An automatic COVID-19 CT segmentation network using spatial and channel attention mechanism

Tongxue Zhou[1,2], Stéphane Canu[3], and Su Ruan[1]

[1] Université de Rouen Normandie, LITIS - QuantIF, Rouen 76183, France
[2] INSA Rouen, LITIS - Apprentissage, Rouen 76800, France
[3] Normandie Univ, INSA Rouen, UNIROUEN, UNIHAVRE, LITIS, France

Abstract. The coronavirus disease (COVID-19) pandemic has led to a devastating effect on the global public health. Computed Tomography (CT) is an effective tool in the screening of COVID-19. It is of great importance to rapidly and accurately segment COVID-19 from CT to help diagnostic and patient monitoring. In this paper, we propose a U-Net based segmentation network using attention mechanism. As not all the features extracted from the encoders are useful for segmentation, we propose to incorporate an attention mechanism including a spatial and a channel attention, to a U-Net architecture to re-weight the feature representation spatially and channel-wise to capture rich contextual relationships for better feature representation. In addition, the focal tversky loss is introduced to deal with small lesion segmentation. The experiment results, evaluated on a COVID-19 CT segmentation dataset where 473 CT slices are available, demonstrate the proposed method can achieve an accurate and rapid segmentation on COVID-19 segmentation. The method takes only 0.29 second to segment a single CT slice. The obtained Dice Score, Sensitivity and Specificity are 83.1%, 86.7% and 99.3%, respectively.

Keywords: Attention mechanism · COVID-19 · CT · Deep learning · Focal tversky loss · Segmentation

1 Introduction

In December 2019, a novel coronavirus, now designated as COVID-19 by the World Health Organization (WHO), was identified as the cause of an outbreak of acute respiratory illness [15, 25]. The pandemic of COVID-19 is spreading all over the world and causes a devastating effect on the global public health. As a form of pneumonia, the infection causes inflammation in alveoli, which fills with fluid or pus, making the patient difficult to breathe [23]. Similar to other coronaviral pneumonia such as Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS), COVID-19 can also lead to acute respiratory distress syndrome (ARDS) [7, 11]. In addition, the number of people infected by the virus is increasing rapidly. Up to April 19, 2020, 2,241,359 cases of

COVID-19 have been reported in over 200 countries and territories, resulting in approximately 152,551 deaths[4], while there is no efficient treatment at present.

Due to the fast progression and infectious ability of the disease, it's urgent to develop some tools to accurate diagnose and evaluate the disease. Although the real-time polymerase chain reaction (RT-PCR) assay of the sputum is considered as the gold standard for diagnosis, while it is time-consuming and has been reported to suffer from high false negative rates [13,21]. In clinical practice, Chest Computed tomography (CT), as a non-invasive imaging approach, can detect certain characteristic manifestations in the lung associated with COVID-19, for example, ground-glass opacities and consolidation are the most relative imaging features in pneumonia associated with SARS-CoV-2 infection. Therefore, Chest CT is considered as a low-cost, accurate and efficient method diagnostic tool for early screening and diagnosis of COVID-19. It can be evaluated how severely the lungs are affected, and how the patients disease is evolving, which is helpful in making treatment decisions [10, 12, 14, 16, 26].

A number of artificial intelligence (AI) systems based on deep learning have been proposed and results have been shown to be quite promising in medical image analysis [2, 6, 17, 20]. Compared to the traditional imaging workflow heavily relies on the human labors, AI enables more safe, accurate and efficient imaging solutions. Recent AI-empowered applications in COVID-19 mainly include the dedicated imaging platform, the lung and infection region segmentation, the clinical assessment and diagnosis, as well as the pioneering basic and clinical research [22]. Segmentation is an essential step in AI-based COVID-19 image processing and analysis for make a prediction of disease evolution. It delineates the regions of interest (ROIs), e.g., lung, lobes, bronchopulmonary segments, and infected regions or lesions, in the chest X-ray or CT images for further assessment and quantification [22]. There are a number of researches related to COVID-19. For example, Zheng et al. [28] proposed a weakly-supervised deep learning-based software system using 3D CT volumes to detect COVID-19. Goze et al. [4] presented a system that utilises 2D slice analysis and 3D volume analysis to achieve the detection of COVID-19. Jin et al. [9] proposed an AI system for fast COVID-19 diagnosis, where a segmentation model is first used to obtain the lung lesion regions, and then the classification model is used to determine whether it is COVID-19-like for each lesion region. Li et al. [12] developed a COVID-19 detection neural network (COVNet) to extract visual features from volumetric chest CT exams for distinguishing COVID-19 from Community Acquired Pneumonia (CAP). Chen et al. [3] proposed to use UNet++ [30] to extract valid areas and detect suspicious lesions in CT images.

