Image Segmentation for COVID-19

Aytekin YILDIZHAN
Computer Engineering Department, Hacettepe University
CMP 717 Final Report, N18147923
aytekinyildizhan@hacettepe.edu.tr

Abstract

Since december 2019, the entire world is facing a situation of huge pandemic crisis due to a newly emerged coronavirus-type disease (COVID-19). Although no vaccine of COVID-19 is available yet, RT-PCR tests take an important place in the diagnosis of COVID-19. Since it suffers from high false negative rates and time consuming, computed tomography (CT) is the one of the influential technique to diagnose COVID-19. But it is hard for the doctors (especially radiologists) to maintain an acceptable level of attention when examining the CT images. Having a computer-aided software from CT images will both save time and help doctors to diagnose. In this paper, we analyze the COVID-19 CT segmentation dataset, determine the infection regions of dataset from the lungs and try to adjust Inf-Net to more generalized method. We use Inf-Net for the semantic segmentation model and we try to improve its baseline. The structure of Inf-Net is created according to Res2Net, and we expand its capabilities. We add these networks which are iResNet50, ResNest50, ResNet50 and ResNext50 to Inf-Net. These ResNet derivatives are easily integrated and when the next network is VGG16, we change Inf-Net's baseline. Finally, we show the quantitative and qualitative results and discussion about future work.

1. Introduction

COVID-19 is a new infectious disease that has affected 9,947,624 individuals all over the world and caused 497,825 deaths, as of June 27 in 2020. This disease can lead to severe pneumonia that can causes death. Also, COVID-19 is highly contagious so it is very important to diagnose COVID-19 in order to limit the spread of the disease. The main tests are based on reverse transcription polymerase chain reaction (RT-PCR). However, this test is a very time-consuming [1]. It takes 4-6 hours to obtain results, which is a long time compared with the rapid spreading rate of COVID-19 [2, 14].

An alternative diagnostic method is that chest radiography imaging (e.g., computed tomography (CT) imaging). There have been several works studying the effectiveness of CT scans in screening and testing COVID-19. However, there are various challenges of this studies. These are different image CT formats (NIfTI, DICOM, jpg, png etc.), different levels of resolution, different image sizes in pixels, with/without annotations at different places in the lungs. And also examining CT slices is really fatiguing process for radiologists and after twenty minutes of continuous work, the operator's attention degrades [3].

Semantic segmentation is one of the fundamental problem of image processing and has been studied a lot in literature. Medical semantic segmentation, which is a subbranch of semantic segmentation, is a vital important field of research in terms of determining where the disease is in the organs. In this paper, we analyze COVID-19 segmentation databases and try to increase Inf-Net's structure ability. In Inf-Net [4], Res2Net is used as backbone and their model is specialized by Res2Net structure. Thus, we aim to increase the abilities of their models to become a more general model. We use COVID-19 CT segmentation dataset [5] with iResNet50 [6], ResNest50 [7], ResNet50 and ResNext50 [8]. Finally we change the structure of Inf-Net and we implement VGG16 [9].

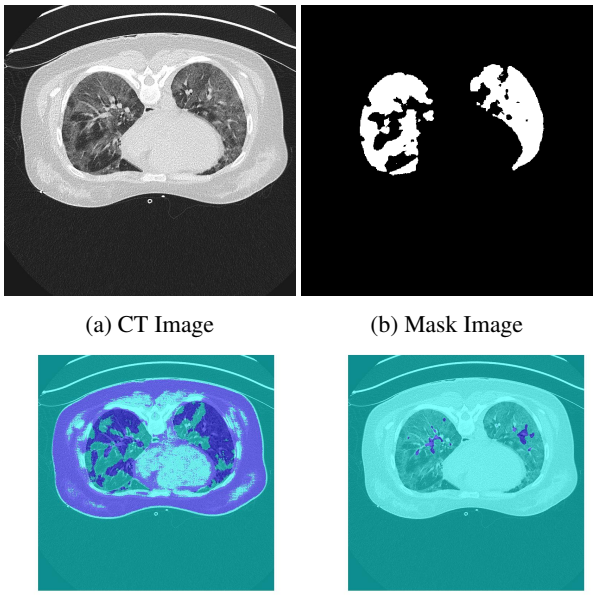
Our paper is constructed as follows: (2) report on related work that has been done in the past, (3) proposed method for CT images (Inf-Net) and (4) results from our experiments with our method, and (5) conclusion.

