. Statistical results achieved after adding each module; RD: residual block; SE: squeeze and excitation; DC: dilated convolution)

This enhanced performance is due to the presence of identity short-cut mapping in each residual block which allows the flow of information from initial layers to last layers and reduces the problem of vanishing and exploding gradients. When each residual block was integrated with SE unit, the model performance was further enhanced by 1.2%. The SE unit has contributed in increasing the feature representation power through excitation operation and suppressing less important information through squeeze operation. The model performance was further boosted by 4.8% when cascaded dilation was added along the encoding path. This is due to the large and variable receptive field integrated in the cascaded scheme to capture the multiscale features.

Worth mentioning, for the benefit of improving the segmentation accuracy, our model has slightly increased the number of trainable parameters with a reasonable aspect. The trainable parameters in the basic U-net, ResDUNet and our model are 4.1, 4.3, and 4.5 million, respectively.

## 6. Conclusion

In this paper we present a robust deep learning segmentation model, that outperformed the basic U-net and ResDUNet models by 8.1% and 1.9%, respectively, with a slight increase in the network parameters. Despite the challenging nature of COVID-19 infected lesions with low contrast and variability in shape and size, our proposed model has achieved a dice similarity measure of 85.3% in reference to experts delineations. This outperformance is achieved by integrating innovative deep learning modules such as residual blocks, squeeze and excitation units, and dilated convolution. Instead of using U-net basic building blocks, we added residual blocks to contribute to feature propagation, helping the flow of low fine details throughout the network without the need for deep architecture. Also, after each residual block, squeeze and excitation unit contributed in removing redundant information, exciting informative features, and strengthening the low-level presentation. Moreover, to alleviate the variation between encoder-decoder features, we proposed the DC module that uses cascaded dilated convolution. This module maintained the spatial resolution of the features and helped capture multi-scale context information to tackle the problem of variability in shape and size of infected lesions. Each module utilized in our model had contributed in improving the model segmentation accuracy and robustness. For future work, we intend to overcome the problem of imbalanced data through investigating and further modifying the loss objective func-

tion utilized for segmentation. Moreover, we aim to build a more generalizable segmentation model designed for multiple imaging modalities with robust segmentation accuracy. Finally, to practically contribute in clinical domain we will focus on designing more simplified model with less trainable parameters.

## References

- [1] Coronavirus disease (covid-19). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. (Accessed on 06/25/2021).
- [2] Covid-19 - medical segmentation. <http://medicalsegmentation.com/covid19/>. (Accessed on 02/26/2021).
- [3] Alyaa Amer, Xujiong Ye, Massoud Zolgharni, and Faraz Janan. Resdunet: Residual dilated unet for left ventricle segmentation from echocardiographic images. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2019–2022. IEEE, 2020.
- [4] Amine Amyar, Su Ruan, Isabelle Gardin, Romain Hérault, Chatelain Clement, Pierre Decazes, and Romain Modzelewski. Radiomics-net: Convolutional neural networks on fdg pet images for predicting cancer treatment response. *Journal of Nuclear Medicine*, 59(supplement 1):324–324, 2018.
- [5] Hema Shekar Basavegowda and Guesh Dagneu. Deep learning approach for microarray cancer data classification. *CAA/ Transactions on Intelligence Technology*, 5(1):22–33, 2020.
- [6] Marcus D Bloice, Peter M Roth, and Andreas Holzinger. Biomedical image augmentation using augmentor. *Bioinformatics*, 35(21):4522–4524, 2019.
- [7] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [8] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [9] Michal Drozdal, Eugene Vorontsov, Gabriel Chartrand, Samuel Kadoury, and Chris Pal. The importance of skip connections in biomedical image segmentation. *CoRR*, abs/1608.04117, 2016.
- [10] Omar Elharrouss, Nandhini Subramanian, and Somaya Al-Maadeed. An encoder-decoder-based method for covid-19 lung infection segmentation. *arXiv preprint arXiv:2007.00861*, 2020.
- [11] Ryuhei Hamaguchi, Aito Fujita, Keisuke Nemoto, Tomoyuki Imaizumi, and Shuhei Hikosaka. Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 1442–1450, 2018.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- [13] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.
- [14] Zixing Huang, Shuang Zhao, Zhenlin Li, Weixia Chen, Lihong Zhao, Lipeng Deng, and Bin Song. The battle against coronavirus disease 2019 (covid-19): emergency management and infection control in a radiology department. *Journal of the american college of radiology*, 17(6):710–716, 2020.
- [15] Nabil Ibtehaz and M. Sohel Rahman. Multiresunet : Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 2020.
- [16] Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.
- [17] Shuo Jin, Bo Wang, Haibo Xu, Chuan Luo, Lai Wei, Wei Zhao, Xuexue Hou, Wenshuo Ma, Zhengqing Xu, Zhuozhao Zheng, et al. Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks. *MedRxiv*, 2020.
- [18] Mohsen Keshani, Zohreh Azimifar, Farshad Tajeripour, and Reza Boostani. Lung nodule segmentation and recognition using svm classifier and active contour modeling: A complete intelligent system. *Computers in biology and medicine*, 43(4):287–300, 2013.
- [19] Pedro Macías Gordaliza, María Arrate Muñoz Barrutia, Mónica Abella García, Manuel Desco Menéndez, Sally Sharpe, and Juan José Vaquero López. Unsupervised ct lung image segmentation of a mycobacterium tuberculosis infection model. 2018.
- [20] Joan Pastor-Pellicer, Francisco Zamora-Martínez, Salvador Espana-Boquera, and María José Castro-Bleda. F-measure as the error function to train neural networks. In *International Work-Conference on Artificial Neural Networks*, pages 376–384. Springer, 2013.
- [21] Xiaolong Qi, Zicheng Jiang, Qian Yu, Chuxiao Shao, Hongguang Zhang, Hongmei Yue, Baoyi Ma, Yuan Cheng Wang, Chuan Liu, Xiangpan Meng, et al. Machine learning-based ct radiomics model for predicting hospital stay in patients with pneumonia associated with sars-cov-2 infection: A multicenter study. *Medrxiv*, 2020.
- [22] Geoffrey D Rubin, Christopher J Ryerson, Linda B Haramati, Nicola Sverzellati, Jeffrey P Kanne, Suhail Raof, Neil W Schluger, Annalisa Volpi, Jae-Joon Yim, Ian BK Martin, et al. The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society. *Chest*, 158(1):106–116, 2020.
- [23] Fei Shan, Yaozong Gao, Jun Wang, Weiya Shi, Nannan Shi, Miaoifei Han, Zhong Xue, and Yuxin Shi. Lung infection quantification of covid-19 in ct images with deep learning. *arXiv preprint arXiv:2003.04655*, 2020.
- [24] Shiwen Shen, Alex AT Bui, Jason Cong, and William Hsu. An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. *Computers in biology and medicine*, 57:139–149, 2015.
- [25] Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Chao-Wei Zhao, and Ming-Ming Cheng. Jcs: An ex-

plainable covid-19 diagnosis system by joint classification and segmentation. *arXiv preprint arXiv:2004.07054*, 2020.

- [26] Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, Qin Ni, Yanfei Chen, Junwei Su, et al. A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering*, 6(10):1122–1129, 2020.
- [27] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.
- [28] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus. Deconvolutional networks. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2528–2535, 2010.