Modified U-Net Based Covid-19 Lesion Segmentation Using CT Scans

xxxx xxxxx
xxxxxxxxxxxx
xxxxxxxxxx
xxxxxxx

xxxx xxxxx
xxxxxx
xxxxxxx
xxxxxx

xxxx xxxx
xxxxxx
xxxxxxx
xxxxx

***Abstract*—Computed Tomography (CT) based analysis will assist doctors in a prompt diagnosis of the Covid-19 infection. Automated segmentation of lesions in chest CT scans helps in determining the severity of the infection. The presented work addresses the task of automated segmentation of Covid-19 lesions. A U-Net framework incorporated with spatial-channel attention modules (contextual relationships), Atrous Spatial Pyramid Pooling module (a wider receptive field) and Deep Supervision (lesion focus, less error propagation) is proposed. Focal Tversky Loss is used to evaluate the outputs at coarser scales while Tversky loss evaluates the final segmentation output. This combination of losses is used to enhance segmentation of the small lesions. The framework is trained on CT scans of 20 subjects of COVID-19 CT Lung and Infection Segmentation Dataset and tested on Mosmed dataset of 50 subjects, where infection has affected less than 25% of lung parenchyma. The experimental results show that the proposed method is effective in segmenting the hard ROIs in Mosmed data resulting in a mean Dice score of 0.57 (9% more than the state-of-the-art).**

***Index Terms*—Covid-19, Lesion Segmentation, U-Net, Atrous Spatial Pyramid Pooling, Spatial-Channel Attention**

I. Introduction

The widespread outbreak of Covid-19 pandemic has affected more than 200 countries infecting more than 235 million people around the world leading to more than 4.8 million fatalities [https://www.worldometers.info/coronavirus/]. Covid-19 affects respiratory tracts and can lead to Covid-19 induced pneumonia[1]. Assessing the extent of the infection in lungs can help the medical professionals assess the severity of the disease. Computed Tomography (CT) scans are commonly used in assessing the infections in the body. Chest CT scans are an effective means to assess the severity of lung infection and hence can aid in diagnosis and severity assessment of Covid-19 infection[2]. Thus, assessing Covid-19 using CT scans, aids in an accurate and timely diagnosis enabling faster treatments and control of the spread of the disease.

Chest CT scan based analyses have been carried out mainly in diagnosis of tumour in lungs[3], Tuberculosis[4], Pulmonary Embolism[5], Pulmonary Fibrosis[6], Emphysema[7] and various types of Pneumonia[8],[9]. Thus chest CT scan based analysis of Covid-19 induced lesions would provide an insight into the extent of severity of the disease. Wang, Guotai, et al.[10] proposed COPLE-Net which was based on self-ensemble of CNNs with an adaptive teacher and an adaptive student mechanism. Additionally bridge layers are used between encoder and decoder to reduce any semantic information loss. The authors also proposed a noise robust Dice Loss to overcome noisy annotations.

Inf-Net was proposed by Fan, Deng-Ping, et al.[11] to segment the infections in CT scans for Covid-19 diagnosis. The network utilized a combination of partial parallel decoder, reverse attention and edge attention to effectively localize the infections in the CT scans. In addition to this, semi-supervised learning strategy was employed to counter the problem limited annotated data. Konar et al [12] utilized Parallel Quantum Inspired Self-supervised Network (PQIS-Net) for segmentation of CT scans forming a semi-supervised shallow learning model. A patch based classification was further carried out to assess Covid-19 infection. A variation of U-Net known as D2A U-Net was proposed by Zhao, Xiangyu, et al.[13] for an effective segmentation of Covid-19 infections. The network utilized hybrid dilated convolution with dual attention strategy, namely, Gate Attention module and Decoder Attention module.

The work presented here utilizes base framework of U-Net with attention mechanism infused at different stages of the decoder after concatenation of transposed convolution output with cropped features from encoder part. Attention mechanism has been introduced at these points of the decoder to have an effective localization by enabling the network to learn channel and spatial interdependencies of the concatenated feature map. Atrous Spatial Pyramid Pooling with residual connection is introduced after the bottleneck layer and also before the final output layer to provide a wider receptive field leading to improved localization of lesions. In addition to these changes, deep supervision [14] is carried out at different stages of decoder path. This helps prevent error propagation from lower layers of decoders to higher layers and also help in tuning the lower layers of decoder to focus on the lesions, thereby enhancing the overall segmentation.

