CovSegNet: An Automated COVID-19 lesion segmentation from CT scans using Deep Learning Techniques

Cezar Mbiethieu, Norbert Tsopze, Engelbert Mephu-Nguifo

CovSegNet: An Automated COVID-19 lesion segmentation from CT scans using Deep Learning Techniques

Cezar Mbiethieu[*1,2], **Norbert Tsopze**[1,4], **Engelbert Mephu-Nguifo**[3]

[1]Department of Computer Science, University of Yaounde I, P.o Box 812 Yaounde, Cameroon
[2]Department of Computer Engineering, College of Technology, University of Bamenda, Bamenda, Cameroon
[3]CNRS, ENSMSE, LIMOS, Clermont Auvergne University, Clermont-Ferrand, France
[4]Sorbonne Universite, IRD, UMMISCO, F-93143, Bondy, France

*E-mail : mbiethieucezar@gmail.com

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 lesions from CT images 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 efficient attention mechanism, to an U-Net architecture to re-weight the feature representation channel-wise to capture rich contextual relationships for better features representation while maintaining a better time complexity. In addition, a novel subsampling techniques for CNN is introduced to significantly increase the amount of information kept by feature maps through subsampling layers and we finally use a focal Tversky loss to deal with small lesion segmentation. The experiment results, evaluated on a COVID-19 CT scans segmentation dataset where 3250 CT slices are available, demonstrate the proposed method can achieve an accurate and rapid segmentation on COVID-19 segmentation. The method takes only a fraction of second to segment a single CT slice.

Keywords

COVID-19; CT scans ; Deep Learning ; U-net ; SubSampling ; Attention mechanism

I INTRODUCTION

The COVID-19 pandemic caused by the virus named SARS-CoV-2 identified in 2019, has been spread all over the world and has caused a devastating effect on the global public health. Researches have shown that the coronavirus is spread through droplets and virus particles released in the air when an infected person breathes, talks, laughs, sings, coughs or sneezes. The traditional method used to diagnose and evaluate the disease is named real-time polymerase chain reaction (RT-PCR) which assays the sputum; it has a high specificity, but is time consuming, laborious and has been reported for high false negative rates Liang [8]. Chest Computed tomography (CT) and Chest X-ray are two imaging approaches used to detect certain characteristics manifestations in the lung infected by COVID-19. CT is more sensitive, especially for early stage of infection. 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 used to evaluate how severely

the lungs are affected, and how the patients disease is evolving, which is helpful in making treatment decisions. The evaluation of localization and geometric features of infection area could provide adequate information of disease progress and help doctors make better treatment.

One of the most widely used method to segment region of interest(ROI) in medical images is a deep learning system called U-net presented by Ronneberger, Fischer, and Brox [1] which has an architecture in "U" form based on encoder - decoder that captures low level and high level resolution features at the encoder phase and combines with the semantic features extracted at the decoder phase using skip connections. However, the features captured from the encoder part are not all useful for the segmentation task. Therefore, it is necessary to find an effective way to fuse features of the encoder and that of the decoder; we focus on the extraction of the most informative features for segmentation. Recently, incorporation of channel attention into convolution blocks has attracted a lot of interests, showing great potential in performance improvement of deep CNNs which extract the most informative features for segmentation. Some of the representative methods include squeeze-and-excitation networks (SENet) of Hu, Shen, and Sun [2] and it variants, which learns channel attention for each convolution block, bringing clear performance gain for various deep CNN architectures. Although these methods have achieved higher accuracy, more often, they bring higher model complexity and suffer from heavier computation burden due to complex architecture. Through the deep analysis of different types of layers in a deep learning architectures, we discovered that we usually skip some important features which could significantly increase the amount of information kept by the feature maps, therefore improving the models accuracy.