2. Related Work

Wang et al. create COVID-Net which detects COVID-19 cases from chest X-ray (CXR) images [1]. COVID-Net is one of the first open source network designs for COVID-19. They use the residual projection-expansion-projection-extension (PEPX) module. It consists of conv + conv + Depth-wise Representation + conv + conv. In depth-wise representation, it learns the spatial characteristics of lungs to minimize complexity while it preserves the representational capacity. They use COVID-19 and non-COVID-19 CXR images for training and their accuracy value is better

than VGG19 and ResNet50. Zhao et al.[2] create a COVID-CT dataset and train a deep CNN on this dataset. They use different approaches to get lung masks and lesion masks (COVID-19 region). They use two-streams, classification stream and segmentation stream. In classification stream, for lung mask they use residual U-Net. When lung masks are created, they concatenate the CT image with the corresponding lung mask. In segmentation stream, concatenated images are fed into the diagnosis model to feature extraction using Atrous Spatial Pyramid Pooling (ASPP). During training, the network predicts whether this image is positive or negative for COVID-19 and the lesion mask. Finally, they get the COVID-19 region. Thier precision values are high but the recall values are not satisfactory. Yan et al. [10] try to develop a new deep CNN which is called COVID-SegNet. It is designed for segmenting the CT images with COVID-19 infections. COVID-SegNet consists of encoder and decoder. Encoder is used for feature extraction and decoder is used for monitoring the segmentation. They increase the ability of encoder by adding Feature Variation (FV) block and Progressive Atrous Spatial Pyramid Pooling (PASPP) block to better extracting the features. Thier dice, precision and recall values are better than FCN, UNet, VNet, and UNet++. Zheng et al. [11] proposed a weakly-supervised deep CNN. They develop a 3D deep learning framework and use ResNet50. Their new CNN is called De-CoVNet and it is detection network not segmentation network. In the future, their network can be used for segmentation. Hofmanninger et al. [12] compare 4 generic deep learning approaches and develop their own network which is called U-net(R231). Also they achieve the second highest score among all submissions to the LOLA11 challenge. Ma et al. [13] create a a well-labelled COVID-19 3D CT dataset and they concentrate the few-shot learning, domain generalization, and knowledge transfer. They use 3D U-Net as network and they compare different datasets. Voulodimos et al. [14] study the efficacy of U-Nets and FCN that require a limited annotated dataset for training and testing. Qiu et al. [15] present MiniSeg which is a lightweight CNN for COVID-19 segmentation. It has just 472K parameters so it is not overfit easily. Also they design Attentive Hierarchical Spatial Pyramid (AHSP) for extracting multi-scale semantic features of CT images. It consists of four dilated DSConvs (depthwise separable convolution) and thus it reduces the parameters. Then, MiniSeg has encoder-decoder mechanism and it uses several AHSP module except first stage is vanilla Convolution Block for encoder part. When it finishes encoder stage, they use Feature Fusion Module (FFM) to fuse the learned feature map at each scale of the encoder stage. Also Parametric ReLU (PReLU) is used instead of ReLU. It provides state-of-the-art method with high efficiency and limited COVID-19 training data.

According to literature, COVID-19 segmentation problem is new and also challenging problem. We have COVID-19 datasets that contain a small amount of data so it is really tough job for increasing the number of CT images in a short time. Thus it is very important to achieve success with small amount of data for segmentation of COVID-19 cases from CT images. In this paper, Inf-Net [4] is chosen as a baseline and we try to increase its capability.



(a) CT Image (b) Mask Image

(c) DeepLab V3+ (d) Seg-Net

Figure 1: Examples of Wrong Segmented CTs

3. Method

Due to the privacy, researchers and hospitals do not want to share the information about CT images. Thus it is hard to obtain CT images. Several dataset has been purposed for open access COVID-19 such as Mosmed-1110, MedSeg-29, Radiopaedia etc. Also several works create their own dataset [1, 2, 13, 14]. In this paper, COVID-19 CT segmentation dataset and obtained from in this link [5]. The dataset contains two individual dataset. We use the second one. In the second dataset, there are 9 NIfTI training files, COVID masks and lung mask. We use training and COVID mask images. When we convert the NIfTI images to jpg (with MATLAB), we have 829 CT images with size 630 x 630 x 3 and a total size 202 Mb. We selected 700 images as training set and 100 images as test set in our previous work (progress report). However in this dataset, there are both positive and negative sample of COVID-19 so we could not get proper results in Figure 1.

