Lightweight Model For The Prediction of COVID-19 Through The Detection And Segmentation of Lesions in Chest CT Scans

Aram Ter-Sarkisov

CitAI, Artificial Intelligence Research Centre
Department of Computer Science
City, University of London

Abstract

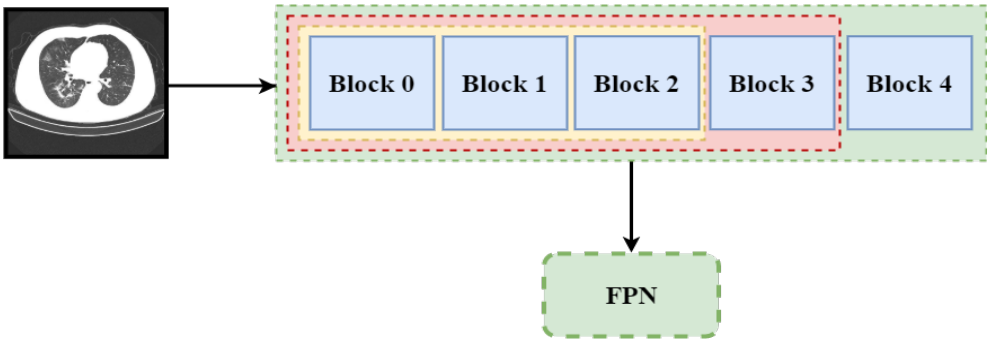
We introduce a lightweight Mask R-CNN model that segments areas with the Ground Glass Opacity and Consolidation in chest CT scans. The model uses truncated ResNet18 and ResNet34 nets with a single layer of Feature Pyramid Network as a backbone net, thus substantially reducing the number of the parameters and the training time compared to similar solutions using deeper networks. Without any data balancing and manipulations, and using only a small fraction of the training data, COVID-CT-Mask-Net classification model with 6.12M total and 600K trainable parameters derived from Mask R-CNN, achieves **91.35**% COVID-19 sensitivity, **91.63**% Common Pneumonia sensitivity, **96.98**% true negative rate and **93.95**% overall accuracy on COVIDx-CT dataset (21191 images). We also present a thorough analysis of the regional features critical to the correct classification of the image. The full source code, models and pretrained weights are available on `https://github.com/AlexTS1980/COVID-CT-Mask-Net`.

1 Introduction

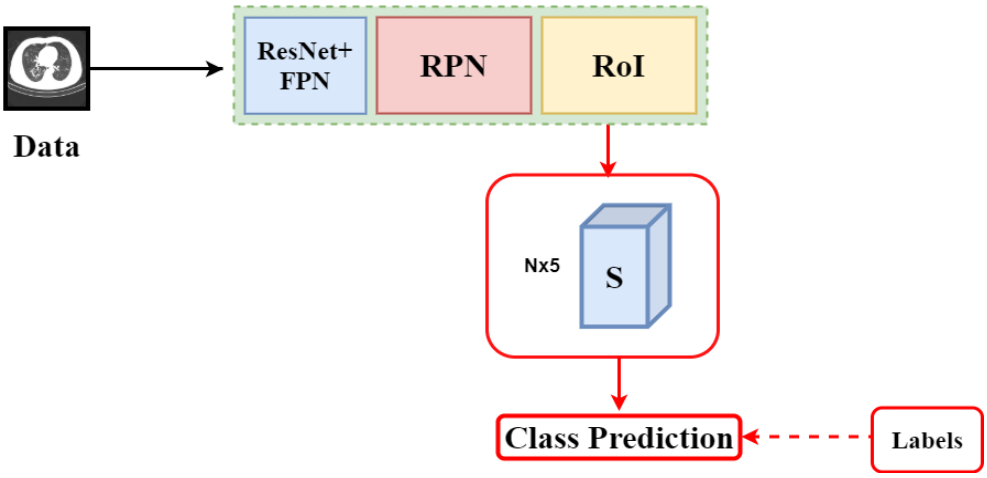
Most Deep Learning algorithms predicting COVID-19 from chest CT scans use one of the three approaches to classification: general-purpose feature extractor such as ResNet or DenseNet, or a specialized one, like COVIDNet-CT mapping the input to the predicted class, [GWW20, BGCB20, LQX$^+$20, YWR$^+$20, SZL$^+$], a combination of feature extraction and a semantic segmentation/image mask, [JWX$^+$20, WGM$^+$20, ZZHX20] and a combination of regional instance extraction and global (image) classification, [TS20a, TS20b].

