International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 9, Issue 9 [September. 2020] PP: 01-07

Instance Segmentation of Lung Infection of Coronavirus in CT scan Lungs

D P Gaikwad[1], Vibhav Joshi[2]
Department of Computer Engineering
AISSSMS College of Engineering, Pune, India

ABSTRACT:*The Covid CT scans are used for treatment of Covid patients and calculating the severity of infection in Lungs. Radiologist use these scans for research as well as for treatment. Knowing the spread of infection in areas of lungs is very important for radiological study. In this paper, we investigate issues involving Mask-RCNN for Instance Segmentation of Lung Infection, Right Lung, Left Lung. We propose an effective instance segmentation model for lung Infection detection on CT scan by using a Mask-RCNN model. We train and evaluate our model on slices of CT scan which are labelled by radiologist. The Dataset has mask of right and left lung and lung infection. With different backbone architectures we acquired a good accuracy on runtime. Prospectively, the proposed methods could provide support to radiology practice in terms of quantitative analysis of lungs, and it could potentially lead to a better understanding of spread of infection of Covid in different parts of lungs.*

Date of Submıssıon: 05-12-2020 Date of Acceptance: 20-12-2020

I. INTRODUCTION

For testing of Covid patients, RT-PCR test have been used but they have a high false negative rate. The spread of infection in lungs has similar features in most of the patients. The RT-PCR take a long time for results, most of the patients had decreased lymphocyte count and high-sensitivity c-reactive protein level. CT-Scan are fast and give fast 3-D model of lungs. Thus, using this model and extracting and processing images from this model and training the Mask-RCNN to predict the location of infection in the lungs of patients. Initial CT-findings in **Covid** cases include bilateral, a peripheral or posterior distribution, mainly in the lower lobes and less frequently in the middle lobe. The main aim of Deep learning model is to predict the ground glass opacification (GGO) which mostly found in the patients of Covid-19. Radiologist can delineate this area by which they can label these images and lung infection. To learn from these features and thus predicting the infection is main aim of our Mask-RCNN model. Mask RCNN uses RPN regional proposal network where first regions are identified where the objects could be. After that we get sliding windows with threshold for probability of the object present. Then on these regions, we perform basic feature extraction with Convolutional Neural Network. At the end bounding box repressors, prediction class, and mask is given out by the Mask RCNN algorithm with the mask of infection, right lung, left lung.

We can localize the infection and detect the features that were labelled by radiologist. Features such as ground glass density Fibrotic bands, Dilated vessels in affected area. The CT scan are in NIFTY format which represents the 3D model of lungs as shown in Figure 1. To use this the radiologist have sliced the model and created mask on these images. We use these and train our model which gives output the mask and label instance of that mask. This mask can be used to infer different results and determine features such as ground class opacification, consolidations and pleural effusions.

Mask RCNN uses FPN Feature Pyramid Network for up sampling and down sampling of images. The network is connected laterally which helps in accuracy of feature extraction. The FPN architecture is used in determining the regions of interest. The data needs to be converted to COCO format and fed into the model. Mask RCNN offers different backbone architectures with different use cases. We compare results of these backbone architectures and determine which architecture is suitable for instance segmentations of images of Covid patients' lungs.