Thus in this paper, we choose the baseline network which is called Inf-Net [4]. Inf-Net structure is shown in

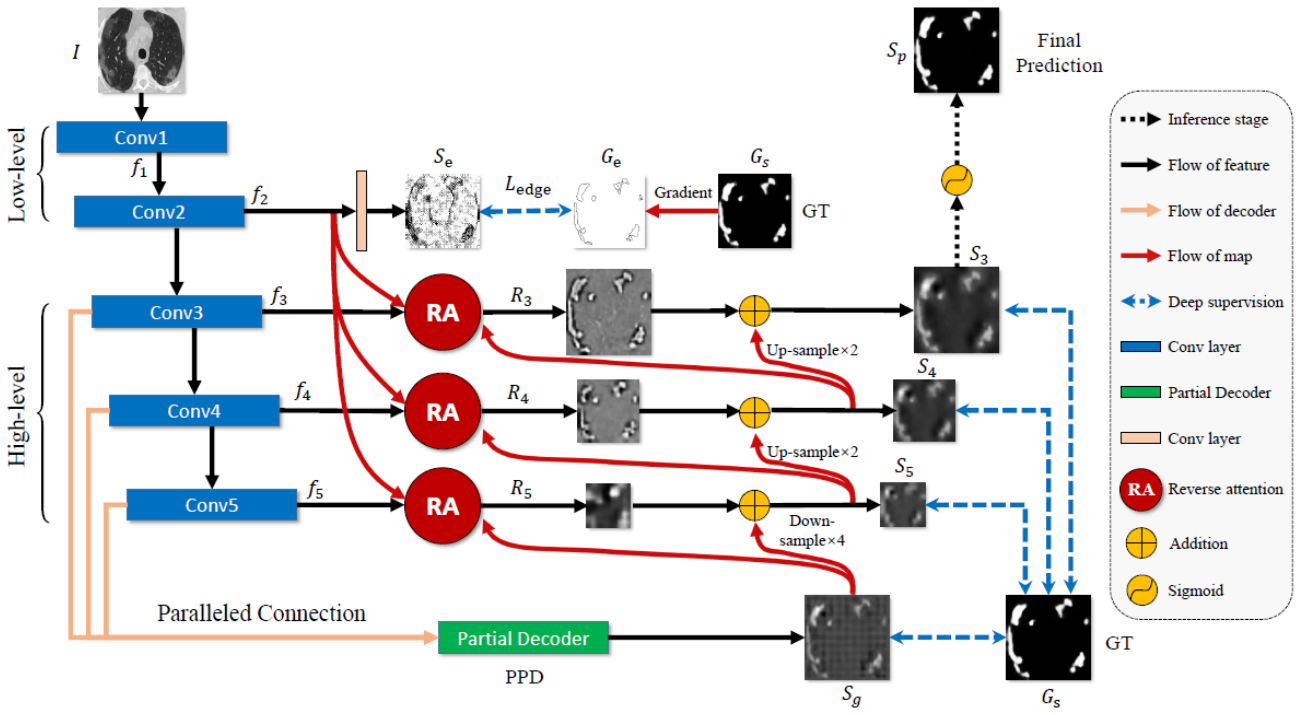


Figure 2: Inf-Net

Figure 2.

3.1. Overview of Inf-Net

Inf-Net is created on the base of Res2Net. The first two convolution layer represents as low-level features and other three convulation layer represents high-level features using the first five convolution layers of Res2Net. Inf-Net's parts are described in detail below.

3.1.1 Egde Attention Mechanism

In Figure 1, f_2 is the low level features. Then it is fed into one convolution layer and the predicted edge feature map is produced, S_e. G_e, is the edge ground-truth map, is calculated using the ground-truth map G_s. Then our system first loss is calculated by using Binary Cross Entropy (S_e, G_e) and L_{edge} is calculated. By using L_{edge}, Inf-Net is better extract the edge properties of CT images.