To address the above issues, we focus our research on finding a method which will establish a tradeoff between a better performance and a low model complexity cost. In this paper, we propose CovSegNet (Covid-19 Segmentation Network), a deep learning based segmentation system made with improved convolution and pooling layers to significantly increase the amount of information kept by the feature maps; we also proposed to used an improved channel attention mechanism which reduces the complexity of the segmentation models while maintaining their accuracy to segment the covid-19 lesions. We revisit the existing layers in deep architectures which perform subsampling and the existing attention mechanism in order to tackle their limitations in medical image segmentation. Our contributions include:

1. The proposition of a novel subsampling technique in encoder-decoder models to significantly increase the amount of information kept by the feature maps.
2. We propose a novel attention mechanism based Deep CNN architecture to segment COVID-19 CT scans.
3. We built a new U-net based architecture which integrates our proposed subsampling and attention mechanism in the classical skip connection, and also we replace the classical subsampling by our proposed lighter subsampling module.

The rest of this paper is organized as follows. In section 2, we provide an overview of the segmentation models and some related literature using attention mechanism. Section 3 represents our proposed CovSegNet. Section 4 will describe the dataset and the experiments to demonstrate the potential power of our proposed method. Finally, we will present the conclusion in section 5.

II RELATED WORK

Most cutting-edge segmentation algorithms are based on deep learning approaches, which are algorithms that feature powerful fitting capacity and require no laborious preprocessing on input

data. For example, U-Net of Ronneberger, Fischer, and Brox [1] and its variants were used to segment lung tissues in chest CT scans. However, those models suffer from the fact that, not all the features extracted from the encoder are useful for segmentation task at hand. Thus, attention mechanisms guide the model to emphasize the most salient features, avoiding useless features that are not beneficial.

With the development of deep neural networks, attention mechanism has been widely used in diverse application domains. It is a module added at the top of a convolution layer in order to guide the model to emphasize the most salient features, avoiding useless feature that are not beneficial for the given task. Related algorithms typically refine feature maps in spatial dimension, channel dimension or both. To focus on the most informative features for segmentation using U-Net, researchers proposed to insert attention module in the skip connection of U-net. For example, Oktay et al. [3] proposed attention gate modules (AG) that are incorporated into the skip connections of the encoder-decoder network, using the information of the decoder feature to better guide information to the encoder feature. Hu, Shen, and Sun [2] introduced a Squeeze-and-Excitation module, where global average pooling is performed on input features to produce channel-wise attention which highlight useful channels and that outperforms the methods in ILSVRC 2017 image classification. Roy, Navab, and Wachinger [4] introduced to use both spatial and channel block (scSE), which concurrently recalibrates the feature representations spatially and channel-wise, and then combines them to obtain the final feature representation. Some of the above mentioned methods have been adjusted to segment COVID-19 infections. Zhou, Canu, and Ruan [10] segmented COVID-19 from CT to help diagnostic and patient monitoring while incorporating spatial and channel attention mechanism to U-Net architecture to capture rich contextual relationships for better feature representation. Li et al. [7] 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).

III METHOD

In this section, we will explain in detail the proposed CovSegNet.

3.1 Overview of CovSegNet architecture

The figure 1 shows the general architecture of the network, which is basically the U-net architecture with additional blocks including the attention mechanism module which replaces the classical skip connection. The traditional subsampling blocks (convolutional layers with stride, pooling layers) are replaced by their corresponding multisampling blocks. In figure 1, Multi_conv2 represents the improved version of 2D convolution layer with stride of 2 and a kernel of size k=3; Multisample Maxpooling is the improved version of maxpooling layer; EATT is the attention mechanism block.

