A DEEP LEARNING APPROACH FOR IMPROVED SEGMENTATION OF LESIONS RELATED TO COVID-19 CHEST CT SCANS

Vlad Vasilescu[1], *Ana Neacșu*[1,2], *Emilie Chouzenoux*[2], *Jean-Christophe Pesquet*[2], *Corneliu Burileanu*[1]

[1] *Speech and Dialogue Laboratory*
University Politehnica of Bucharest, Bucharest, Romania
[2] *Université Paris-Saclay, CentraleSupélec,*
Centre de Vision Numérique, Inria, Gif-sur-Yvette, France

ABSTRACT

The current coronavirus pandemic (COVID-19) became a world-wide threat, infecting more than 42 million people since its outbreak in early 2020. Recent studies show that analyzing chest CT scans plays an essential role in assessing disease progression and facilitates early diagnosis. Automatic lesion segmentation constitutes a useful tool to complement more traditional healthcare system strategies to address the COVID-19 crisis. We introduce *MASC-Net*, a novel deep neural network that automatically detects COVID-19 related infected lung regions from chest CT scans. The proposed architecture consists of a multi-input encoder-decoder that aggregates high-level features extracted with variable-size receptive fields.

Index Terms— COVID-19, Computer Tomography, lesion segmentation

1. INTRODUCTION

Since the first cases of infection in China, the ongoing COVID-19 global pandemic has affected more than 42 million people (October 2020 estimation), claiming the lives of more than one million others [1, 2, 3]. To face this dramatic situation, the community of researchers worldwide started to provide active solutions to help the healthcare system to fight against this growing threat. In multiple studies, radiologists have reported the presence of specific lesions in the CT scans of COVID-19 patients, the two most frequently encountered ones being *Ground Glass Opacities* and *Pulmonary Consolidation* [4, 5]. In particular, it appears that an accurate volumetric estimation of these lesions is useful to evaluate the severity of the infection and provides valuable information in terms of prescription of the most suitable clinical treatment [6]. In many countries, it is indeed of paramount importance to manage efficiently the available resources in terms of intensive care beds. Unfortunately, accurate quantification of chest CT lung lesions is very time-consuming even for experienced radiologists, so that large-scale manual analysis is virtually impossible [7].

The use of deep learning methods in analyzing chest CT scans in the COVID-19 context is an active research area, focusing on tasks such as lung segmentation, COVID-19-related lesion segmentation, and the classification of these lesions. One of the most challenging tasks is lesion segmentation, which aims at localizing all lesions, indiscriminately. The authors of [8] propose a semi-supervised system, while [9] introduces a modified *VNet* [10] structure to detect COVID-19 from 3D volumes. Since handling 3D structures may be unnecessarily costly from both resource and time viewpoints, some recent works focus on segmenting lesions from CT slices. In [11] *Inf-Net*, a semi-supervised segmentation system is described, which is based on a partial parallel decoder structure, using multi-level deep supervision to infer the segmentation mask at different encoding levels. For other CT slice-based approaches, we refer the reader to [12, 13, 14].

In this work, we propose a new neural network architecture called *MASC-Net (Multi-level Atrous Spatial Convolution Network)* to improve automatic lesion segmentation in chest CT scans. It employs convolution layers with different dilation rates, allowing the network to aggregate information regarding areas of different sizes, followed by an encoder module that progressively extracts higher dimensional representations by processing local information layer by layer. This approach allows for better pixel class separation in the computed higher dimensional space.

The rest of the paper is organized as follows. Section 2 presents our main contribution, which is a new neural network approach for lesion segmentation, by detailing each part of the proposed system. Section 3 describes the experimental setup and compares the obtained results with those produced by other existing approaches in the literature. Finally, Section 4 contains some concluding remarks.