U-net [18] is the most widely used encoder-decoder network architecture for medical image segmentation, since the encoder captures the low-level and high-level features, and the decoder combines the semantic features to construct the final result. However, not all features extracted from the encoder are useful for segmentation. Therefore, it is necessary to find an effective way to fuse features,

---

[4] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

we focus on the extraction of the most informative features for segmentation. Hu et al. [5] introduced the Squeeze and Excitation (SE) block to improve the representational power of a network by modelling the interdependencies between the channels of its convolutional features. Roy et al. [19] introduced to use both spatial and channel SE blocks (scSE), which concurrently recalibrates the feature representations spatially and channel-wise, and then combine them to obtain the final feature representation. Inspired by this work, we incorporate an attention mechanism including both spatial attention and channel one to our segmentation network to extract more informative feature representation to enhance the network performance.

In this paper, we propose a deep learning based segmentation with the attention mechanism. A preliminary conference version appeared at ISBI 2020 [29], which focused on the multi-model fusion issue. This journal version is a substantial extension, including (1) An automatic COVID-19 CT segmentation network. (2) A focal tversky loss function (different from the paper of ISBI) which is introduced to help to segment the small COVID-19 regions. (3) An attention mechanism including a spatial and a channel attention is introduced to capture rich contextual relationships for better feature representations.

The paper is organized as follows: Section 2 offers an overview of this work and details our model, Section 3 describes experimental setup, Section 4 presents the experimental results, Section 5 discusses the proposed method and concludes this work.

2 Method

2.1 The proposed network architecture

Our network is mainly based on the U-Net architecture [18], in which we integrate an attention mechanism, res_dil block and deep supervision. The encoder of the U-Net is used to obtain the feature representations. The feature representation at each layer are input into an attention mechanism, where they will be re-weighted along channel-wise and space-wise, and the most informative representations can be obtained, and finally they are projected by decoder to the label space to obtain the segmentation result. In the following, we will describe the main components of our model: encoder, decoder, and res_dil block, deep supervision and attention mechanism. The network architecture scheme is described in Fig. 1.

2.2 Encoder and decoder

The encoder is used to obtain the feature representations. It includes a convolutional block, a res_dil block followed by skip connection. In order to maintain the spatial information, we use a convolution with stride $= 2$ to replace pooling operation. It's likely to require different receptive field when segmenting different regions in an image. All convolutions are 3×3 and the number of filter is increased from 32 to 512. Each decoder level begins with upsampling layer followed by a convolution to reduce the number of features by a factor of 2. Then
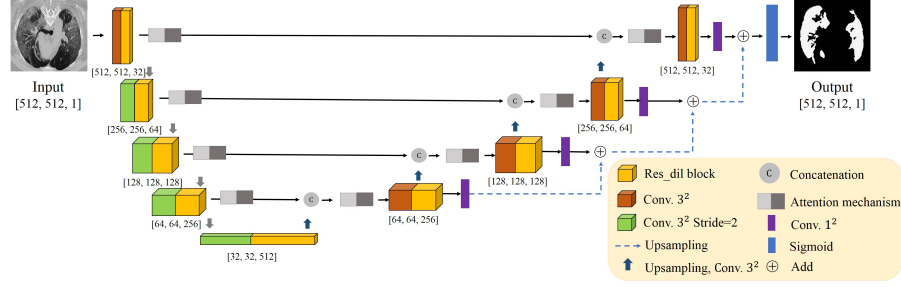
Fig. 1. The architecture of the proposed network. The network takes a CT slice as input and directly outputs the COVID-19 region.

the upsampled features are combined with the features from the corresponding level of the encoder part using concatenation. After the concatenation, we use the res_dil block to increase the receptive field. In addition, we employ deep supervision [8] for the segmentation decoder by integrating segmentation layers from different levels to form the final network output, shown in Fig. 2.

2.3   Res_dil block and deep supervision

It's likely to require different receptive field when segmenting different regions in an image. Since standard U-Net can not get enough semantic features due to the limited receptive field, inspired by dilated convolution [27], we proposed to use residual block with dilated convolutions on both encoder part and decoder part to obtain features at multiple scales, the architecture of res_dil is shown in Fig. 2. The res_dil block can obtain more extensive local information to help retain information and fill details during training process.