3.2 Novel Subsampling technique: Multisampling

Deep learning architectures employ subsampling techniques to scale down spatial dimension lengths of the feature maps in order to increase the global receptive field of neurons in proceeding layers. Some subsampling layers include: pooling layers, convolution layers with stride, dropout, normalization. The spatial resolution r_t of a feature map after passing through a CNN layer is: $r_t = r_{t-1}/k^d$ where r_{t-1} is the resolution of the feature map before the subsampling layer is applied, k is the stride length, and d is the dimensionality of the CNN (1D, 2D or 3D)

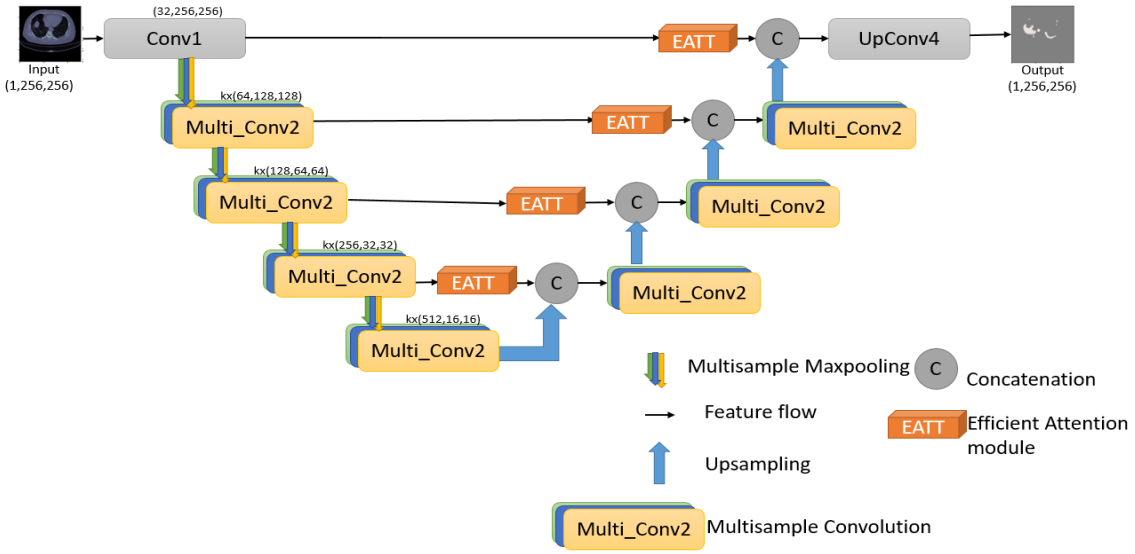

Figure 1: Architecture of the proposed network. A CT slice is taken as input and the COVID-19 infected region is displayed as output

. A two dimensions convolution layer (2D CNN) with a stride length of 2 reduces spatial resolution by a ratio of 4, bottlenecking the capacity of proceeding feature maps. By implementing the following subsampling strategy, we preserve the benefits of traditional subsampling layers while generating a more informative forward pass producing higher-resolution feature maps, better gradient updates for deep layers during training. We increase the number of samples taken from feature maps at subsampling layers and therefore preserve more information for late layers in the network.