2. PROPOSED METHOD

The proposed architecture is detailed in Figure 1. It has an encoder-decoder-like structure, with skip connections, that inputs the original CT scan slice and the histogram-equalized

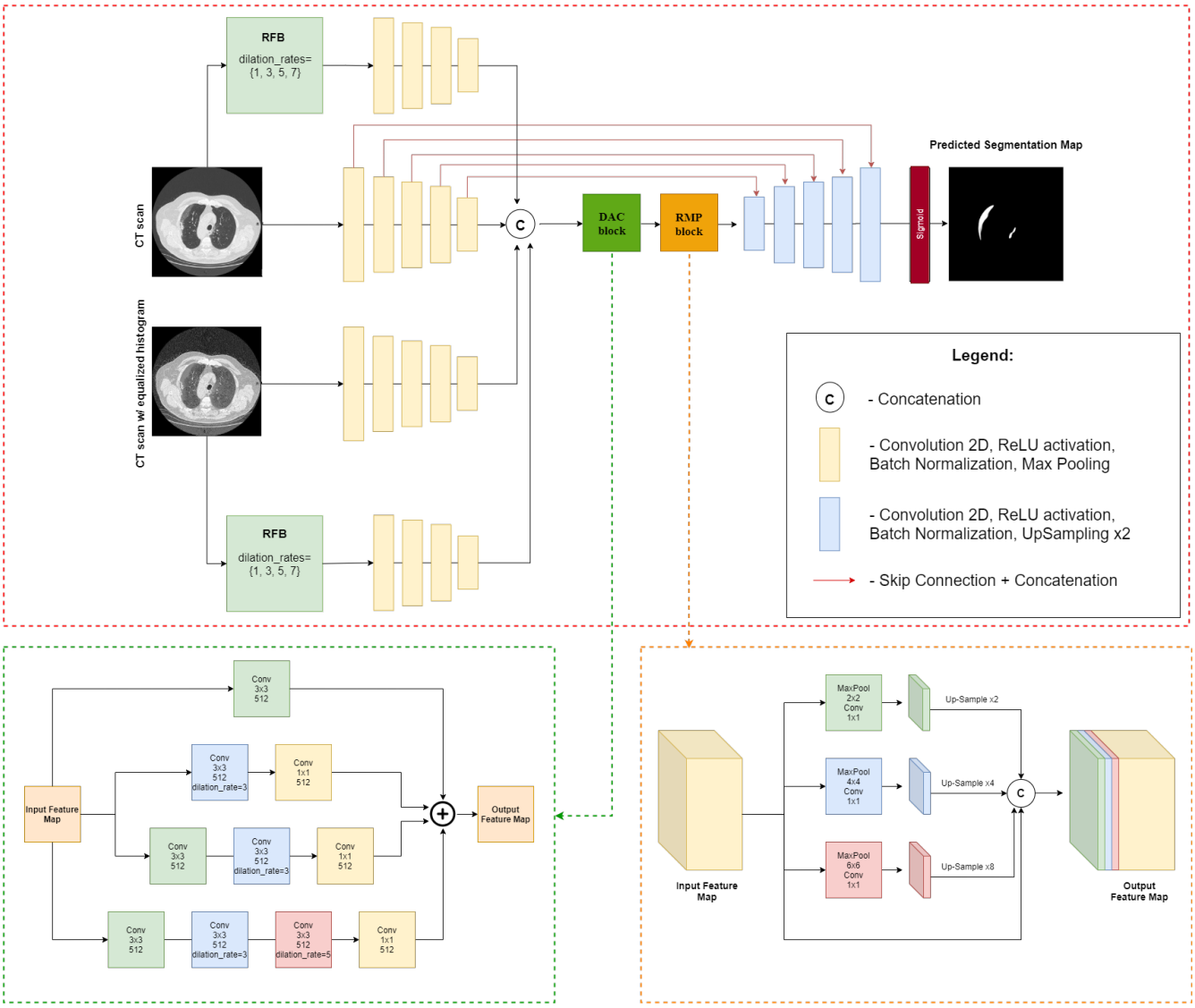

Fig. 1: Proposed architecture of Multi-level Atrous Spatial Convolution Network (MASC-Net). Top-figure represents the overall configuration of the network, an encoder-decoder structure with different feature extraction blocks. Bottom-left and bottom-right figures detail the Dense Atrous Convolution (DAC) and Residual Multi-kernel Pooling (RMP) blocks, respectively.