3.1.2 Parallel Partial Decoder

Parallel Partial Decoder (PPD) uses high-level features of CT images since low-level features consume more resources [16]. PPD takes f_3, f_4, f_5 and these are fed into several conv and BN layer then it produces a coarse global map S_g. PPD works like a coarse extractor. Figure 3 shows the PPD structure.

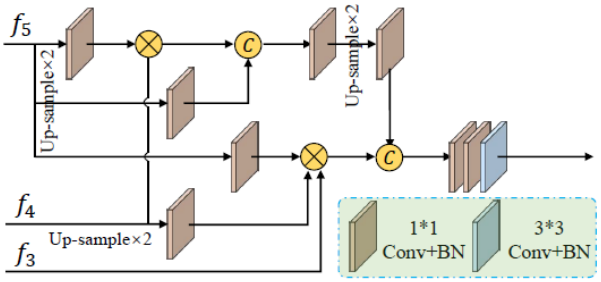


Figure 3: Parallel Partical Decoder

3.1.3 Reverse Attention Mechanism

There are three Reverse Attention(RA) blocks in Inf-Net. It takes three inputs. These are high-level features (f_3, f_4, f_5), EA features (f_2) and down/upsample of coarse global map (S_g). RA works like a fine extractor. . RA rectifies the coarse estimation into more accurate prediction map. It estimates infection COVID-19 regions. Figure 4 shows the RA mechanism

3.1.4 Calculation Losses

In Figure 2, the loss function is L_{seg}. It is equals to summation of the weighted IoU (L_{IoU}) and BCE (Binary

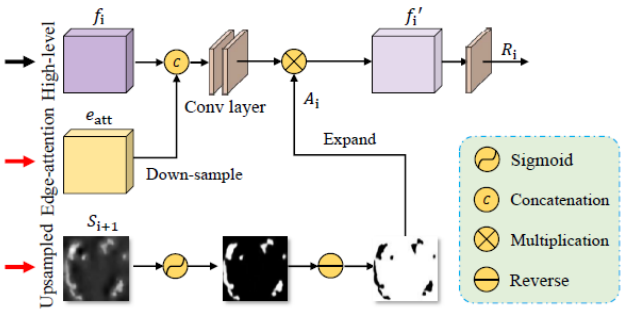


Figure 4: Reverse Attention Mechanism

Cross Entropy). Firstly we have to calculate $L_{seg}(S_g - G_s), L_{seg}(S_3 - G_s), L_{seg}(S_4 - G_s), L_{seg}(S_5 - G_s)$. Also all S_3, S_4, S_5 are up-sampled to obtain the same size of S_g. Total loss is equal to all summation of L_{seg} and L_{edge}.

4. Experimental Results

In this paper, Inf-Net is selected as baseline model. 50 CT images are selected as a training set and 50 CT images are selected as a testing set from positive COVID-19 sample [5]. All size of images are set by 352 x 352. Our work is written with Pytorch v1.5.1 and CUDA v10.1. Res2Net is used for backbone in original work and we enlarge this system with different backbones such as iResNet50 [6], ResNest50 [7], ResNet50 and ResNext50 [8]. For VGG16[9], we use just one convolution layer for low-level features because in Figure 2, first convolution layer of Res2Net is not used for any segmentation process. As a result, in VGG16 one convolution layer is used for low-level feature and three convolution layers are used for high-level features using first four convolution blocks of VGG16. In Figure 2, there is no explanation of f_1 and the feature of f_1 is not used. Our hyperparameters are shown in Table 1.

TYPE OF PARAMETER	VALUE
Mini batch size	2
Learning Rate	0.0001
Training Batch Size	352 x 352
Max. Epochs	100
Optimizer	Adam
Decay Learning Rate	50

Table 1: Hyperparameters of The Model

4.1. Quantitative Results

We use six different metrics such as Dice similarity coefficient (Dice), Sensitivity (Sen.), Specificity (Spec.), Structure Measure (Str. M.), Enhanced-alignment Measure (EAM) and Mean Absolute Error (MAE). Dice similarity coefficient measures is the similarity metric between COVID-19 area of ground truth and COVID-19 area of the prediction score. Sensitivity means the number of correctly identified positives with respect to the number of positives. Specificity means the number of correctly identified negatives with respect to the number of negatives. Structure Measure measure the structural similarity between a prediction map and ground truth of CT image [4]. Enhanced-alignment Measure is proposed a metric for evaluating both local and global similarity between prediction map and ground truth of CT image [4]. And final metric MAE measures the pixelwise error between prediction map and ground truth of CT image. These metrics are written with MATLAB 2019b. Res2Net is a backbone method. The quantitative metric results are shown in Table 2. Highest netric value is colored red.