To demonstrate that the proposed res_dil can enlarge the receptive field mathematically, we let $F : Z^2 \to R$ be a discrete function, $\Omega_r = [-r, r]^2 \in Z^2$ and let $k : \Omega_r \to R$ be a discrete filter size $(2r + 1)^2$. The discrete convolution operator $\star$ can be described as follows:

$$(F \star k) = \sum_{m=-r}^{r} \sum_{n=-r}^{r} F(x - m, y - n)k(m, n) \tag{1}$$

Let l be a dilation factor and the l-dilated convolution operation $\star_l$ can be defined as:

$$(F \star_l k) = \sum_{m=-r}^{r} \sum_{n=-r}^{r} F(x - lm, y - ln)k(m, n) \tag{2}$$

We assume F_0, F_1,..., $F_{n-1} : Z^2 \to R$ are a discrete functions, and k_0, k_1,..., $k_{n-2} : Z^2 \to R$ are discrete 3×3 filters. In addition, we apply the filters with exponentially increasing dilation factors, such as 2^0, 2^1,... 2^{n-2}. Then, the discrete function F_{i+1} can be described as:

$$F_{i+1} = F_i \star_{2^i} k_i, i = 0, 1, ..., n - 2 \qquad (3)$$

According to the definition of receptive field, the receptive field size of each element in F_{i+1} is $(2^{i+2} - 1) \times (2^{i+2} - 1)$, which is a square of exponentially increasing size. So we can obtain a 15×15 receptive field by applying our proposed res_dil block with the dilation factor 2 and 4, respectively, while the classical convolution can only obtain 7×7 receptive field, see Fig. 3.
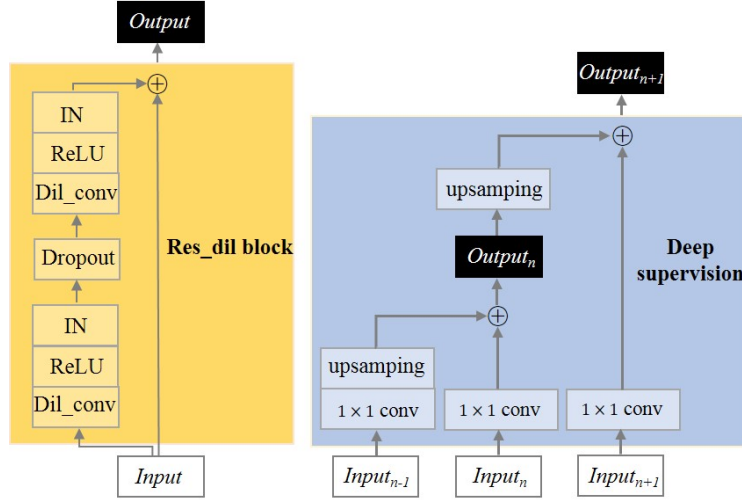
Fig. 2. The architecture of our proposed Res_dil block (left) and Deep supervision (right). IN refers instance normalization, Dil_conv the dilated convolution (rate = 2, 4, respectively). We refer to the vertical depth as level, with higher levels being higher spatial resolution. In the deep supervision part, $Input_n$ refers the output of res_dil block of the n_{th} level in the decoder, $Output_n$ refers the segmentation result of the n_{th} level in the decoder.