version of it, together with other relevant features. All this encoded information is concatenated into a single feature map, that is passed through a Dense Atrous Convolution (DAC) block, followed by a Residual Multi-kernel Pooling (RMP) block. The resulting features are fed to the decoder part. Following the decoding phase, the final predictions are computed by thresholding the output. Each block of the proposed network is detailed below.

2.1. Receptive Field Block (RFB)

The RFB consists of multiple branches with different kernel sizes and dilated convolution layers. This type of structure has been used in accurate and fast object detection, by taking into account variable-distanced pixels, while maintaining roughly the same number of parameters for the convolution operation. In our approach, we use a RFB structure comprised of 4 branches that contains 4 different dilation rates, e.g. 1,3,5,7 (1 corresponds to the standard 2D Convolution) and a residual connection. For each of our inputs we apply an RFB in order to extract relevant features characterising regions with 4 distinctive sizes.

2.2. Dense Atrous Convolution block (DAC) and Residual Multi-kernel Pooling block (RMP)

Inspired by [15], we pass our concatenated high-level feature map to a DAC block in order to further extract significant information by using 4 branches with different dilated convolution rates, followed by 1×1 Convolution blocks to achieve the same dimension for the filter space in each branch. The proposed RMP encodes the resulting features by using multiple pooling layers with different receptive field sizes and a

1×1 Convolution operation to reduce the computational cost. The results from these 3 pooling branches are then concatenated alongside the original feature map in order to produce the output which is then passed to our decoder module.

2.3. Encoder-Decoder

The encoder part of our proposed architecture which acts as a feature extractor takes as input the original CT scan, the equalized histogram version of it, and the 2 feature maps generated by passing each input slice through an RFB, successively down-sampling the input images to a small size abstract feature space. The reason for using the CT slice along with its equalized version of it as input lies in the fact that some lesion areas are of low-contrast, relative to the majority of annotated lesions. Each Encoder block consists of a 2D Convolution layer with a 3×3 kernel, a padding of 1, and various filters to extract feature representations from slices, followed by *ReLU* activation function, a *Batch Normalization* step, and a 2×2 *Max Pooling* layer. The 4 produced high-level feature maps are then concatenated and passed-down to the DAC and the RMP blocks, followed by 5 decoding blocks, each being composed by layers similar to those that our encoder blocks contain, except that the *Max Pooling* operation is replaced by a 2×2 up-sampling operation using *bilinear interpolation* in order to restore the high-level features back to the input slice resolution. While 2D Convolution is used when extracting relevant information in our encoding layers, max-pooling leads to a loss of semantic information. In order to overcome this limitation, we add skip connections from convolution layers in encoding blocks to the mirrored up-sampling layers from decoding blocks. The encoded feature maps are thus concatenated with the up-sampled high-level decoded maps to account for the loss of information due to successive down-sampling.

3. EXPERIMENTS AND RESULTS

3.1. Experimental setup

Let $V \in \mathbb{R}^{D \times H \times W}$ denote a volume which corresponds to a chest CT. We aim at classifying each voxel value $V_{i,j,k} \in \mathbb{R}$ with $(i, j, k) \in \{1, \ldots, D\} \times \{1, \ldots, H\} \times \{1, \ldots, W\}$ into one of the following classes: non-lesion and lesion. After this segmentation, each lesion voxel could be classified into different lesion types. We used as input data *slices* $S \in \mathbb{R}^{H \times W}$ extracted from volume V.