	Dice	Sen.	Spec.	Str. M.	EAM	MAE
Res2Net	0.713	0.716	0.952	0.781	0.867	0.073
iRes2Net	0.690	0.658	0.963	0.781	0.848	0.073
ResNest	0.708	0.701	0.958	0.795	0.868	0.072
ResNet	0.702	0.678	0.961	0.785	0.858	0.071
ResNext	0.711	0.712	0.954	0.798	0.866	0.072
VGG16	0.733	0.699	0.966	0.782	0.893	0.063

Table 2: Quantitative Metric Results

As you can see from Table 2, VGG16 has the highest four metric values (Dice, Spec., EAM and MAE).

4.2. Qualitative Results

Figure 5 shows the three example CT image and their masks. Figure 6 shows the visual comparison results.

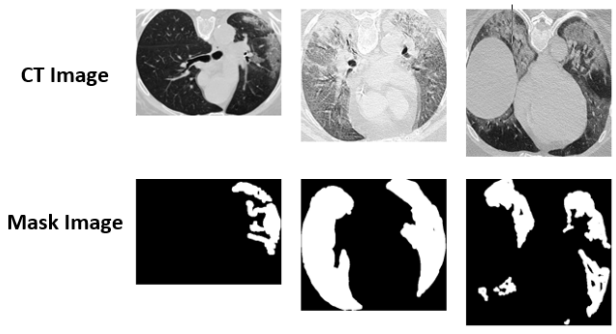


Figure 5: Example CT Images and Their Masks

As you can see in Figure 6, iResNet50, ResNest50, ResNet50, ResNext50 and VGG16 are applied succesfully. There is no big difference in terms of the human vision system.