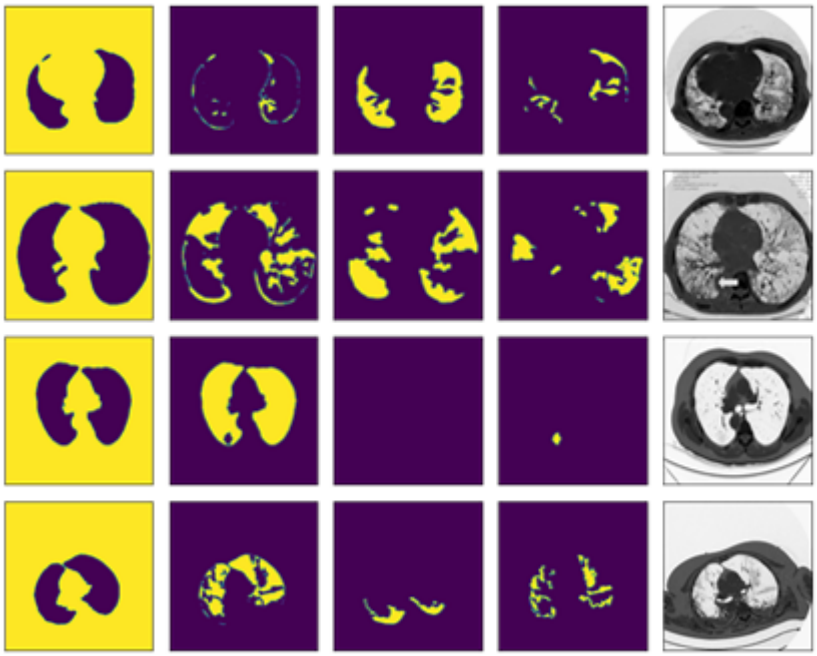

Fig. 1. Segmentation mask of Right Lung, Left Lung, Lung Infection of Covid CT Scan of Lungs

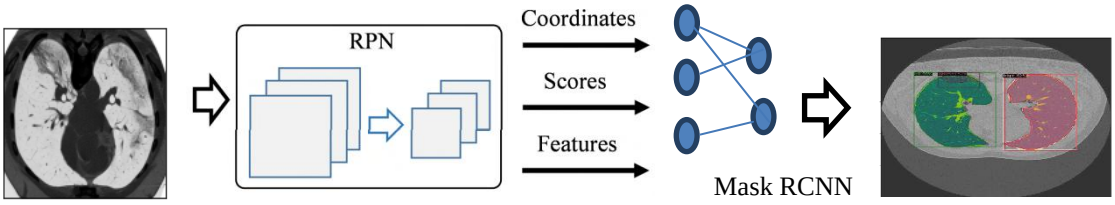

Fig. 2. Our urinary sediment detection pipeline.

II. LITERATURE SURVEY

In [1] Inf-Net is used on CT slices with parallel partial decoder for feature extraction. It uses semi-supervised segmentation for segmentation of images. It focuses on location-oriented model for first determining location of infection. Semi-supervised approach helps the model to randomly select propagation of feature which increases accuracy.3D U-Net architecture is used in [2] for training the model on CT 20 scans. In this method, up sampling is done with decoder and then at the end SoftMax is applied. Thus, at each level of 3D Scan a 2D image is down sampled and features are extracted, in decoder phase these features maps are up sample to original size. Spatial information is lost in this approach which is achieved by architecture where we use Feature Pyramid Network with lateral connections. In [3] Harmony search optimization with thresholding is used to detect severity of lung infection Threshold filter is applied to extract only those regions of lungs which may have highest probability of infection. Then these regions are enhanced by search optimization technique given in paper. Feature extraction is done on this area and these feature maps are then enhanced further. Deep learning approach is applied [4] for automatic segmentation of infection regions on chest CT. To enhance this delineation human-in-the-loop strategy is used. VB-Net a 3-D Convolutional Neural network is the backbone architecture which uses the same up-sampling and down sampling techniques for feature extraction.

[5] uses DECoVNet a 3D training and testing the CT scans. First 2D UNet form a mask volume after which these masks are resampled for spatial resolution. The output of UNet with lung masks is given to DeCoVNet with ground truth labels. A weakly supervised algorithm is used to minimize the efforts of radiologist in annotating the images. [6] They have used U-Net with encoders to amplify feature representation and increase the spatial relationship. In encoder feature representation at each layer are input into an attention mechanism. Finally, decoder obtains the segmentation result. [7] The U-Net architecture is used enabling to obtain segmentation mask in same resolution as images. Region of Interest can be obtained by another CNN combining it with Attention Gates to eliminate the preselecting network. Adversarial scheme is used which is called Structure Correcting Adversarial Network. A pretrained Fully Connected Network is used as a General Adversarial Network. Same structure is followed in [8] where U-Net is used for segmentation.

For data preprocessing and creating segmentation pipeline for data [9] uses MIScnn with data augmentation and training model. With 3d U-Net they train model for segmentation on data of NIFTY format. [10] uses DRE-Net for training model which is pre-trained ResNet on which Feature Pyramid Network is used to get most important features[11] nnU-Net is used for manual training for biomedical image segmentation. Similarly, in [12] U-Net is used with attention mechanism. In this all features are not obtained by encoder. A scSe based attention mechanism in which bot encoder and decoder share feature information.