We trained our model on an online dataset[1] which contains 9 annotated volumes of chest CT scans (from different patients). We split the dataset into 2 distinct sub-sets for training (6 volumes, 200 slices) and testing (3 volumes, 173 slices), using only slices that contain lesions. Each volume has been truncated to $[-1000, 1000]$ on the Hounsfield scale, then normalized in the range $[0, 1]$, and reshaped in order to obtain slices of 512×512 pixels.

To further assess the performance of our approach, we tested it on another dataset, consisting of slice by slice annotated CTs, acquired with a different scan. The test dataset, detailed in [16], contains 10 annotated volumes from Coronacases Initiative. In our experiments, we selected slices that contain infected areas, totaling approximately 500 slices.

The proposed structure was trained with ADAM optimizer [17] with a learning rate $\gamma = 10^{-3}$ and exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$, using *dice loss*, a popular choice in segmentation tasks [18]. The weights were initialized using *Xavier uniform* distribution.

3.2. Evaluation metrics

To estimate the performance of our proposed method, we report several evaluation metrics, including Dice Coefficient Score (DSC), Sensitivity (i.e. Recall), Specificity, Precision and F-score. These metrics are widely used for assessing the accuracy in different medical tasks and are defined as follows:

$$\text{DSC} = \frac{2|A \cap B|}{|A| + |B| + \epsilon}, \tag{1}$$

$$\text{Sensitivity}: \quad r = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2}$$

$$\text{Specificity}: \quad s = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{3}$$

$$\text{Precision}: \quad p = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4}$$

$$\text{F-Score} = \frac{2pr}{p + r}. \tag{5}$$

Here-above A, B, TP, TN, FN, FP denote the ground-truth, the predicted mask, the number of true positives, true negatives, false negatives, false positives, respectively. In our case, a positive refers to a pixel that is labeled as lesion, while negatives represent non-lesion pixels.

3.3. Results and discussion

We experimented with different versions of MASC-Net. Firstly, we trained an Encoder-Decoder model, feeding the the network with the original CT slice, its equalized histogram version, alongside the features extracted by the RFB computed from the normalized CT slice. Secondly, we added an additional branch which encodes features obtained by the RFB on the equalized histogram CT slice. This resulted in a significant increase of the mean DSC on validation set, asserting the relevance of the additional features extracted from the equalized histogram CT slice. For the final structure of MASC-Net, we added skip connections from encoding to decoding blocks to account for information loss during down-sampling operation. We also added 2 feature encoding blocks, respectively DAC and RMP blocks which operate on the high-level concatenated encoded feature-maps.