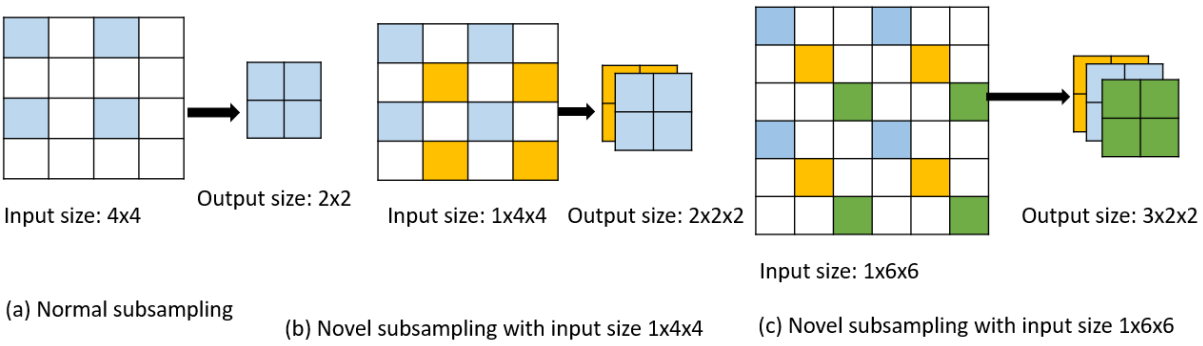

Figure 2: Subsampling strategies

In the above figure 2(a), with the traditional subsampling, the feature map is divided into $k \times k$ sampling windows, then a convolution or pooling operation is lined up with the top left element of the window and the result is a single 2×2 feature window. Our idea is not to limit to lining up the convolution or pooling operation only to the top left element, but we have k^2 possibilities (representing all the elements of the $k \times k$ window) of subsampling. One can decide to choose n samples of the k^2, generalizing the spatial resolution of the feature map r_t to higher dimension: $\mathrm{r}_t = n.\frac{r_{t-1}}{k^d}$.

In figure 2(b), we show an example of 2D subsampling with kernel size of 2x2 where our choice of samples is made on the top left and bottom right element of the window. A 2x2x2 size feature map is then generated with $n = 2$. The multisampling is applied to all the subsampling layers

and the output of the final convolution layer is made of many different submaps giving room to the post processing operations to generate feature vector (3D convolution can be used to combine features across the submap dimension).

3.3 Attention mechanism module

3.3.1 Description

The channel attention block of Hu, Shen, and Sun [2] (Squeeze Excitation block named SENET) is performed through Equation 1 below which shows that there is a dimensionality reduction which can reduce the model complexity, but it destroys the direct correspondence between channels and its weights. Equation 1 first projects channel features into a low-dimensional space and then maps them back, making correspondence between channel and its weight to be indirect.

$$z' = W_1(\delta(W_2 z)) \tag{1}$$

with $W_1 \in R^{C\times C/r}, W_2 \in R^{C/r\times C}$ being weights of two fully-connected layers and the ReLU operator $\delta(.)$. Wang [9] demonstrates that avoiding dimensionality reduction helps to learn effective channel attention.

Inspired by Wang [9], we propose to solve the above problem by avoiding the dimensionality reduction. The idea is to consider the following: Z = $\delta(Wy)$ where W is a matrix with $C \times C$ parameters allowing to model cross-channel (C channels) interaction which is proven to be beneficial to learn channel attention, each row i of the matrix models the relation between the channel i and the remaining **C-1** channels . Proceeding as such requires mass of parameters $(C \times C)$ leading to high model complexity, especially when the number of channel **C** is large. An efficient and effective method to cross-channel interaction consists of modelling the relation between the channel i and its k neighbors, making an abstraction with other channels. This can be done by equation 2 proposed by Wang [9]:

$$w = \delta(C1D_k(y)), \tag{2}$$

where C1D indicates 1D convolution and k is the kernel size of the convolution.

The equation 2 is graphically represented by figure 3 named efficient attention mechanism (EATT) module, which only involves k parameters. The objective of the EATT module is to approximately capture local cross-channel interaction, the coverage of the interaction k can be determined automatically. A high value of k also increases the model complexity. After aggregating convolution features using global average pooling without dimensionality reduction, EATT module first performs 1D convolution then followed by a Sigmoid function to learn channel attention. Our EATT module is applied in U-net by replacing the skip connection by the EATT module as it is done by Hu, Shen, and Sun [2]. The architecture of our proposed attention module is shown in Figure 3.

IV EXPERIMENTATIONS

4.1 Dataset

We used a publicly available COVID-19 CT scans dataset published by Jun et al. [5], which contains 20 labeled COVID-19 CT lung scans volume (with 3520 slices of 630x630 gray scale