Each approach has certain drawbacks regardless of the achieved accuracy of the model. These drawbacks include a small size of the dataset [BGCB20], limited scope (only two classes: COVID-19 and Common Pneumonia (CP), COVID-19 and Control, COVID-19 and non-COVID-19), [SZL$^+$, ZZHX20], large training data requirement [GWW20], large model size [LQX$^+$20, TS20a]. In [TS20a] the drawback of using a large amount of data was addressed by training a Mask R-CNN [HZRS16] model to segment areas with lesions in chest CT scans. Then, the model was augmented with a classification head that predicts the class of the image. This allowed for using a much smaller dataset for training than, e.g. [GWW20] at the cost of the size of the model, which has 34.14M total and 2.45M trainable parameters.

In this paper we overcome this drawback by attempting several variants of two different backbone models, ResNet18 and ReNet34 [HZRS16] with a single Feature Pyramid Network (FPN) layer connected to the last backbone layer. The sizes of models vary from 4.02M to 24.63M parameters (segmentation model) and 4.25M to 24.86M (classification model), with only 0.6M trainable parameters in the classification model in [TS20a]. Segmentation models with a truncated ResNet34+FPN backbone (last block of layers deleted) with 11.74M parameters achieved a mean average precision (mAP) of 0.4524, which is at par with the top 25 results of MS COCO segmentation leaderboard, `https://cocodataset.org/#detection-leaderboard`. The classification model using this backbone, with 11.74M total



(a) ResNet18/34+FPN architecture. The original models have the same architecture as in [HZRS16]. Both models have the same number of blocks, but blocks are of different size, with ResNet34 having twice as many layers in each block as ResNet18. Green: full backbone ResNet model, red: first truncated model, yellow: second truncated model. Feature Pyramid Net [LMB+14] (FPN) consists of one input and one output layer and is always connected to the last layer in the backbone net.



(b) COVID-CT-Mask-Net with the ResNet backbone net. ResNet+FPN is presented in Figure 1a. RPN and RoI are connected to the FPN feature output only (one feature layer). Mask modules are deleted from the model.

Figure 1: Architecture of the backbone nets (Figure 1a) and the lightweight COVID-CT-Mask-Net (Figure 1b). Best viewed in color.

parameters, of which only 0.6M are trainable, achieved a 91.76% COVID-19 sensitivity and 92.89% overall accuracy. An even smaller model, truncated ResNet18+FPN (6.12M parameters in the segmentation model and 6.35M in the classification model, of which also 0.6M are trainable) achieved mAP of 0.4269, COVID-19 sensitivty of 91.35% and overall accuracy of 93.95%.

2 Data and Models

We use the same datasets and train/validation/test splits as in [TS20a, TS20b] for a fair comparison. The raw chest CT scan data is taken from CNCB-COVID repository, [ZLS+20], `http://ncov-ai.big.ac.cn/download`. For the segmentation problem, the train/validation split is 500/150. All results reported in Table 2 were obtained on the validation split. The train/validation/test splits for the classification model are taken from COVIDx-CT [GWW20]: 3000 images were sampled randomly from the train split (over 60000 images) and used to train all COVID-CT-Mask-Net classifiers. Validation and test splits were used in full (21036 and 21192 images resp.).

Model	#Total parameters	#Trainable parameters	Training	Validation	Test	Ratio Train/Test
Mask R-CNN (ResNet50+FPN)	31.78M		600	150	-	-
Mask R-CNN (ResNet18+FPN)	14.52M		600	150	-	-
Mask R-CNN (ResNet18$_{T1}$+FPN)	6.12M				-	-
Mask R-CNN (ResNet18$_{T2}$+FPN)	4.02M				-	-
Mask R-CNN (ResNet34+FPN)	24.63M				-	-
Mask R-CNN (ResNet34$_{T1}$+FPN)	11.45M				-	-
Mask R-CNN (ResNet34$_{T2}$+FPN)	4.68M				-	-
COVID-CT-Mask-Net (ResNet50+FPN)	34.14M	2.36M	3K	20.6K	21.1K	7.06
COVID-CT-Mask-Net (ResNet18+FPN)	14.75M	0.6M	3K	20.6K	21.1K	7.06
COVID-CT-Mask-Net (ResNet18$_{T1}$+FPN)	6.35M					
COVID-CT-Mask-Net (ResNet18$_{T2}$+FPN)	4.25M					
COVID-CT-Mask-Net (ResNet34+FPN)	24.86M					
COVID-CT-Mask-Net (ResNet34$_{T_1}$+FPN)	11.74M					
COVID-CT-Mask-Net (ResNet34$_{T2}$+FPN)	4.92M					
COVIDNet-CT (best) [GWW20]	1.8M	1.8M	60K	20.6K	21.1K	0.353
COVNet [LQX$^+$20]	25.61M	25.61M	3K	370	438	0.129
ResNet18 [BGCB20]	11.69M	11.69M	528		90	0.17