[1] https://medicalsegmentation.com/covid19/

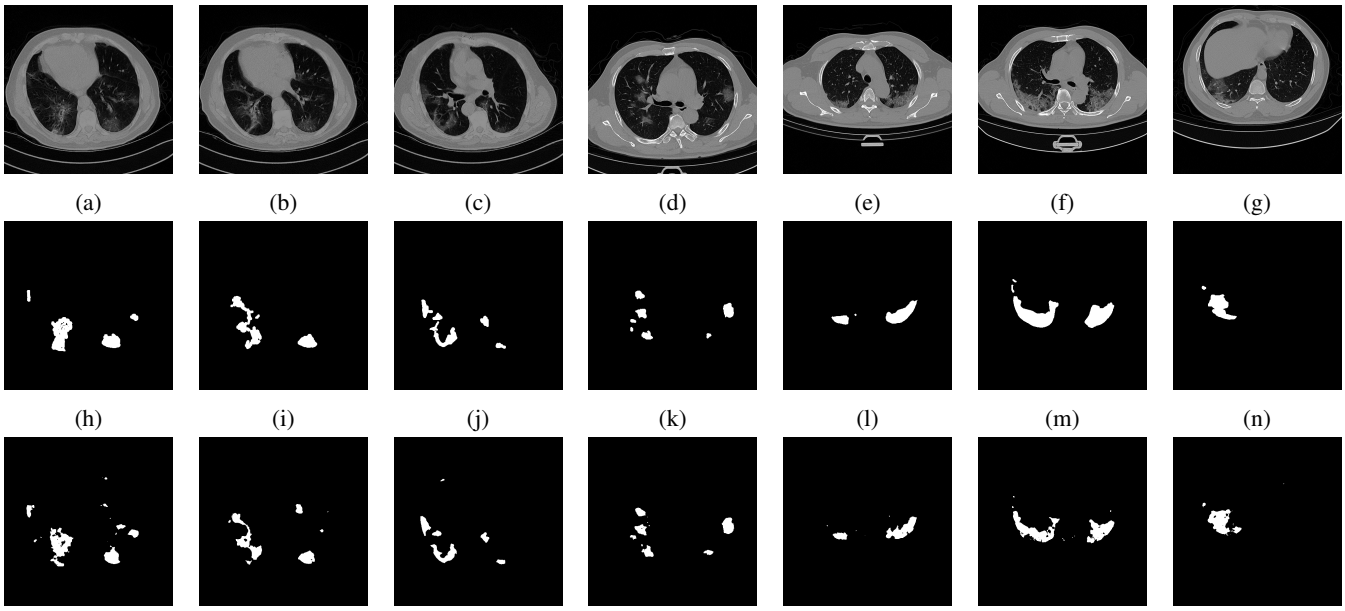

Fig. 2: Validation examples from both datasets. (a), (b), (c) are examples from dataset[1], (d), (e), (f), (g) from dataset [16]. First row: normalized CT slices. Second row: corresponding infection masks. Third row: infection masks predicted by *MASC-Net*

	DSC	Sensitivity	Specificity	Precision	F-score
U-Net	0.492	0.456	0.996	0.657	0.492
Inf-Net	0.664	0.674	**0.997**	**0.789**	0.632
Encoder-Decoder + RFB CT scan	0.515	0.495	0.996	0.564	0.477
Encoder-Decoder + RFB CT scan + RFB CT scan w/ e.h	0.665	0.802	0.996	0.608	0.639
MASC-Net	**0.707**	**0.803**	0.996	0.671	**0.707**

Table 1: Average performance measures on test set from dataset[1]; *e.h.* stands for *equalized histogram*

Table 1 summarizes our results computed as a mean over the test set. We compare the performance of our proposed architecture with other state-of-the-art ones used for lesion segmentation, namely UNet [19] and InfNet [11]. Note that all models were trained on a small number of scans. Even in this setting, the performance of MASC-Net in terms of DSC is comparable with recent structures trained on much larger datasets; see for example [13].

To test the generalization capabilities of each system in the context of training with limited data, we tested all the models on a different dataset, in an attempt to simulate how such system would function in real-life conditions, e.g. a hospital were CTs from different scanners may need to be analysed. The results detailed in Table 2 show that MASC-Net outperforms the other models in terms of DSC, Sensitivity, Precision, and F-Score, proving the adaptability of the architecture to new data. The performance of our system can be compared with results in [11], where the reported DSC value was 0.59. Some visual results of masks predicted using MASC-Net are presented in Figure 2.

	DSC	Sen.	Spec.	Precision	F-score
U-Net	0.485	0.377	**0.999**	0.810	0.485
Inf-Net	0.516	0.438	0.998	0.747	0.512
MASC-Net	**0.655**	**0.575**	0.998	**0.821**	**0.657**

Table 2: Average performance measures on test dataset [16]

4. CONCLUSION

This paper has presented MASC-Net, a novel lung CT lesion segmentation architecture using a four-level encoder-decoder module and multiple feature extractor blocks. Trained on a small dataset, accurate results are obtained for lesion segmentation tasks. Indeed, the proposed structure outperforms state-of-the-art models in terms of DSC and Sensitivity. We also provided performance measures on a second larger dataset, showing that MASC-Net exhibits good generalization capabilities, an essential aspect in medical segmentation tasks. An estimation of the COVID-19 disease extent could be deduced by estimating the volume of the segmented lesions.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by dataset[1] and [16]. Ethical approval was not required as confirmed by the license attached with the open access data.