2.4   Attention mechanism

In U-net shaped network, not all the features obtained by the encoder are effective for segmentation. In addition, not only the different channels (filters) have various contributions but also different spatial location in each channel can give different weights on feature representation for segmentation. To this end, we introduced a "scSE based " attention mechanism in both encoder and decoder to take into account the most informative feature representations along channel-wise and spatial-wise for segmentation, the architecture is described in Fig. 4.
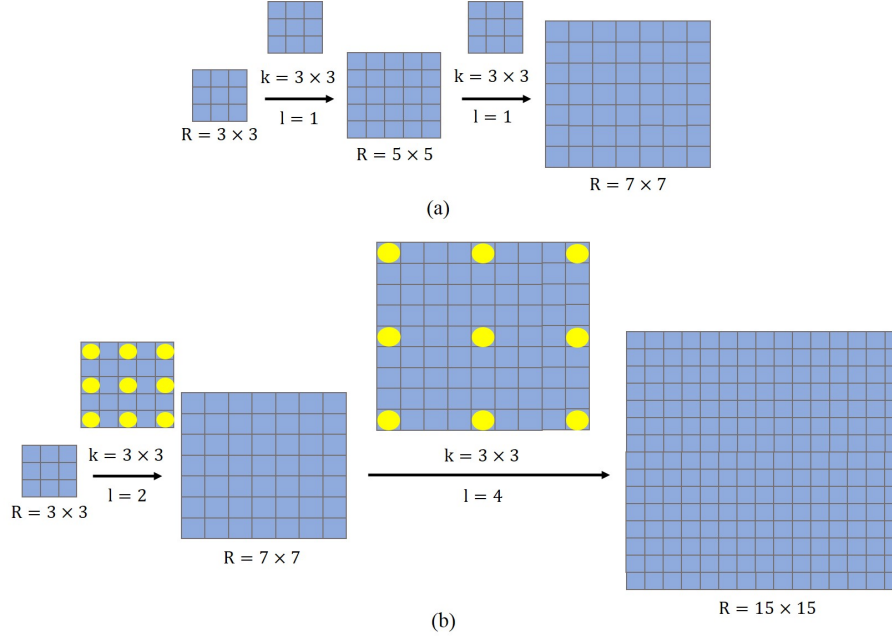
Fig. 3. The illustration of receptive field, R denotes the receptive field, k denotes the convolution kernel size, and l denotes the dilated factor. (a) a convolution network which consists of two $k = 3 \times 3$ and $l = 1, 1$ convolutional layers. (b) a convolution network which consists of two $k = 3 \times 3$ and $l = 2, 4$ dilated convolutional layers.

The individual feature representations from each channel are first concatenated as the input representation $Z = [z_1, z_2, ..., z_n]$, $Z_k \in R^{H \times W}$, n is the number of channel in each layer. To simplify the description, we take $n = 32$.

In the channel attention module, a global average pooling is first performed to produce a tensor $g \in R^{1 \times 1 \times 32}$, which represents the global spatial information of the representation, with its kth element

$$g_k = \frac{1}{H \times W} \sum_{i}^{H} \sum_{j}^{W} Z_k(i,j) \qquad (4)$$

Then two fully-connected layers are applied to encode the channel-wise dependencies, $\hat{g} = W_1(\delta(W_2 g))$, with $W_1 \in R^{32 \times 16}$, $W_2 \in R^{16 \times 32}$, being weights of two fully-connected layers and the ReLU operator $\delta(\cdot)$. $\hat{g}$ is then passed through the sigmoid layer to obtain the channel-wise weights, which will be applied to the input representation Z through multiplication to achieve the channel-wise representation Z_c, the $\sigma(\hat{g}_k)$ indicates the importance of the i channel of the representation:

$$Z_c = [\sigma(\hat{g_1})z_1, \sigma(\hat{g_2})z_2, ..., \sigma(\hat{g_{32}})z_{32},] \qquad (5)$$

In the spatial attention module, the representation can be considered as $Z = [z^{1,1}, z^{1,2}, ..., z^{i,j}, ..., z^{H,W}]$, $Z^{i,j} \in R^{1 \times 1 \times 32}$, $i \in 1, 2, ..., H$, $j \in 1, 2, ..., W$, and then a convolution operation $q = W_s \star Z$, $q \in R^{H \times W}$ with weight $W_s \in R^{1 \times 1 \times 32 \times 1}$, is used to squeeze the spatial domain, and to produce a projection tensor, which represents the linearly combined representation for all channels for a spatial location. The tensor is finally passed through a sigmoid layer to obtain the space-wise weights and to achieve the spatial-wise representation Z_s, the $\sigma(q_{i,j})$ that indicates the importance of the spatial information (i, j) of the representation:

$$Z_s = [\sigma(q_{1,1})z^{1,1}, ..., \sigma(q_{i,j})z^{i,j}, ..., \sigma(q_{H,W})z^{H,W}] \tag{6}$$

The fused feature representation is obtained by adding the channel-wise representation and space-wise representation:

$$Z_f = Z_c + Z_s \tag{7}$$

The attention mechanism can be directly adapted to any feature representation problem, and it encourages the network to capture rich contextual relationships for better feature representations.
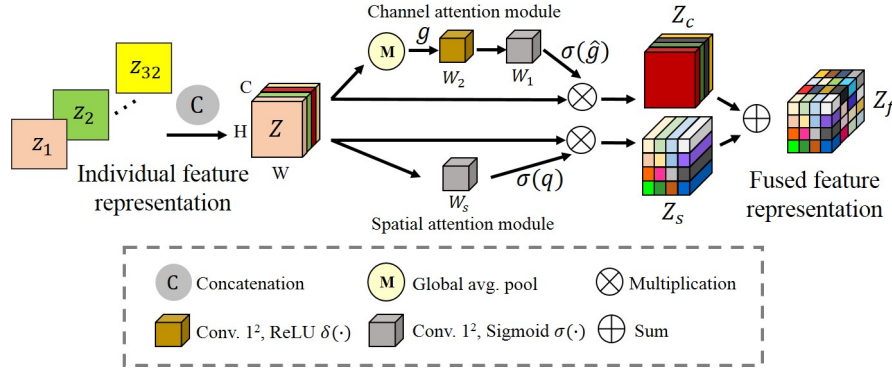
Fig. 4. The architecture of attention mechanism. The individual feature representations $(z_1, z_2, ..., z_{32})$ are first concatenated as Z, and then they are recalibrated spatially and channel-wise to achieve the Z_s and Z_c, final they are added to obtain the rich fused feature representation Z_f.

2.5   Loss function

In the medical community, the Dice Score Coefficient (DSC), defined in (5), is the most widespread metric to measure the overlap ratio of the segmented region and the ground truth, and it is widely used to evaluate segmentation

performance. Dice Loss (DL) in (6) is defined as a minimization of the overlap between the prediction and ground truth.

$$DSC_c = \frac{\sum_{j=1}^{N} p_{ic} g_{ic} + \epsilon}{\sum_{j=1}^{N} p_{ic} + g_{ic} + \epsilon} \tag{8}$$

$$DL_c = \sum_c (1 - DSC_c) \tag{9}$$

where N is the number of pixels in the image, c is the set of the classes, p_{ic} is the probability that pixel i is of the tumor class c and $p_{i\bar{c}}$ is the probability that pixel i is of the non-tumor class $\bar{c}$. The same is true for g_{ic} and $g_{i\bar{c}}$, and ϵ is a small constant to avoid dividing by 0.

One of the limitation of Dice Loss is that it penalizes false positive (FP) and false negative (FN) equally, which results in segmentation maps with high precision but low recall. This is particularly true for highly imbalanced dataset and small regions of interests (ROI) such as COVID-19 lesions. Experimental results show that FN needs to be weighted higher than FP to improve recall rate. Tversky similarity index [24] is a generalization of the DSC which allows for flexibility in balancing FP and FN:

$$TI_c = \frac{\sum_{j=1}^{N} p_{ic} g_{ic} + \epsilon}{\sum_{j=1}^{N} p_{ic} g_{ic} + \alpha \sum_{j=1}^{N} p_{i\bar{c}} g_{ic} + \beta \sum_{j=1}^{N} p_{ic} g_{i\bar{c}} + \epsilon} \tag{10}$$

Another issue with the DL is that it struggles to segment small ROIs as they do not contribute to the loss significantly. To address this, Abraham et al. [1] proposed the Focal Tversky Loss function (FTL).

$$FTL_c = \sum_c (1 - TI_c)^{1/\gamma} \tag{11}$$

where γ varies in the range $[1, 3]$. In practice, if a pixel is misclassified with a high Tversky index, the FTL is unaffected. However, if the Tversky index is small and the pixel is misclassified, the FTL will decrease significantly. To this end, we used FTL to train the network to help segment the small COVID-19 regions.

3   Experimental setup

3.1   Dataset and preprocessing

The two datasets used in the experiments come from Italian Society of Medical and Interventional Radiology: COVID-19 CT segmentation dataset [5]. Dataset-1 includes 100 axial CT images from 60 patients with Covid-19. The images have been resized, greyscaled and compiled into a single NIFTI-file. The image