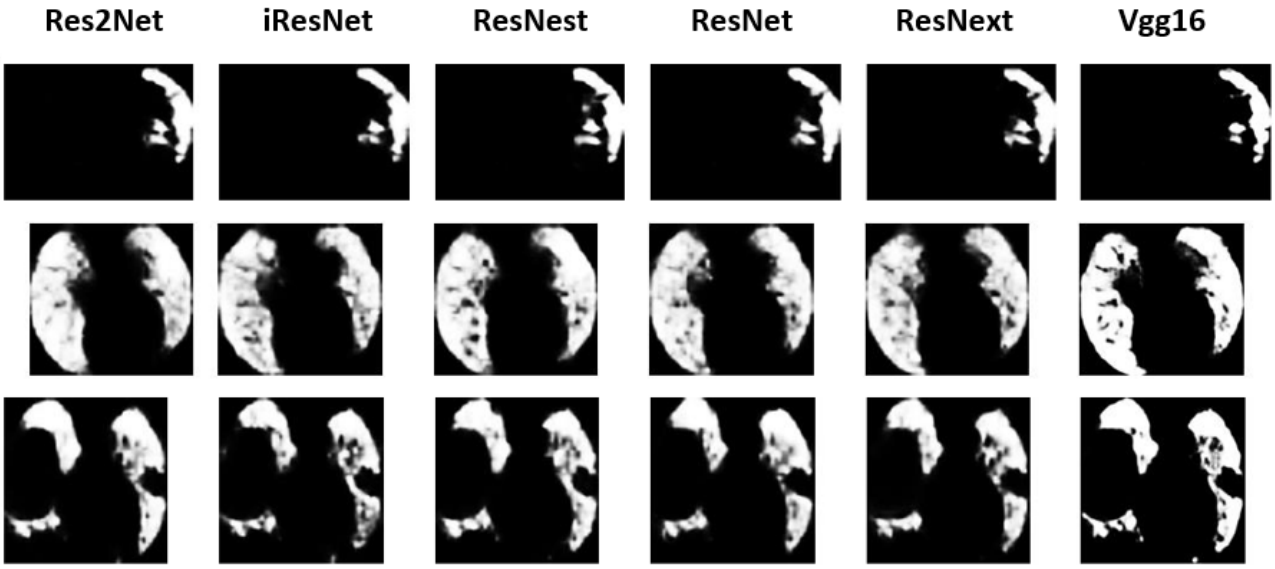


Figure 6: Visual Comparison of Lung Segmentation

5. Conclusion

In this paper, we try to improve the capability of Inf-Net. Inf-Net is based on the block of Res2Net. We enlarge it with five different backbones. It is shown that Inf-Net supports many backbone models which are iResNet50, ResNest50, ResNet50, ResNext50 and VGGNet16. While very first five convolutional layer is used for Res2Net, we use very first four convolutional layer of VGG16 since first convolutional layer is not used for segmentation in Figure 2. With this experiment, VGG16 has the highest metric value among them. The other ResNet based backbones may be applied with very first four convolutional layer. Also in Figure 6, five different backbones are implemented successfully (Res2Net is backbone). If the number of layer of backbone is increased, the metric value will increase too (Res2Net50 $-$ >Res2Net100 etc.). This system may be support lightweight architecture and faster inference, like MobileNet, SqueezeNet by increasing/decreasing the number of RA (reverse attention module).

COVID-19 disease are increasing day by day. There is no vaccine yet. So treatment is applied according to the segmentation (infected) area. As a computer scientists, we have to help healthcare professionals somehow. The purpose of such studies is to help the radiologist to detect COVID-19 regions from CT images rather than more citing by another works. Therefore, more COVID-19 CT images are collected, converted to the same type (DICOM, NIfTI etc.) and tested. If these are done, the work with the segmentation of COVID-19 will increase.

References

[1] Wang L., Lin, Z.Q. and Wong, A. COVID-Net: "A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images". arXiv preprint, May 11, 2020, arXiv:2003.09871v4.

[2] Zhao, J., Zhang, Y., He, X. and Xie, P. "COVID-CT-Dataset: A CT Scan Dataset about COVID-19". arXiv preprint, March 30, 2020, arXiv:2003.13865v1

[3] Haering, N., Venetianer, P.L., Lipton, A. "The evolution of video surveillance: An overview". Mach. Vis. Appl. 2008, 19, 279–290.

[4] Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L., "Inf-net: Automatic covid-19 lung infection segmentation from ct scans." arXiv, preprint arXiv:2004.14133, 2020.

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

[6] Duta, I.C., Liu, L., Zhu, F., and Shao, L., "Improved Residual Networks for Image and Video Recognition." ArXiv, abs/2004.04989, 2020

[7] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Zhang, Z., Lin, H., Sun, Y., He, T., Mueller, J., Manmatha, R., Li, M., and Smola, A.J., "ResNeSt: Split-Attention Networks." arXiv, abs/2004.08955, 2020.

[8] Xie, S., Girshick, R., Dollár, P., Tu, Z. and He, K., "ResNeXt: Aggregated Residual Transformations

for Deep Neural Networks", arXiv, abs/2004.08955, 2016.

[9] https://download.pytorch.org/models/vgg16-397923af.pth

[10] Yan, Q., Wang, B., Gong, D., Luo, C., Zhao, W., Shen, J., Shi, Q., Jin, S., Zhang, L. and You, Z. "COVID-19 Chest CT Image Segmentation – A Deep Convolutional Neural Network Solution", arXiv preprint, Apr 26, 2020, arXiv:2004.10987v2.

[11] Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W. and Wang, X., "Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label", medRxiv preprint, March 26, 2020, doi: https://doi.org/10.1101/2020.03.12.20027185

[12] Hofmanninger, J., Prayer, F., Pan, J., Rohrich, S., Prosch, H., and Langs, G. "Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem". arXiv preprint, Jan 31, 2020, arXiv:2001.11767v1

[13] Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z., Nie, Z., and Yang, X., "Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation". ArXiv, abs/2004.12537.

[14] Voulodimos, A., Protopapadakis, E., Katsamenis, I., Doulamis, A. and Doulamis, N., "Deep learning models for COVID-19 infected area segmentation in CT images", Cold Spring Harbor Laboratory Press, doi: 10.1101/2020.05.08.20094664, 2020

[15] Qiu, Y., Liu, Y. and Xu, J., "MiniSeg: An Extremely Minimum Network for Efficient COVID-19 Segmentation". ArXiv, abs/2004.09750, 2020

[16] Wu, Z., Su, L. and Q. Huang, "Cascaded partial decoder for fast and accurate salient object detection," CVPR, 2019, pp. 3907–3916.