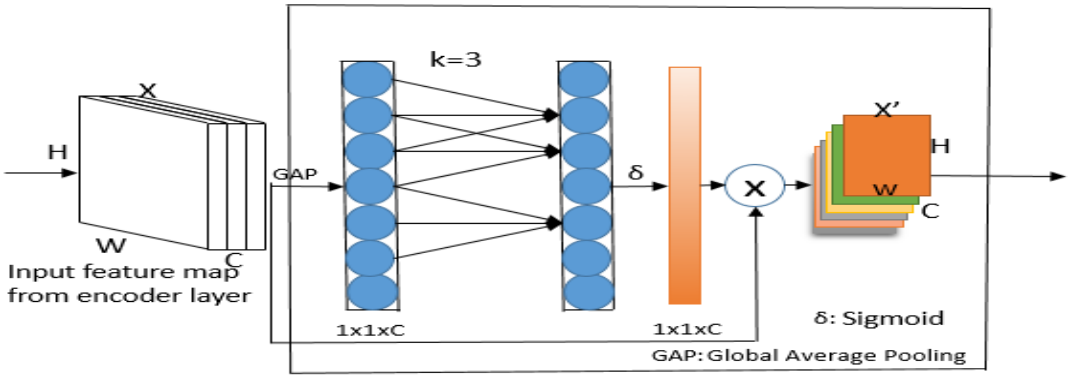

Figure 3: Efficient attention mechanism

Table 1: Quantitative analysis of infection regions on our dataset. scAG-U-net refers to the model proposed in Khanh, Dao, Ho, Yang, Baek, Lee, Kim, and Yoo [6] and scSE-U-net is the model proposed in Roy, Navab, and Wachinger [4]

Model	Dice	Recall	Precision
scAG-U-net	90.67 ± 0.06	96.54 ± 0.022	94.86 ± 0.05
scSE-U-net	90.55 ± 0.07	96.83 ± 0.02	94.97 ± 0.05
CoviSegNet	90.57 ± 0.06	96.03 ± 0.02	95.02 ± 0.04
CoviSegNet + Subsampling	**90.73 ± 0.06**	**96.91± 0.02**	**95.90± 0.04**

image) of patients diagnosed with COVID-19 as well as segmentations of lungs and infections made by experts. Left lung, right lung, and infections are labeled by two radiologists and verified by an experienced radiologist. Each slice has it corresponding infection mask.

4.2 Implementation details

In the medical community, the Dice Score Coefficient (DSC) 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. The Focal Tversky Loss function (FTL) addresses the limitation of Dice loss (DL) that penalizes false positive (FP) and false negative (FN) equally, which results in segmentation maps with high precision but low recall especially true and disadvantageous in the case of COVID-19 lesions which have small regions of interests. To this end, we used FTL to train the network to help segment the small COVID-19 regions.

The models are optimized with the Adam optimizer with learning rate of 5e-4 and batch size of 8. The models are trained until they cannot achieve further improvement and we use Early stopping strategy of size 15 to avoid model overfitting. The evaluation metrics used to validate the effectiveness of the proposed method are Dice similarity coefficient, Precision and Recall. The execution is been done on a GPU Tesla P100-PCIE machine, with a memory size of 16GB. The number of output channel is respectively 32, 64, 128, 256, and 512 at each level of the downsampling phase.

Table 2: Comparison of different attention methods on Covid-19 in terms of network parameters (Param.), floating point operations per second (FLOPs), training speed (per second)

Model	Time	#Param.	#FLOPS
scAG-U-net	56.47	8.65M	16.47 G
scSE-U-net	47.48	8.98M	16.44 G
CoviSegNet	**42.49**	8.63M	16.43 G
CoviSegNet + Subsampling	43.82	**8.63M**	**16.42 G**

V EXPERIMENT RESULTS

We compared the performance of the proposed network with two publicly available models namely spatial - channel attention gate U-Net (scAG-U-net) Khanh, Dao, Ho, Yang, Baek, Lee, Kim, and Yoo [6] and spatial-channel squeeze-excitation U-net (scSE-U-net) Roy, Navab, and Wachinger [4] executed on the same computing platform. The evaluation metrics include both efficiency (i.e., network parameters, floating point operations per second and training speed) and effectiveness(Dice loss, recall and precision). Table 1 presents the evaluation metrics of different algorithms (our proposal and some state of art algorithms). The results presented in Table 2 show that the computer resources needed to execute CoviSegNet is lower than the the state-of-the-art counterpart.