6. ACKNOWLEDGMENTS

No funding was received for conducting this study. The authors have no relevant financial or non-financial interests to disclose.

7. REFERENCES

[1] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei *et al.*, "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, no. 7798, pp. 265–269, 2020.

[2] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu *et al.*, "A novel coronavirus from patients with pneumonia in China, 2019," *New England J. Med.*, vol. 382, no. 8, pp. 727–733, 2020.

[3] L. T. Phan, T. V. Nguyen, Q. C. Luong, T. V. Nguyen, H. T. Nguyen, H. Q. Le, T. T. Nguyen, T. M. Cao, and Q. D. Pham, "Importation and human-to-human transmission of a novel coronavirus in Vietnam," *New England J. Med.*, vol. 382, no. 9, pp. 872–874, 2020.

[4] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R. L. Hesketh, L. Yang *et al.*, "Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia," *Radiology*, 2020.

[5] M.-P. Revel, A. P. Parkar, H. Prosch, M. Silva, N. Sverzellati, F. Gleeson, A. Brady, E. S. of Radiology, and E. S. of Thoracic Imaging, "COVID-19 patients and the radiology department–advice from the european society of radiology (ESR) and the european society of thoracic imaging (ESTI)," *European Radiology*, vol. 30, no. 9, pp. 4903–4909, Sep. 2020.

[6] N. Lassau, S. Ammari, E. Chouzenoux, H. Gortais, P. Herent, M. Devilder, S. Soliman, O. Meyrignac, M.-P. Talabard, J.-P. Lamarque *et al.*, "AI-based multi-modal integration of clinical characteristics, lab tests and chest CTs improves COVID-19 outcome prediction of hospitalized patients," 2020, https://www.medrxiv.org/content/10.1101/2020.05.14.20101972v3.

[7] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," *IEEE Rev. in Biomed. Eng.*, 2020, (to appear).

[8] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for COVID-19 from chest CT using weak label," 2020, https://www.medrxiv.org/content/10.1101/2020.03.12.20027185v2.

[9] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, and Y. Shi, "Lung infection quantification of Covid-19 in ct images with deep learning," *arXiv preprint arXiv:2003.04655*, 2020.

[10] F. Milletari, N. Navab, S. Ahmadi, and V-Net, "Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. Int. Conf. 3D Vis.*, California, USA, 25-28 Oct. 2016, pp. 565–571.

[11] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-Net: Automatic COVID-19 lung infection segmentation from CT images," *IEEE Trans. Med. Imag.*, 2020.

[12] S. Chaganti, A. Balachandran, G. Chabin, S. Cohen, T. Flohr, B. Georgescu, P. Grenier, S. Grbic, S. Liu, F. Mellot *et al.*, "Quantification of tomographic patterns associated with covid-19 from chest ct," *arXiv preprint arXiv:2004.01279*, 2020.

[13] G. Chassagnon, M. Vakalopoulou, E. Battistella, S. Christodoulidis, T.-N. Hoang-Thi, S. Dangeard, E. Deutsch, F. Andre, E. Guillo, N. Halm *et al.*, "AI-Driven CT-based quantification, staging and short-term outcome prediction of COVID-19 pneumonia," *arXiv preprint arXiv:2004.12852*, 2020.

[14] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, "A noise-robust framework for automatic segmentation of Covid-19 pneumonia lesions from CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2653–2663, 2020.

[15] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, 2019.

[16] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng, W. Hao, M. Sen, Y. Xiaoyu, N. Ziwei, L. Chen, T. Lu, Z. Yuntao, Z. Qiongjie, D. Guoqiang, and H. Jian, "COVID-19 CT Lung and Infection Segmentation Dataset," Apr. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3757476

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Int. Conf. Learning Representations*, 7–9 May 2015.

[18] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 240–248.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing Computer-Assisted Intervention Conf.* Springer, 2015, pp. 234–241.