II. Dataset Description

Two independent datasets are used here, one for training and another for testing.

Dataset 1, COVID-19 CT Lung and Infection Segmentation Dataset[15], consists of CT scans from 20 subjects (in .nii format) with corresponding infection masks as well as lung masks. Since this dataset contains a large number of annotated slices, training is carried out using this dataset.

For testing, Dataset 2, Mosmed Dataset[16], is used. This

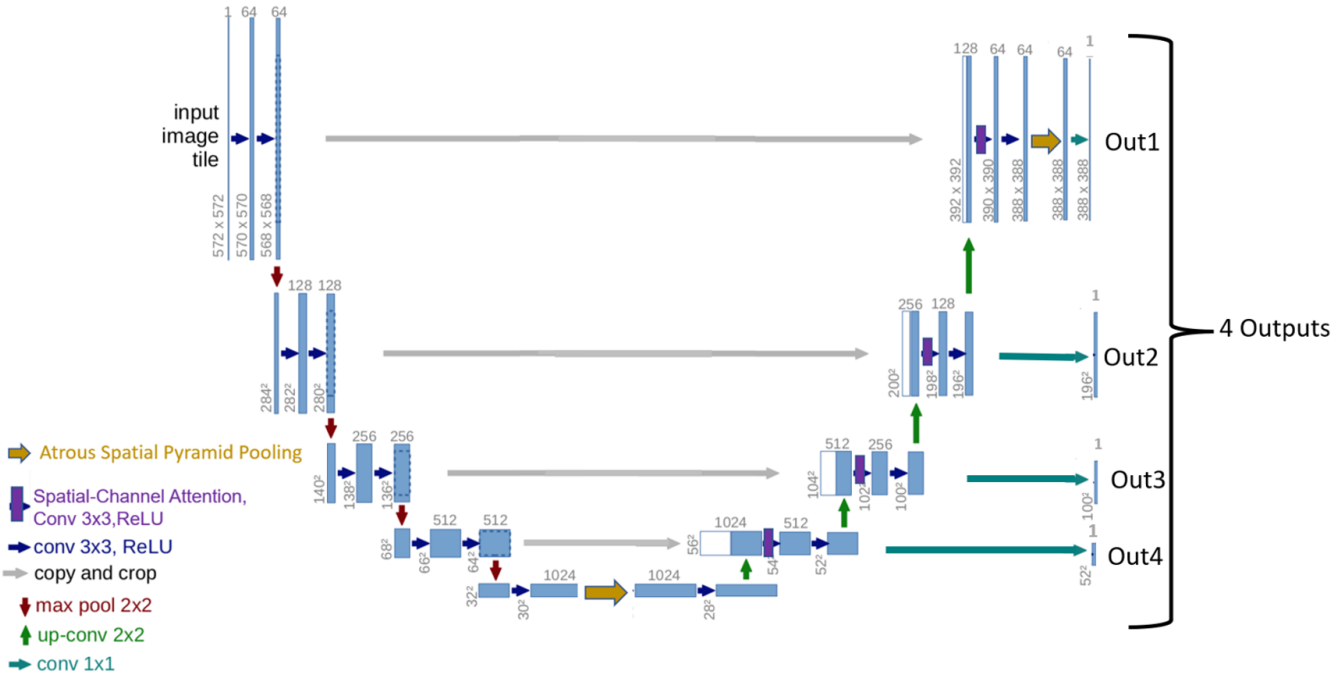

Fig. 1. Block diagram of Proposed Framework

.

dataset contains CT scans of 1110 subjects (in .nii format) but only 50 subjects have pixel-level annotations. Hence CT scans of 50 subjects are used for testing. The annotated slices belong to the group where less than 25% of lung parenchyma are infected, making the segmentation difficult due to the small ROIs.