5.1 Quantitative analysis

Detailed comparison among different models in our experiments is shown in Table 1. Our proposed model outperforms scSE-U-net and scAG-U-net in terms of training time, Dice, recall and precision metrics. As these models are identical in model encoder, it appears that the proposed attention mechanism contributes very well in lesion segmentation. The utilization of attention mechanism helps to accurately detect and segment infected tissues and ignore non infected tissues. The number of false positive is reduced, increasing the rate of the recall. The CovSegNet captures both the large and tiny visual structures which is useful to segment infection lesion with different size. It is noted that the number of parameters of our proposed model is lower than that of others attention U-net models presented in this work. More to that the computation cost is considerably reduced due to lower number of operations. The CovSegNet executed without the proposed subsampling techniques is fastest in terms of training time because of it low number of operations, but it is less accurate comparing to CovSegNet+Novel subsampling, which shows the efficiency of the novel subsampling techniques used in efficient features extractor.

5.2 Qualitative analysis

From the visualization of our results in Figure 5, we notice that our model outperforms other models. The models scSE-U-net and scAG-U-net provide good results but there are limited in segmenting lesions with non-smooth borders. Our proposed method is sensitive to non-smooth and tiny lesions as shown in Figure 4. Additionally, others methods tend to provide some segments which don't exist in the ground truth image.

VI CONCLUSION

We have presented in this paper, a novel segmentation network CovSegNet to efficiently segment COVID-19 lesions. The multisampling technique and the EATT module can easily be integrated into the popular U-Net architectures while minimizing the computational cost and increasing the performance of the model. By taking advantage of the information of the encoder and decoder features, our proposed attention mechanism captures efficient information about the location and the shape of the lesion. Experiments results show that the proposed method improves the segmentation results while reducing the model parameters and computation cost. We limited our study to the channel attention mechanism while expecting further works to be carried on the spatial attention mechanism that will allow to focus on the salient region that are beneficial to find the location of the object and determine the target structure of the of the object.

References

[1] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Edited by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi. Cham: Springer International Publishing, 2015, pages 234–241.

[2] J. Hu, L. Shen, and G. Sun. "Squeeze-and-Excitation Networks". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pages 7132–7141.

[3] O. Oktay, J. Schlemper, L. L. Folgoc, M. J. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. "Attention U-Net: Learning Where to Look for the Pancreas". In: *ArXiv* abs/1804.03999 (Apr. 2018).

[4] A. G. Roy, N. Navab, and C. Wachinger. "Concurrent Spatial and Channel 'Squeeze & Excitation' in Fully Convolutional Networks". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Edited by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Cham: Springer International Publishing, 2018, pages 421–429.

[5] M. Jun, W. Yixin, A. Xingle, G. Cheng, Y. Ziqi, C. Jianan, Z. Qiongjie, D. Guoqiang, H. Jian, H. Zhiqiang, N. Ziwei, and Y. Xiaoping. "Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation". In: *arXiv preprint arXiv:2004.12537* (2020).

[6] T. L. B. Khanh, D.-P. Dao, N.-H. Ho, H.-J. Yang, E.-T. Baek, G. Lee, S.-H. Kim, and S. B. Yoo. "Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging". In: *Applied Sciences* 10.17 (2020).

[7] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, and J. Xia. "Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy". In: *Radiology* 296.2 (2020), E65–E71.

[8] T. Liang. *Handbook of covid-19 prevention and treatment*. Zhejiang: Zhejiang University School of Medicine, 2020.

[9] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu. *ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks*. 2020. arXiv: 1910.03151 [cs.CV].

[10] T. Zhou, S. Canu, and S. Ruan. "Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism". In: *International Journal of Imaging Systems and Technology* 31.1 (Nov. 2020), pages 16–27.

A ANNEX 1

See Figure 4.