III. RESIDUAL NETWORK AND RPN NETWORK

Traditionally, to increase the accuracy more and more layers were added to the neural network. The problem raised of vanishing gradients. The deeper networks started converging and accuracy started decreasing. This was not caused by overfitting but due to stacking the layers one above the other. To overcome this problem in [13] deep residual learning framework was introduced. In this, a skip connection or shortcut connection was used to learn the identity mapping. The residual mapping H(x) -x was introduced as skip connection. Thus, the original output F(x) + x was propagated further in feedforward neural network. The entire network still can be trained by the gradient descent algorithm; we use ResNet as backbone architecture with RPN to produce regions of interest.

In [14] Region proposal network is used in Fast RCNN and Faster RCNN to predict if part of images has what probability of object present and that position. Region Proposal Networks are convolutional neural network which slide on feature maps. The network predicts anchors which score with sliding the convolutional neural networks. After this threshold is set for scores and RoIs (Region of Interest) are selected with top 2K anchors for further processing of mask. Further, these anchors are given for class, bounding box repressors. In Fast RCNN, Approximate joint training or Non-approximate joint training is used to train both RPN and Fast-RCNN. The RoI pooling layer is applied to feature maps coming from RPN network which give us the Region of Interest. In Faster RCNN, 4-Sep alternating training approach is used to train the network. First RPN is trained independently, second separate detection network is trained on proposals produced by step 1. In third step, only RPN is fined tuned keeping another network unchanged. In the Fourth step, keeping the shared convolutional neural network same and only Faster RCNN layers are trained. The anchors created in RPN show objects in different sizes and aspect ratio. Twenty thousand (20K) anchors from each image is send further for object bounding box repressors. After arranging according to classification score non maximum suppression is applied with threshold and only highest scoring bounding box are retained. Thus, at the end, we get bounding box and scores for them. Faster RCNN performs very well on PASCAL VOC and MS COCO dataset for object detection. Thus, with technique of RPN (region proposal network) Faster RCNN is lot faster and is used for real time object detection. Further, we use this architecture for our mask prediction with additional parameter of mask with BBox and class score to predict. Using this technique of sharing convolutional neural network on proposing regions reduced a lot of computation time and hence with good GPUs Faster RCNN produces good results.FPNnetwork[15] Feature Pyramid Network was further improvement Faster-RCNN architecture.

To increase the range to detect objects over large range pyramid architecture was used. These are layers of ConvNets which are stacked on top of each other for multi-scale feature representation. This hierarchy produces different feature maps at each scale with different spatial resolution. First a bottom-up pathway is used where at each level feature maps are passed to upper layer. This block are ResNets where residual blocks of ConvNets are used for feature extraction. In top-down pathway, the features are up sampled by increasing spatial resolution with lateral connection with the bottom-up feature map. This creates fine resolution feature maps at each level and we preserve the spatial and sematic resolution at each level.FPN can be implemented to replace RPN which is sliding window protocol. With a sliding window on each level of pyramid we can produce anchors on each specific level. Each anchor has training labels based on Intersection over Union. As parameters are shared on each level all feature pyramid levels have similar semantics quality which produces good performance are used in architectures of Fast RCNN where Region of Interest are produced with help of FPN of different scales. With ResNet as Backbone, FPN is implemented with attached predictor heads to all RoIs of all levels pooling is used to extract feature maps before final classification and bounding box regression layers

3.1 Instance Segmentation

Instance Segmentation is next objective after semantic segmentation. In Semantic Segmentation each pixel has to be given a class according to it location and the class label it belongs to. In instance segmentation each instance of class needs to be distinguished and a binary mask should be given for each instance. This task is difficult because there may be multiple instances of same class label very close by. To preserve the spatial information about the instance and predicting mask, class score, bounding box is achieved by Mask RCNN.

3.2 Mask R-CNNnetwork

[16] For object detection we used Faster RCNN with region proposal network. For CT scan we are using Mask RCNN for instance segmentation. In instance segmentation each mask has class bale related to it. Each instance of each class has a different class, so the goal is to localize each independent object and class related to it. Mask RCNN is same as faster RCNN with only difference is predicting segmentation mask on each RoI with bounding box and class confidence score. In Mask RCNN instead of RoIPool we use RoIAlign so that pixel to pixel alignment is restored and preserve spatial locations. Mask RCNN surpasses all different models on COCO dataset with 200ms on GPU.