III. Proposed Methodology

The proposed segmentation framework is shown in Figure1. The base framework is that of U-net. U-net has been modified by adding Attention module in each layer of the decoder. This is carried out to ensure that the model learns spatial and channel inter-dependencies effectively. Atrous Spatial Pyramid Pooling with residual connection has been added at the bottleneck layer as well as before the final output layer to enable the network to obtain the semantic information at multiple scales using different rates of Atrous convolution. In addition to these, the output at each level of decoder network is compared with the Ground Truth to ensure that the network focuses on the relevant features and any error occurring at the initial layers of decoder is not propagated to final output layer. The proposed network is trained using Dataset 1 and tested on Dataset 2 to ensure generalizability.

Adam optimizer is used in this network with learning rate of 0.0002 and betas = (0.9, 0.999). L2 regularization is carried out using weight decay of 0.0004.

A. U-Net

The basic U-Net model [17] is composed of encoder part (contracting path) and decoder part (expanding path) with skip connections between them. The skip connections incorporate finer details learned by the encoder into the decoder to enable effective localization for segmentation.

Encoder path is a series of convolution layers with ReLU and Max pooling layers. Convolution block consisting of two 3x3 unpadded convolutions, each followed by ReLU, is repeated multiple times as seen from the Fig. 1. Between each convolution block a 2x2 Max pooling layer is utilized for downsampling.

In the decoder part, the features first pass through a transposed convolution layer and the output is concatenated with the features from encoder layer. To match the dimensions of the feature maps of decoder, the encoder features are cropped to the dimension of decoder features. This is then followed by two 3x3 convolution and ReLU. This is repeated multiple times symmetric to encoder path. The output layer is a 1x1 convolution which outputs a single feature map since a binary segmentation is carried out in this work.

B. Attention Module

The size of CT image is 512*512 which is total of 2,62,144 pixels. Lesions occupy very small fraction of the pixels and are in diverse sizes with minimal of 150 pixels and maximum of 14080 pixels in the considered data, that is $\sim$ 0.05% to 5.37% of the image, resulting in high imbalance of region of interest (lesions) that the network needs to focus on. To make sure that the network would not miss out on the smaller lesions, Attention was introduced.

Attention modules ensure that the network focuses more on

the important details than on learning unnecessary background information[18]. The important information in this work is the lesion regions as opposed to the lung regions and the background. The attention mechanism carried out in this work is similar to that in CBAM[19], that is, Channel and Spatial attention mechanism carried out sequentially.

1) Channel Attention Module: The input feature map of 'n' channels undergoes adaptive average pooling and adaptive Max pooling in parallel resulting in two 'n' dimension channel vectors. Each of these vectors are separately passed through a 2D convolution layer that reduces the channel by a factor 'r'. Here 'r' is taken as 16. This is followed by ReLU activation and another 2D convolution layer that outputs the original channel dimension of 'n'. The two output vectors are summed and passed through a sigmoid layer to get the 'n' dimension channel attention map. This map is multiplied to the original input feature map to obtain channel weighted feature map. This feature map is then used by Spatial Attention Module to generate spatial attention map.

2) Spatial Attention Module: The channel weighted input feature map undergoes Max pooling and Average pooling in parallel along the channel axis. The outputs are then concatenated and undergo a 2D 7x7 convolution with padding=3 followed by a sigmoid layer. This results in 2D spatial attention map which is then multiplied to channel weighted input feature map. This results in spatial-channel weighted input features. Thus the input feature maps are refined to focus on *what* is important and *where* it is located.

As explained earlier, in this work, the important parts are the lesions and their locations. The attention module is applied in the decoder section after concatenation of the cropped feature map from encoder path with the output of transposed convolution in the decoder path.

C. Atrous Spatial Pyramid Pooling

Multi-scale analysis provides a means to assess the information contained in the feature map at different scales. This analysis increases the receptive field providing context information at various scales. In Atrous Spatial Pyramid Pooling (ASPP) [20], atrous convolutions are utilized to assess the feature maps at multiple scales. Dilation rates of 1,2,4,6 and 12 are used in this work.

The feature map undergoes 5 different 3x3 atrous convolutions with batch normalization and ReLU activation layers in parallel with the rates mentioned earlier. Appropriate paddings of 1,2,4,6 and 12 are utilized respectively. The outputs of these atrous convolutions is concatenated and further given to 1x1 2D convolution followed by batch normalization and ReLU activation to reduce the channels to original dimension. The resulting output is then added elementwise with the original input feature map forming a residual connection.