Table 1: Comparison of the models' sizes and data splits used for training, validation and testing. $T1$ and $T2$ refer to the truncated models (1 and 2), see Figure 1a. FPN is used in all our models because it helps with the reduction in the total number of parameters and improves the final result. The number of trainable parameters in the classifiers with ResNet18 and ResNet34 backbones varies insignificantly.

Apart from the subtraction of the global mean and division by the global standard deviation, no other data manipulations were applied to either dataset.

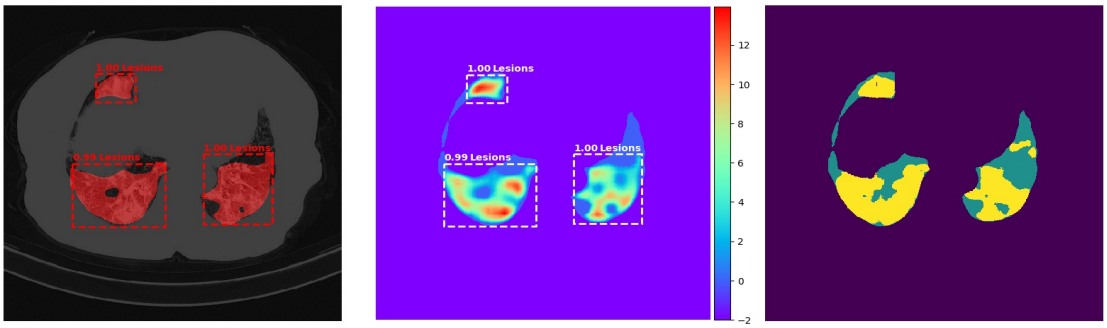
The main contribution of this paper is the training of the lightweight segmentation and classification models with ResNet18+FPN and ResNet34+FPN backbones to produce results that beat or approach those of the full-sized ResNet50+FPN models with 4 FPN layers for both tasks. In all backbone nets the last (problem-specific) fully connected and average pooling layers were removed. For the full list of model sizes and comparison to the benchmarks, see Table 1. We consider three versions of each model:

1. Full model. This is the baseline for each experiment, in Figure 1a it is the model that contains all blocks (green), and FPN module is connected to the last fourth block. FPN input is downsized from 512 to 256 maps.

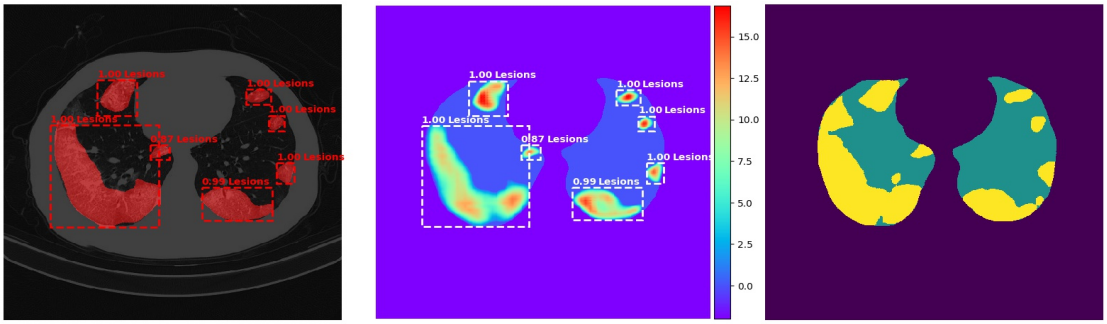
2. ResNet 18/34$_{T1}$: the first truncated model. The last ResNet block is removed, FPN is connected to Block 3, and FPN has the same number of maps (256) as the last block in ResNet.

3. ResNet 18/34$_{T2}$: the second truncated model. The last two blocks in ResNet are removed, and FPN is connected to ResNet Block 2. FPN upsizes the input from 128 to 256 maps.