Output of Mask RCNN is binary mask for each RoI. The total loss L=Lcls +Lbox + Lmask with class confidence score as Lcls, bounding box loss Lbox, and mask loss as Lmask. The mask extracts spatial structure by pixel-to pixel correspondence. The most important difference is RoIPool which is used to extract the feature map which is quantized and RoI is transformed to discrete granularity of feature map. In Mask RCNN, we use RoIAlign which reduces this quantization and properly aligns the feature maps. This can be compared to RoIWrap which does bilinear resampling. In Mask RCNN ResNet 50 and its derivatives are used as backbone architecture. Another component of Mask RCNN is FPN Feature pyramid network which uses top-down approach with lateral connections. Thus, using ResNet and FPN it gives good accuracy and fast processing on GPU,

IV. PROPOSED METHODOLOGY

4.1 Data Preprocessing

First the dataset of Covid CT scan is in NIFTY format with image, mask of right lung, left lung and infection mask. This data is the converted to COCO dataset format given in Table 1. The Binary mask are converted to segmentation which is and dictionary is prepared of all images and their respective mask of each instance. Dataset Catalog and Metadata Catalog is prepared which hold this COCO data and is passed to function which will be used by our model. In dataset, we get the image and mask of right lung, left lung and mask infection. We are using Detectron2 library which implements Mask RCNN. In this, the data is given to model in COCO format. The data given to us needs to be transformed in this format for training it in the model. The COCO data format is tabulated I table 1.

TABLE 1: COCO DATASET FORMAT

Keys	Description
file_name	Path of file where image exists
height, width	Shape of image
image_id (str or int)	Unique id
annotations (list[dict])	The class which are there for instance segmentations
bbox (list[float]	Bounding Box co-ordinates
bbox_mode (int, required)	Mode of Bounding box
category_id (int, required)	Class label
segmentation (list[list[float]] or dict)	Instance Mask in this format
Iscrowd	Boolean if instances are crowded

To convert the given mask to COCO dataset format the mask of right lung, left lung and infection is processed. Contours for the mask are obtained using find-contours method from PILLOW library which is Python Image Processing Library. Then scan every pixel and only take positive Boolean value of mask as it has mask of that class label. After getting these contours, we find height and width which are needed in COCO format. Next, we obtain segmentation property of COCO by converting these contours to polygon data structure. These polygons are converted to segmentation in which every x and y co-ordinate of mast is included. We get the bounding box by parsing these contours and we define the bounding box mode defining the type of format.

1) BoxMode.XYXY_ABS – Where x0,y0 and x1,y1 are endpoints of box
2)BoxMode.XYWH_ABS - Where x, y is co-ordinates and W, H are width and heightAfter getting all these parameters we append it in record dictionary for all images. After that form Dataset Catalog and Metadata Catalog which is given to model at runtime.

4.2 Training the model

We select the model for mask RCNN from model zoo of Detectron2 library which is .yaml format. The DatasetCatalog and MetadataCatalog which were formed in data preprocessing are passed to model. The dataloder with number of workers is fixed. Learning rate which will decide our gradient descent algorithm can

be varied according to the AP. As the Region of Interest for our 3 i.e. left lung, right lung, infection we set ROI_HEADS to 3 in Figure 2. The threshold score which will select the mask only if the class score is greater than this threshold. The test Dataset is passed to predictor and we can visualize it using Detectron2 libraries.
The Hyper parameters for the best model are as follows:

DATALOADER.NUM_WORKERS = 2
MODEL.WEIGHTS = model_zoo.get_checkpoint_url("COCO-InstanceSegmentation/mask_rcnn_R_50_FPN_3x.yaml")
SOLVER.IMS_PER_BATCH = 2
SOLVER.BASE_LR = 0.00025
SOLVER.MAX_ITER = 5000
.MODEL.ROI_HEADS.NUM_CLASSES = 3

V. EXPERIMENTAL RESULTS AND ANALYSIS

The dataset and Metadata Catalog need to be registered by the COCO instance environment. Average Precision is calculated for each instance comparing the mask of ground truth and prediction of model. The best accuracy, we achieved is given in Table 2. In Instance Segmentation, Average Precision above 70 is good accuracy according to COCO standards. The metrics for instance segmentation is Average Precision where ground Truth masks and predicted mask are compared. We see that AP for mask for left lung and right lung is good as it has constant features. The infection masks have AP of 78 which is good according to the COCO format. In Figure 5,6,7 we can see the output on input image where get bounding box, label score and mask for each ROI head which is our class label.