In this work the intention of ASPP with residual connection is to ensure a wider receptive field with reduced network parameters.

D. Loss Functions

The network assesses the effectiveness of the decoder at various stages as shown in the Figure 1. To ensure an effective segmentation, Focal Tversky Loss [21] is employed for training with the final output utilizing Tversky Loss [22] as recommended by Abraham et al[21] to prevent over-suppression while providing a strong error signal and reducing sub-optimal convergence.

Tversky Index is defined as

$$TI = \frac{\sum_{i=1}^{N} p_{ic_1} g_{ic_1} + \epsilon}{\sum_{i=1}^{N} p_{ic_1} g_{ic_1} + \alpha \sum_{i=1}^{N} p_{ic_2} g_{ic_1} + \beta \sum_{i=1}^{N} p_{ic_1} g_{ic_2} + \epsilon} \quad (1)$$

where $g_{ic} \in 0, 1$ and $p_{ic} \in 0, 1$ are ground truth and predicted label respectively. N is the total number of pixels and c_1 represents lesion class while c_2 represents non-lesion class. The hyper-parameters α and β are utilized to add weights to misclassified pixels and hence improves recall in case of class imbalance.

The Tversky Loss is calculated by minimizing $\sum(1 - TI)$.

The Focal Tversky Loss is then defined as

$$FTL = \sum (1 - TI)^{\frac{1}{\gamma}} \quad (2)$$

where γ enables the loss function to incorporate a balance between segmenting easy background and the difficult small ROIs, thus controlling the non-linearity of the loss. For this work, α, β and γ are 0.7, 0.3 and $\frac{4}{3}$ respectively. Giving γ =1, results in Tversky Loss (TL).

Thus the overall loss of the proposed architecture is

$$Loss = 0.5 * D_1 + 0.2 * D_2 + 0.2 * D_3 + 0.1 * D_4 \quad (3)$$

where

$D_1 = TL(Out1)$

$D_2 = FTL(Out2)$

$D_3 = FTL(Out3)$

$D_4 = FTL(Out4)$

and $Out1$, $Out2$, $Out3$, $Out4$ are the outputs of the network obtained at different stages as shown in the Figure 1.

To evaluate the effectiveness of the segmentation and to compare with other works which have utilized the same test data, Dice Score [23] is used.

IV. Experimental Results

Dataset 1 is used for training while Dataset 2 is used to evaluate the performance of the proposed network. Dataset 2 is chosen as slices have less than 25% lung parenchyma infection making the segmentation task difficult.

The training and test CT slices are thresholded using a lung window of (-1000 HU, 400 HU) and are normalized to [0,1]. The annotated slices are considered in this work and closed lung slices have been excluded. The slices are resized to 572x572 either by zero padding or cropping (since it was observed that lungs regions are not affected by cropping). Corresponding masks are also resized. All images are contrast

adjusted.
Training slices further undergo data augmentations of

1) Horizontal Flip
2) Vertical Flip
3) Rotation by 90 (clockwise and anti-clockwise)
4) Gamma Correction of 0.75

All the augmentations are carried out with a probability of 0.5.

The proposed framework is evaluated on Dataset 2 with Dice score. It is observed that a mean Dice Score of 0.57 is obtained by the network in the segmentation task. Fig. 2 (Rows 1, 2 and 3) shows the segmented images carried out by the framework. It can be observed that the proposed network is able to segment small lesions. However Fig. 2 (Rows 4, 5 and 6) shows failed case scenarios of the network. As can be observed from the failed cases, the amount of lesion is minute and not dense. This results in the network overlooking the lesion for some denser lung parenchyma structure for lesions.

TABLE I
COMPARISON WITH OTHER WORKS USING MOSMED DATA AS TEST DATA USING DICE SCORE

Literature	Method	Dice
Bressem, et al [24]	3D U-Net + pretrained encoder Resnet18	0.40
Lizzi, et al [25]	3D U-Net + active contour lung segmentation	0.42
Zhang, et al. [20]	CoSinGAN	0.47
Proposed	U-Net + Attention + ASPP + Deep Supervision	0.57

Table I shows the comparison of proposed framework with other works that have used Mosmed data for testing. It can be observed that the proposed methodology outperforms other works significantly indicating the effectiveness of the network in segmenting lesions where only less than 25% of lung parenchyma are infected.