For the training and evaluation of the segmentation model we used only one positive class, 'Lesion', obtained by merging the masks for the Ground Glass Opacity (GGO) and Consolidation (C) areas, see [TS20b]. For the training and evaluation of the classification model, we use the labeling convention from COVIDx-CT and CNCB: 0 for the Control class, 1 for Common Pneumonia and 2 for COVID-19.

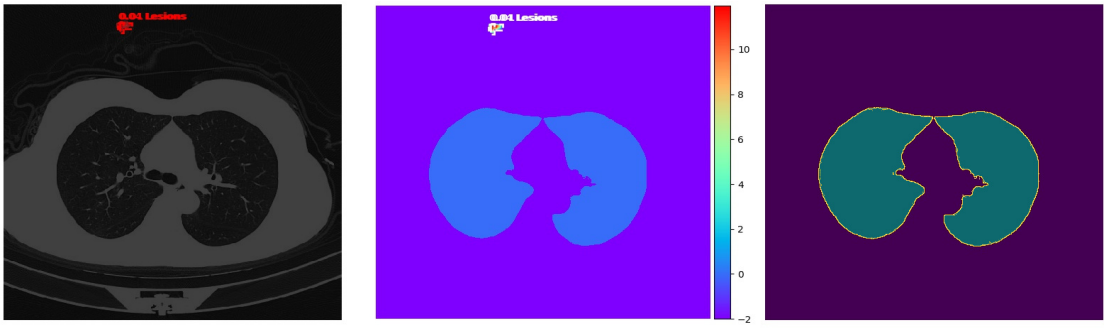


(a) CT Scan Slice #88.



(b) CT Scan Slice #107.

(c) CT Scan Slice #48.



(d) CT Scan Slice #46.

Figure 2: Segmentation results of ResNet34$_{T1}$ model across a number of CT scan slices (different levels of the lungs). Images in each Figure 2a-2d pertain to the same scan slice. Figures 2a and 2b are COVID-19 positive, Figures 2c and 2d are Control/Negative. Column 1: Input images superimposed with the final mask prediction, bounding box, class and confidence scores for each instance. Column 2: Regional (mask) score maps. Outputs from each RoI are *independent of each other*, meaning that they were obtained from different RoIs independently and combined in the same score map. To avoid the image clutter, only the highest-ranking predictions are displayed. Column 3: Ground truth lesion and lungs masks. Column 4: true labels in dark green (0: Control, 2: COVID-19) and class scores predicted by COVID-CT-Mask-Net in red. Best viewed in color.

Model	AP@0.5IoU	AP@0.75IoU	AP@[0.5 : 0.95]	Training
ResNet18+FPN	0.4111	0.2827	0.2990	128.75
ResNet18$_{T1}$+FPN	0.6262	0.4659	0.4269	108.8
ResNet18$_{T2}$+FPN	0.5633	0.3761	0.3325	126.9
ResNet34+FPN	0.5357	0.3333	0.3465	119.23
ResNet34$_{T1}$+FPN	**0.6787**	**0.4848**	**0.4524**	134.01
ResNet34$_{T2}$+FPN	0.5538	0.4020	0.3536	129.2
ResNet50+FPN[TS20b]	0.6192	0.4522	0.4468	137.38

Table 2: Average precision of segmentation models and training time (in minutes).

Model	COVID-19	Pneumonia	Normal	Overall	Training
ResNet18+FPN	87.51%	77.31%	74.57%	78.18%	151.2
ResNet18$_{T1}$+FPN	91.35%	91.63%	**96.98**%	**93.95**%	106.6
ResNet18$_{T2}$+FPN	84.05%	85.81%	93.01%	88.66%	113.66
ResNet34+FPN	86.98%	**94.27**%	71.12%	82.45%	163.81
ResNet34$_{T1}$+FPN	**91.76**%	91.70%	94.36%	92.89%	144.73
ResNet34$_{T2}$+FPN	89.25%	93.32%	92.11%	91.99%	112.6
COVID-CT-Mask-Net [TS20b]	93.88%	95.06%	96.91%	95.64%	460