---

[5] http://medicalsegmentation.com/covid19/

size is 512×512 pixels. The images have been segmented by a radiologist using three labels: ground-glass, consolidation and pleural effusion. Dataset-2 includes 9 volumes, total 829 slices, where 373 slices have been evaluated and segmented by a radiologist as COVID-19 cases. We resize these images from 630×630 pixels to 512×512 pixels same as Dataset-1. And an intensity normalization is applied to both datasets. Since there are severe data imbalance in the dataset. For example, in Dataset-1, only 25 slices have pleural effusion, which is the smallest region among all the COVID-19 lesion regions (see the green region in Fig. 4). In Dataset-2, only 233 slices have consolidation, which takes up a small amount of pixels in the image (see the yellow region in Fig. 4). We take all the lesion labels as a COVID-19 lesion. Because of the small number of data in both two datasets, we combine the two datasets as our final training dataset. Here, we give some example images of the COVID-19 CT segmentation dataset in Fig. 5.
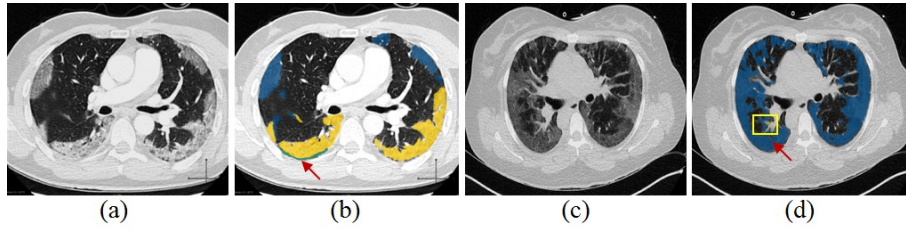
Fig. 5. Example images of the COVID-19 CT segmentation dataset. (a) and (c): CT image from Dataset-1 and Dataset-2, (b) and (d): The Ground truth of (a), (c), respectively, ground-glass is shown in blue, consolidation is shown in yellow and pleural effusion is shown in green.

3.2 Implementation details

Our network is implemented in Keras with a single Nvidia GPU Quadro P5000 (16G). The network is trained by focal tversky loss and is optimized using the Adam optimizer, the initial learning rate = 5e-5 with a decreasing learning rate factor 0.5 with patience of 10 epochs. Early stopping is employed to avoid over-fitting if the validation loss is not improved over 50 epochs. We randomly split the dataset into 80% training and 20% testing.

3.3 Evaluation metrics

Segmentation accuracy determines the eventual success or failure of segmentation procedures. To measure the segmentation performance of the proposed methods, three evaluation metrics: Dice, Sensitivity and Specificity are used to obtain quantitative measurements of the segmentation accuracy.

1) Dice: It is designed to evaluate the overlap rate of prediction results and ground truth. Dice ranges from 0 to 1, and the better predict result will have a larger Dice value.

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{12}$$

2) Sensitivity(also called the true positive rate, the recall): It measures the proportion of actual positives that are correctly identified:

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

3) Specificity(also called the true negative rate): It measures the proportion of actual negatives that are correctly identified:

$$Specificity = \frac{TN}{TN + FP} \tag{14}$$

where TP represents the number of true positive voxels, TN represents the number of true negative voxels, FP represents the number of false positive voxels, and FN represents the number of false negative voxels.

4   Experiment results

In this section, we conduct extensive comparative experiments including quantitative analysis and qualitative analysis to demonstrate the effectiveness of our proposed method.

4.1   Quantitative analysis

To assess the performance of our method, and to analyze the impact of the proposed components of our network, we did an ablation study with regard to the attention mechanism and Focal Tversky Loss function (FTL), we refer our proposed network without the attention mechanism to baseline, the results are shown in Table 1. We can observe the baseline trained with DL achieves dice score, sensitivity and specificity of 80.4%, 75.7%, 99.8%, respectively. However, using the focal tversky loss can aide the network to focus more on the false negative voxels, which increases 0.87% of dice score, 3.43% of sensitivity for Baseline and 1.96% of dice score, 13.04% of sensitivity for our proposed network. We can also observe in Table 1 that integrating the attention mechanism to the segmentation network can boost the performance, since we can see an increase of 1.37% of dice score and 1.32% of sensitivity for 'Baseline+DL' and 2.47% of dice score and 10.73% of sensitivity for 'Baseline+FTL'. The main reason is that the attention mechanism can help to emphasis on the most important feature representation for segmentation. In addition, the proposed network trained by FTL combines the benefits of attention mechanism with FTL to outperform all other methods with dice = 83.1%, sensitivity= 86.7% and specificity= 99.3%.

Table 1. Comparison of different methods on COVID-19 CT segmentation dataset, Baseline denotes our proposed network without the attention mechanism, bold results show the best score.

Model	Parameters	Dice (%)	Sensitivity (%)	Specificity (%)
Baseline + DL	$\alpha = 0.5,\ \beta = 0.5$	80.4	75.7	**99.8**
Baseline + FTL	$\alpha = 0.7,\ \beta = 0.3,\ \gamma = 4/3$	81.1	78.3	99.7
Ours + DL	$\alpha = 0.5,\ \beta = 0.5$	81.5	76.7	99.7
Ours + FTL	$\alpha = 0.7\ \beta = 0.3,\ \gamma = 4/3$	**83.1**	**86.7**	99.3

4.2   Qualitative analysis

In order to evaluate the effectiveness of our model, we randomly select several examples on COVID-19 CT segmentation dataset and visualize the results in Fig. 6. From Fig. 6, we can observe that the baseline trained by DL could give a rough segmentation result, while it fails to segment many small lesion regions. With the application of focal tversky loss, it can help to improve the segmentation result with a much better result. In addition, the attention mechanism can help to further refine the segmentation result. The proposed network trained by FTL can achieve the result closest to the ground truth. The obtained results have demonstrated that leveraging the attention mechanism and FTL can generally enhance the COVID-19 segmentation performance.