TABLE II
ABALATION STUDIES.

Framework	Mean Dice Score
U-Net + Attention	0.52
U-Net + Attention + ASPP	0.54
U-Net + Attention + Deep Supervision	0.55
U-Net + Attention+ ASPP + Deep Supervision	0.57

Various combinations of proposed modules with base framework are shown in Table II along with the proposed framework. It can be observed that U-Net with Attention results in 0.52 Dice score while U-Net with Attention along with ASPP increases the Dice score to 0.54. If Deep Supervision is used instead of ASPP, the Dice score increases to 0.55. But overall highest mean Dice score of 0.57 is observed only when U-Net is combined with Attention, ASPP and Deep Supervision. Thus each of the added module enhances the overall segmentation performance.

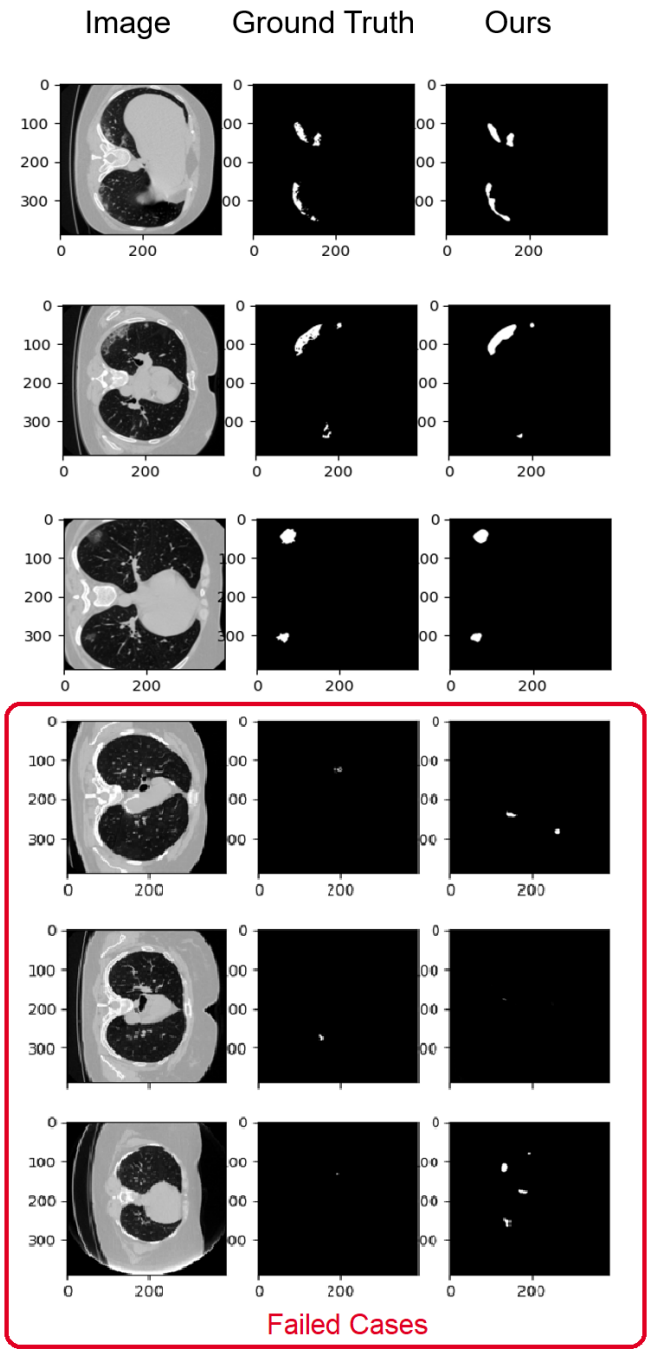

Fig. 2. Row 1, 2 and 3 : Examples of successful segmented results. Row 4, 5 and 6: Failed cases. It can be observed that the network fails when lesions are minute and not dense.