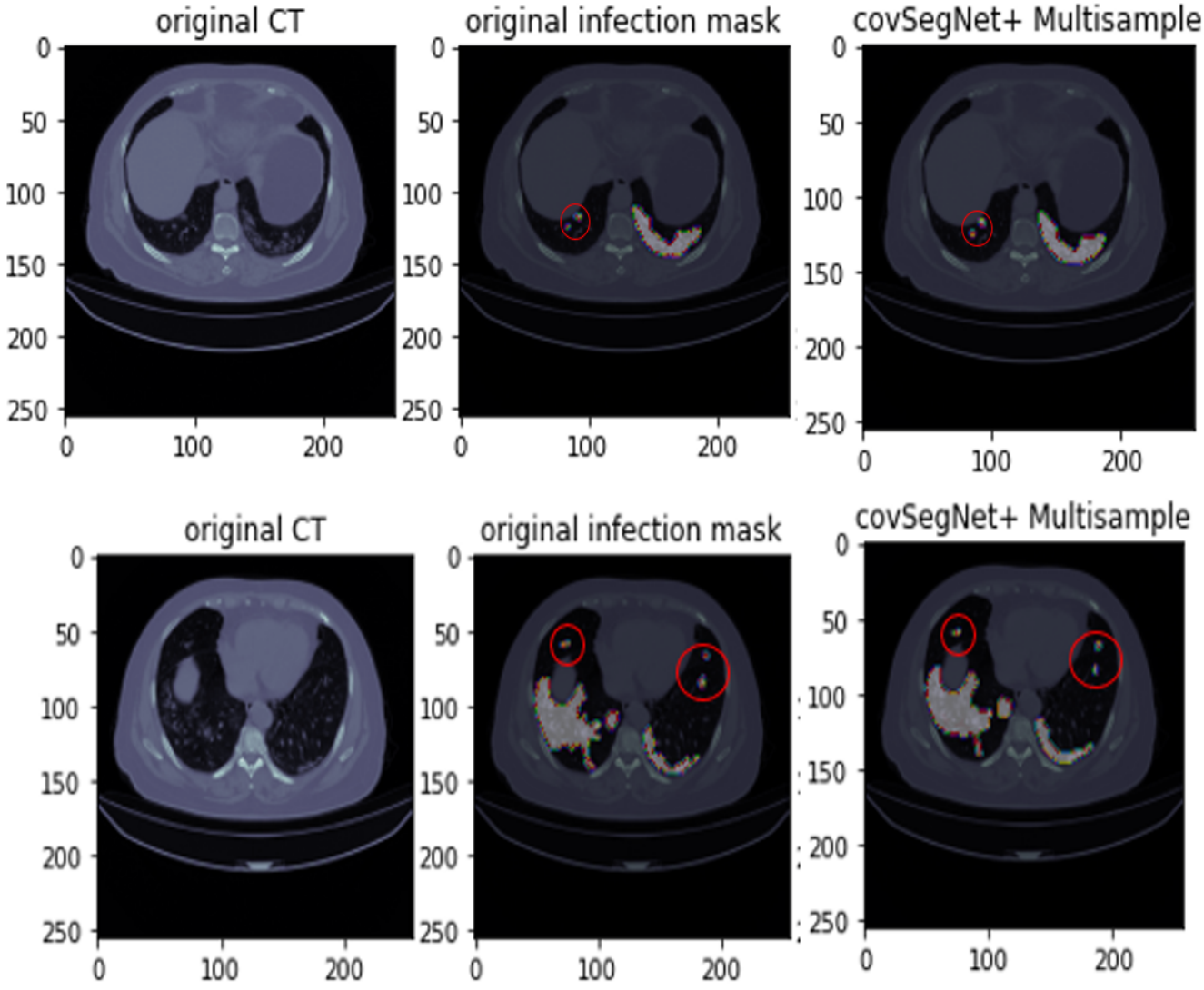

Figure 4: Detection of tiny lesions

B ANNEX 2

See Figure 5.