5   Conclusion

In this paper, we have presented a U-Net based segmentation network using attention mechanism. Since most current segmentation networks are trained with dice loss, which penalize the false negative voxels and false positive voxels equally, contributing a high specificity but low sensitivity. To this end, we applied the focal tversky loss to train the model to improve the small ROI segmentation performance. Moreover, we improve the baseline by incorporating an attention mechanism including a spatial attention and a channel attention in each layer to capture rich contextual relationships for better feature representations. We evaluated our proposed network on COVID-19 CT segmentation datasets, and the experiment results demonstrate the the superior performance of our method. However, the study is limited by the small dataset, in the future we would like to apply a larger training dataset to refine our model, and achieve more competitive results.

References

1. Abraham, N., Khan, N.M.: A novel focal tversky loss function with improved attention u-net for lesion segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 683–687. IEEE (2019)
2. Bui, T.D., Lee, J.J., Shin, J.: Incorporated region detection and classification using deep convolutional networks for bone age assessment. Artificial intelligence in medicine **97**, 1–8 (2019)
3. Chen, J., Wu, L., Zhang, J., Zhang, L., Gong, D., Zhao, Y., Hu, S., Wang, Y., Hu, X., Zheng, B., et al.: Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. medRxiv (2020)
4. Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P.D., Zhang, H., Ji, W., Bernheim, A., Siegel, E.: Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. arXiv preprint arXiv:2003.05037 (2020)
5. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
6. Hu, Q., Souza, L.F.d.F., Holanda, G.B., Alves, S.S., Silva, F.H.d.S., Han, T., Rebouças Filho, P.P.: An effective approach for ct lung segmentation using mask region-based convolutional neural networks. Artificial Intelligence in Medicine p. 101792 (2020)
7. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al.: Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. The Lancet **395**(10223), 497–506 (2020)
8. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In: International MICCAI Brainlesion Workshop. pp. 287–297. Springer (2017)
9. Jin, S., Wang, B., Xu, H., Luo, C., Wei, L., Zhao, W., Hou, X., Ma, W., Xu, Z., Zheng, Z., et al.: Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks. medRxiv (2020)
10. Lei, J., Li, J., Li, X., Qi, X.: Ct imaging of the 2019 novel coronavirus (2019-ncov) pneumonia. Radiology **295**(1), 18–18 (2020)
11. Li, H., Liu, S.M., Yu, X.H., Tang, S.L., Tang, C.K.: Coronavirus disease 2019 (covid-19): current status and future perspective. International Journal of Antimicrobial Agents p. 105951 (2020)
12. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., et al.: Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. Radiology p. 200905 (2020)
13. Liang, T., et al.: Handbook of covid-19 prevention and treatment. Zhejiang: Zhejiang University School of Medicine (2020)
14. Ng, M.Y., Lee, E.Y., Yang, J., Yang, F., Li, X., Wang, H., Lui, M.M.s., Lo, C.S.Y., Leung, B., Khong, P.L., et al.: Imaging profile of the covid-19 infection: radiologic findings and literature review. Radiology: Cardiothoracic Imaging **2**(1), e200034 (2020)
15. Organization, W.H., et al.: Who director-general's opening remarks at the media briefing on covid-19-11 march 2020. Geneva, Switzerland (2020)
16. Pan, F., Ye, T., Sun, P., Gui, S., Liang, B., Li, L., Zheng, D., Wang, J., Hesketh, R.L., Yang, L., et al.: Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia. Radiology p. 200370 (2020)