The weights given to the Loss function in Equation 3 are found experimentally. It is observed that when equal weights are given, the overall mean Dice score decreases to 0.51 while when weights of 0.9, 0.05,0.03 and 0.02 are given to D_1, D_2, D_3 and D_4, the Dice score obtained is 0.55. The highest mean

Dice score is obtained when weights are distributed as 0.5, 0.2, 0.2 and 0.1 for D_1, D_2, D_3 and D_4 respectively.
The proposed method fails when minute and light lesions are present. One possible way to overcome this would be to incorporate patch-based analysis in parallel to the presented framework. Additionally, a volume based segmentation may also prove to be effective in analyzing such lesions.

V. Conclusion

This work presents a U-Net based segmentation framework incorporating Spatial-Channel Attention to extract rich contextual information and Atrous Spatial Pyramid Pooling to provide a wider receptive field to enhance localization of lesions. Additionally a Deep Supervision is used to tune the decoder to focus on the lesions. The proposed network is trained on COVID-19 CT Lung and Infection Segmentation Dataset with Focal Tversky Loss and Tversky Loss. The evaluation of the framework is carried out on Mosmed data. The proposed framework results in a mean Dice score of 0.57 indicating the effectiveness of the framework in segmenting hard ROIs of Mosmed data. The framework needs to be analyzed further using patch based and volumetric segmentation to improve the overall performance.

References

[1] Sabine François, Carole Helissey, Sophie Cavallero, Michel Drouet, Nicolas Libert, Jean-Marc Cosset, Eric Deutsch, Lydia Meziani, and Cyrus Chargari, "Covid-19-associated pneumonia: radiobiological insights," *Frontiers in Pharmacology*, vol. 12, 2021.

[2] Damiano Caruso, Marta Zerunian, Michela Polici, Francesco Pucciarelli, Tiziano Polidori, Carlotta Rucci, Gisella Guido, Benedetta Bracci, Chiara De Dominicis, and Andrea Laghi, "Chest ct features of covid-19 in rome, italy," *Radiology*, vol. 296, no. 2, pp. E79–E85, 2020.

[3] Yutong Xie, Yong Xia, Jianpeng Zhang, Yang Song, Dagan Feng, Michael Fulham, and Weidong Cai, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest ct," *IEEE transactions on medical imaging*, vol. 38, no. 4, pp. 991–1004, 2018.

[4] Xiaohong W Gao, Carl James-Reynolds, and Edward Currie, "Analysis of tuberculosis severity levels from ct pulmonary images based on enhanced residual deep learning architecture," *Neurocomputing*, vol. 392, pp. 233–244, 2020.

[5] Shih-Cheng Huang, Tanay Kothari, Imon Banerjee, Chris Chute, Robyn L Ball, Norah Borus, Andrew Huang, Bhavik N Patel, Pranav Rajpurkar, Jeremy Irvin, et al., "Penet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric ct imaging," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–9, 2020.

[6] Wenxi Yu, Hua Zhou, Jonathan G Goldin, Weng Kee Wong, and Grace Hyun J Kim, "End-to-end domain knowledge-assisted automatic diagnosis of idiopathic pulmonary fibrosis (ipf) using computed tomography (ct)," *Medical Physics*, 2021.

[7] Stephen M Humphries, Aleena M Notary, Juan Pablo Centeno, Matthew J Strand, James D Crapo, Edwin K Silverman, David A Lynch, and Genetic Epidemiology of COPD (COPDGene) Investigators, "Deep learning enables automatic classification of emphysema pattern at ct," *Radiology*, vol. 294, no. 2, pp. 434–444, 2020.

[8] Gengfei Ling and Congcong Cao, "Atomatic detection and diagnosis of severe viral pneumonia ct images based on lda-svm," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 11927–11934, 2019.

[9] Wei Chen, Xuanqi Xiong, Bin Xie, Yuan Ou, Wenjing Hou, Mingshan Du, Yongling Chen, Kang Chen, Jing Li, Li Pei, et al., "Pulmonary invasive fungal disease and bacterial pneumonia: a comparative study with high-resolution ct," *American journal of translational research*, vol. 11, no. 7, pp. 4542, 2019.