TABLE: 2 AVERAGE PRECISION FOR EACH CLASS LABELS

Label	**Average Precision (AP)**
Left Lung	83.56
Right Lung	85.21
Infection	78.37

The dataset is trained on different model with different backbone architectures as shown in Table 3. All backbone architectures are for COCO dataset format. The BackBone architectures are combinations of ResNet 50 and ResNet 101 with FPN, C4, DC5 as explained earlier. The Resnet 101 has more layers and skip connections as compared to ResNet 50. We have got the best accuracy is using ResNet 50. There is no any guarantee to increase accuracy on increasing layers of BackBone architecture. With hyperparameters explained above, R50-FPN gives best accuracy followed by other architectures.As shown in figure 4, figure 5 and figure 6, we get mask, bounding box and class score for right lung, left lung and infection.

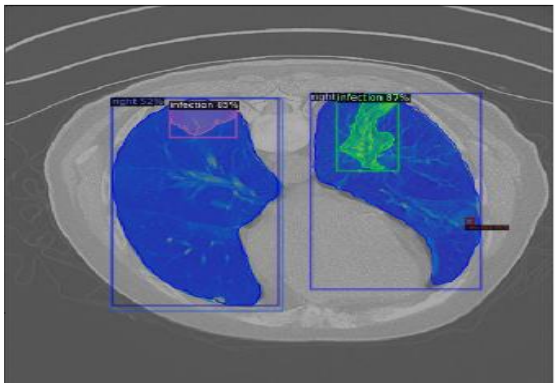

Fig. 4: Result of Model with mask on right lung, left lung and mask with bounding box score and class confidence.

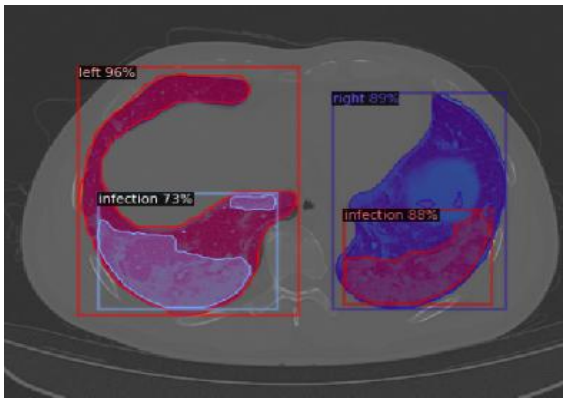

Fig. 5: Result of Model with mask on right lung, left lung and mask with bounding box score and class confidence.

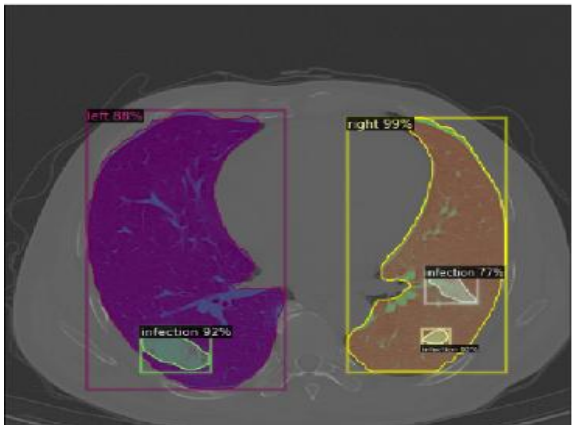

Fig. 6: Result of Model with mask on right lung, left lung and mask with bounding box score and class confidence.

The output of model is as follows:
instances: Instances object with the following fields:
pred_boxes: Boxes object storing N boxes, one for each detected instance.
scores: Tensor, a vector of N confidence scores.
pred_classes: Tensor, a vector of N labels in range [0, num_categories).
pred_masks: Tensor of shape (N, H, W), masks for each detected instance.
pred_keypoints: a Tensor of shape (N, num_keypoint, 3). Each row in the last dimension is (x, y, score). Confidence scores are larger than 0.
sem_seg: Tensor of (num_categories, H, W), the semantic segmentation prediction.

To implement Mask RCNN in Detectron, we have 3 different backbone combinations as follows:
1. FPN: Runs with ResNet and FPN
2. C4: Uses a convolutional 4 backbone and convolutional 5 head as used in Faster RCNN
3. DC5: Uses ResNet convolutional 5 backbone with dilations in it.