17. Piantadosi, G., Sansone, M., Fusco, R., Sansone, C.: Multi-planar 3d breast segmentation in mri via deep convolutional neural networks. Artificial Intelligence in Medicine **103**, 101781 (2020)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
19. Roy, A.G., Navab, N., Wachinger, C.: Concurrent spatial and channel squeeze & excitationin fully convolutional networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 421–429. Springer (2018)
20. Savelli, B., Bria, A., Molinara, M., Marrocco, C., Tortorella, F.: A multi-context cnn ensemble for small lesion detection. Artificial Intelligence in Medicine **103**, 101749 (2020)
21. Shan+, F., Gao+, Y., Wang, J., Shi, W., Shi, N., Han, M., Xue, Z., Shen, D., Shi, Y.: Lung infection quantification of covid-19 in ct images with deep learning. arXiv preprint arXiv:2003.04655 (2020)
22. Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., Shen, D.: Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. arXiv preprint arXiv:2004.02731 (2020)
23. Shi, F., Xia, L., Shan, F., Wu, D., Wei, Y., Yuan, H., Jiang, H., Gao, Y., Sui, H., Shen, D.: Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification. arXiv preprint arXiv:2003.09860 (2020)
24. Tversky, A.: Features of similarity. Psychological review **84**(4), 327 (1977)
25. Wu, J.T., Leung, K., Leung, G.M.: Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. The Lancet **395**(10225), 689–697 (2020)
26. Ye, Z., Zhang, Y., Wang, Y., Huang, Z., Song, B.: Chest ct manifestations of new coronavirus disease 2019 (covid-19): a pictorial review. European Radiology pp. 1–9 (2020)
27. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
28. Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W., Wang, X.: Deep learning-based detection for covid-19 from chest ct using weak label. medRxiv (2020)
29. Zhou, T., Ruan, S., Canu, S.: A multi-modal fusion network based on attention mechanism for brain tumor segmentation. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI 2020). Iowa, USA, April 3-7 (2020)
30. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11. Springer (2018)
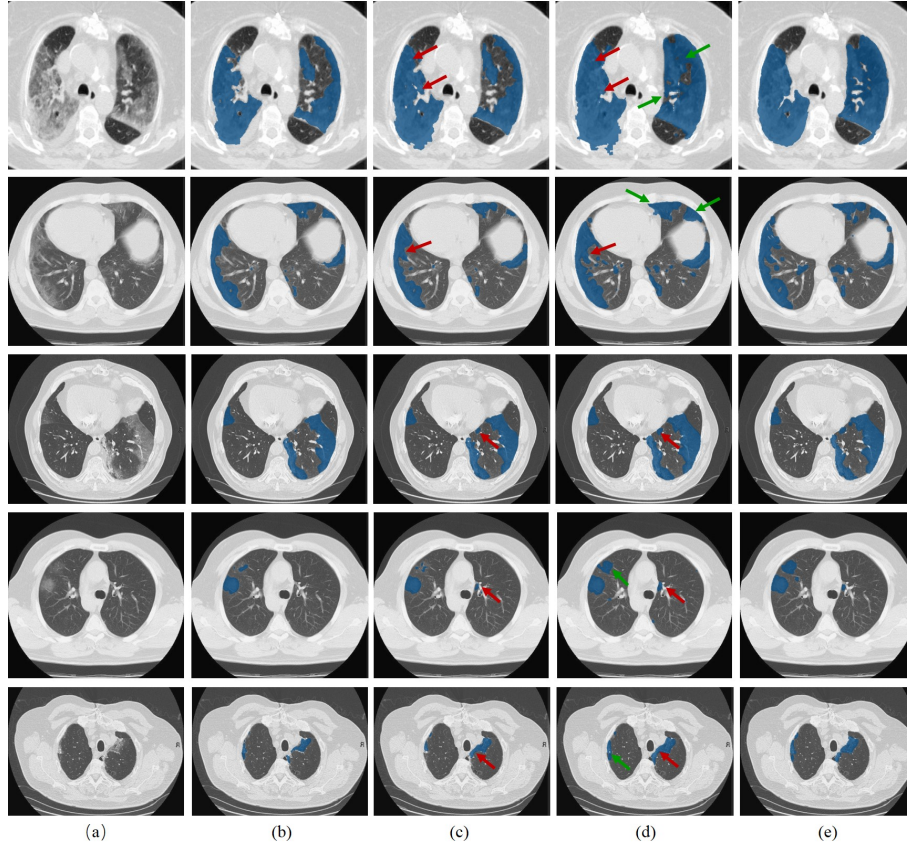
Fig. 6. Segmentation results of some examples on COVID-19 CT dataset. The first three examples are with many COVID-19 lesion regions, the last two examples are with few COVID-19 lesion regions. (a) CT image, (b) Baseline trained by dice loss, (c) Baseline trained by focal tversky loss, (d) Proposed network trained by focal tversky loss, (e) Ground truth, red arrow emphasizes the improvement of using focal tversky loss (from (b) to (c)), green arrow emphasizes the improvement of applying attention mechanism (from (c) to (d)).