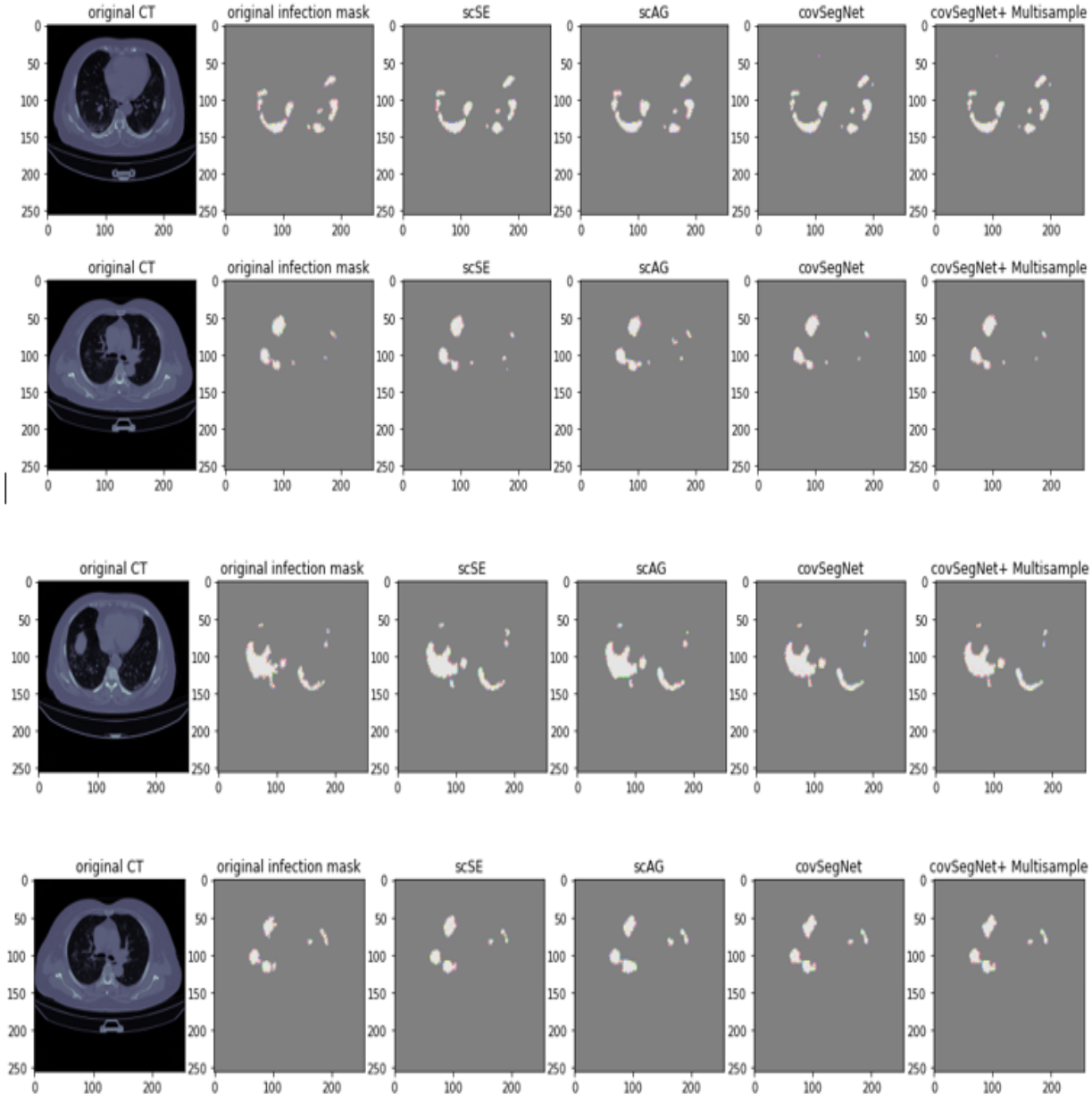

Figure 5: Representative results from our dataset using different U-Net architectures.