***T*ABLE 3: C**OMPARATIVE RESULTS WITH DIFFERENT ARCHITECTURES

Different Backbone Architecture	Average Precision
R50-C4	56.43
R50-DC5	54.2
R50-FPN	78.37
R50-C4	69.3
R50-DC5	68.44
R101-C4	71.23
R101-DC5	65.62
R101-FPN	69.34
X101-FPN	65.8

VI. CONCLUSIONS

Total 2151 images were trained with 823 instances of left lung,821 instances of right lung and 607 instances of infection. With Average Precision of 78 for infection we have got a pretty good accuracy of detecting the infection and its location and spread based on output of model given above. Multifocal areas of ground glass and consolidation can pe calculated by performing operation on mask of infection and its spread in respective lung.

Instance Segmentation with Mask RCNN helps to get each instance of infection in lung with appropriate location which can be interpreted by radiologist for research and diagnosis. With proper backbone architecture and hyperparameters set we can achieve prediction in no time on CT scan and record the features in lungs that are important for diagnosis. The pleural effusions and ground glass consolidations features obtained from Mask RCNN model helps in treatments and further diagnosis. As CT scan are better than X-Rays these scans can be used with Mask RCNN. By distinguishing each instance of infection in lung the most accurate interpretations can be obtained for treatment's end-to-end architecture of Mask RCNN with features such as RPN, RoIAlign and FPN increase accuracy and preserve spatial information of original image and getting the right regions of interest for the segmentation. With three processes i.e. class prediction, bounding box regression and binary mask prediction for each instance happening parallelly the loss is propagated very efficiently giving higher average precision.

REFERENCES

[1] D. P. Fan et al., "Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images," in IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2626-2637, Aug. 2020, doi: 10.1109/TMI.2020.2996645.

[2] Dominik Müller1, Iñaki Soto Rey1 and Frank Kramer1 "Automated Chest CT Image Segmentation of COVID-19 Lung Infection based on 3D U-Net " arXiv 2007.04774

[3] V. Rajinikanth, N. Dey, A. N. J. Raj, A. E. Hassanien, K. C. Santosh, and N. S. M. Raja, "Harmony-search and otsu based system for coronavirus disease (COVID-19) detection using lung CT scan images," 2020, arXiv:2004.03431. [Online]. Available: http://arxiv.org/abs/ 2004.03431

[4] F. Shan, Y. Gao et al., "Lung infection quantification of COVID-19 in CT images with deep learning," arXiv, 2020.

[5] C. Zheng, X. Deng et al., "Deep learning-based detection for COVID-19 from chest CT using weak label," medRxiv, 2020.in C, Chen W, Cao Y, Xu Z, Zhang X, Deng L, et al. Development and Evaluation of an AI System for COVID-19 Diagnosis. MedRxiv 2020:2020.03.20.20039834. doi:10.1101/2020.03.20.20039834.

[6] Zhou T, Canu S, Ruan S. An automatic COVID-19 CT segmentation based on U-Net with attention mechanism 2020:1– 14.

[7] Gaál G, Maga B, Lukács A. Attention U-Net Based Adversarial Architectures for Chest X-ray Lung Segmentation 2020:1–7.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI. Cham, Switzerland: Springer, 2015, pp. 234–241.

[9] Müller D, Kramer F. MIScnn: A Framework for Medical Image Segmentation with Convolutional Neural Networks and Deep Learning 2019..

[10] Y. Song et al., "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," MedRxiv, Feb. 2020, doi: 10.1101/2020.02.23.20026930.

[11] F. Isensee, P. F. Jäger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Automated design of deep learning methods for biomedical image segmentation," 2019, arXiv:1904.08128. [Online]. Available: http://arxiv.org/abs/1904.08128

[12] Tongxue Zhou, Stéphane Canu, Su Ruan "An automatic COVID-19 CT segmentation network using spatial and channel attention mechanism" arXiv:2004.06673.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun "Deep Residual Learning for Image Recognition" arXiv:1512.03385. Histology Images," IEEE Transactions on Medical Imaging, pp. 1196–1206, 2016.

[14] R. Girshick, "Fast R-CNN: Fast Region-based Convolutional Networks for object detection," IEEE International Conference on Computer Vision, pp. 1440–1448,2016.

[15] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 936-944, doi: 10.1109/CVPR.2017.106..

[16] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick "Mask R-CNN "arXiv:1703.06870.