[10] Guotai Wang, Xinglong Liu, Chaoping Li, Zhiyong Xu, Jiugen Ruan, Haifeng Zhu, Tao Meng, Kang Li, Ning Huang, and Shaoting Zhang, "A noise-robust framework for automatic segmentation of covid-19 pneumonia lesions from ct images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2653–2663, 2020.

[11] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2626–2637, 2020.

[12] Debanjan Konar, Bijaya K Panigrahi, Siddhartha Bhattacharyya, Nilanjan Dey, and Richard Jiang, "Auto-diagnosis of covid-19 using lung ct images with semi-supervised shallow learning network," *IEEE Access*, vol. 9, pp. 28716–28728, 2021.

[13] Xiangyu Zhao, Peng Zhang, Fan Song, Guangda Fan, Yangyang Sun, Yujia Wang, Zheyuan Tian, Luqi Zhang, and Guanglei Zhang, "D2a u-net: Automatic segmentation of covid-19 ct slices based on dual attention and hybrid dilated convolution," *Computers in biology and medicine*, p. 104526, 2021.

[14] Guodong Zeng, Xin Yang, Jing Li, Lequan Yu, Pheng-Ann Heng, and Guoyan Zheng, "3d u-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3d mr images," in *International workshop on machine learning in medical imaging*. Springer, 2017, pp. 274–282.

[15] Ma Jun, Ge Cheng, Wang Yixin, An Xingle, Gao Jiantao, Yu Ziqi, Zhang Minqing, Liu Xin, Deng Xueyuan, Cao Shucheng, Wei Hao, Mei Sen, Yang Xiaoyu, Nie Ziwei, Li Chen, Tian Lu, Zhu Yuntao, Zhu Qiongjie, Dong Guoqiang, and He Jian, "COVID-19 CT Lung and Infection Segmentation Dataset," Apr. 2020.

[16] Sergey P Morozov, Anna E Andreychenko, Ivan A Blokhin, Pavel B Gelezhe, Anna P Gonchar, Alexander E Nikolaev, Nikolay A Pavlov, Valeria Yu Chernina, and Victor A Gombolevskiy, "Mosmeddata: data set of 1110 chest ct scans performed during the covid-19 epidemic," *Digital Diagnostics*, vol. 1, no. 1, pp. 49–59, 2020.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[18] Nils Gessert, Julia Krüger, Roland Opfer, Ann-Christin Ostwaldt, Praveena Manogaran, Hagen H Kitzler, Sven Schippling, and Alexander Schlaefer, "Multiple sclerosis lesion activity segmentation with attention-guided two-path cnns," *Computerized Medical Imaging and Graphics*, vol. 84, pp. 101772, 2020.

[19] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[20] Pengyi Zhang, Yunxin Zhong, Yulin Deng, Xiaoying Tang, and Xiaoqiong Li, "Cosingan: learning covid-19 infection segmentation from a single radiological image," *Diagnostics*, vol. 10, no. 11, pp. 901, 2020.

[21] Nabila Abraham and Naimul Mefraz Khan, "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)*. IEEE, 2019, pp. 683–687.

[22] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour, "Tversky loss function for image segmentation using 3d fully convolutional deep networks," in *International workshop on machine learning in medical imaging*. Springer, 2017, pp. 379–387.

[23] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B Blaschko, "Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 92–100.

[24] Keno K Bressem, Stefan M Niehues, Bernd Hamm, Marcus R Makowski, Janis L Vahldiek, and Lisa C Adams, "3d u-net for segmentation of covid-19 associated pulmonary infiltrates using transfer learning: State-of-the-art results on affordable hardware," *arXiv preprint arXiv:2101.09976*, 2021.

[25] Francesca Lizzi., Francesca Brero., Raffaella Cabini., Maria Fantacci., Stefano Piffer., Ian Postuma., Lisa Rinaldi., and Alessandra Retico., "Making data big for a deep-learning analysis: Aggregation of public covid-19 datasets of lung computed tomography scans," in *Proceedings of the 10th International Conference on Data Science, Technology and Applications - DATA,*. INSTICC, 2021, pp. 316–321, SciTePress.