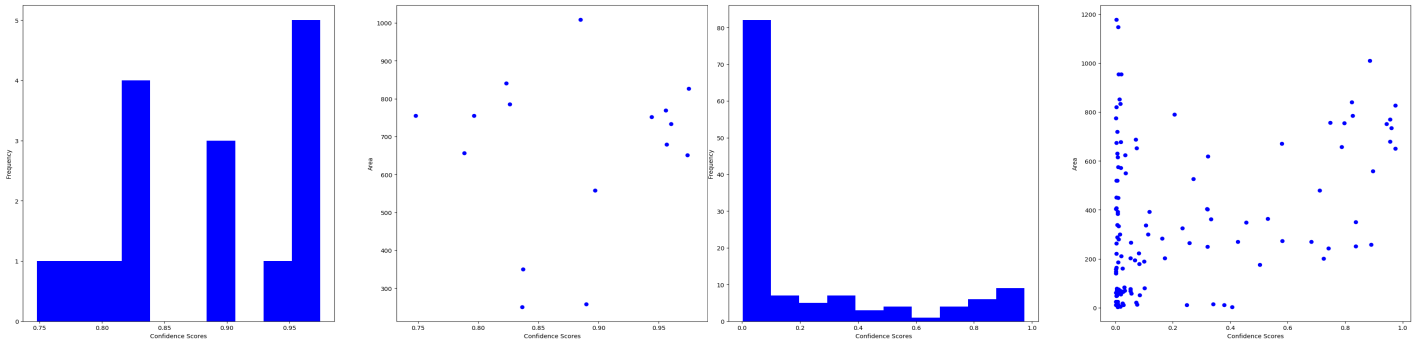
Table 3: Class sensitivity and overall accuracy results on COVIDx-CT test data (21192 images) and the training time (in minutes).

3 Experimental results

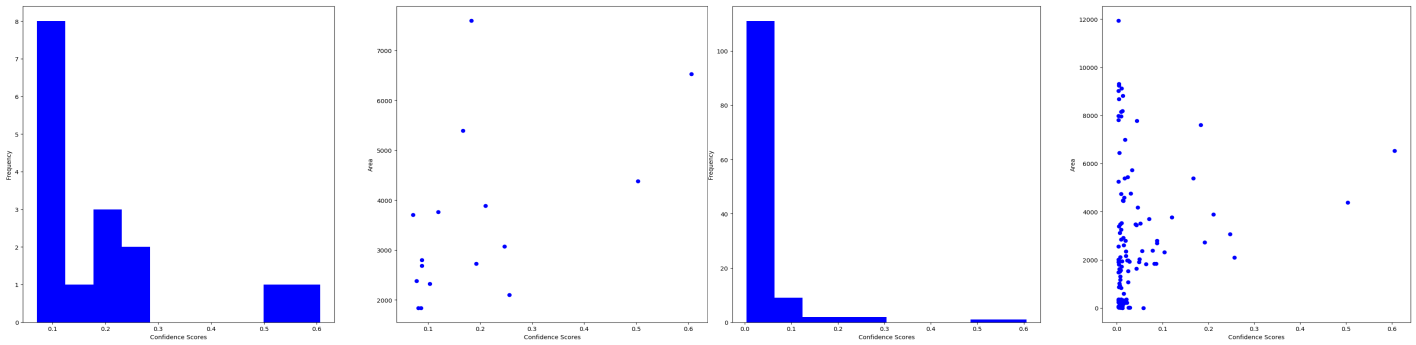
For the explanation of the accuracy metrics and comparison, see [LMB+14], as we adapt MS COCO's average precision (AP) at two Intersect over Union (IoU) threshold values and mean AP across 10 IoU thresholds between 0.5 and 0.95 with at 0.05 step. The hyperparameters of the classification model are the same as in the best model in [TS20b], with the NMS threshold of 0.75 and RoI score$_\theta = -0.01$, except that we reduce the RoI batch size from 256 to 128 and the total model size from 34.14M and the number of trainable parameters from 2.45M (ResNet50+FPN) to 6.12M and 0.6M respectively (ResNet18$_{T1}$+FPN) with only about 2% drop in the COVID-19 sensitivity and 1.5% drop in overall accuracy. For the comparison to larger models, see [TS20b]. Results for training full and truncated lightweight models are presented in Table 2. The best segmentation model we trained, ResNet34 with a deleted last block (ResNet34$_{T1}$+FPN) with 11.45M parameters outperforms the best model in [TS20b], ResNet50+FPN, which is almost 3 times larger. The classification model derived from it also achieves the highest COVID-19 sensitivity among the lightweight models, 91.76%. The second-best segmentation model, ResNet18$_{T1}$+FPN, achieves 0.4269 overall accuracy, with only 6.12M parameters. The classification model derived from it achieves the highest overall accuracy of 93.95% and the second-best COVID-19 sensitivity among the lightweight models of 91.35%.

We experimented with a number of additional hacks for each model:

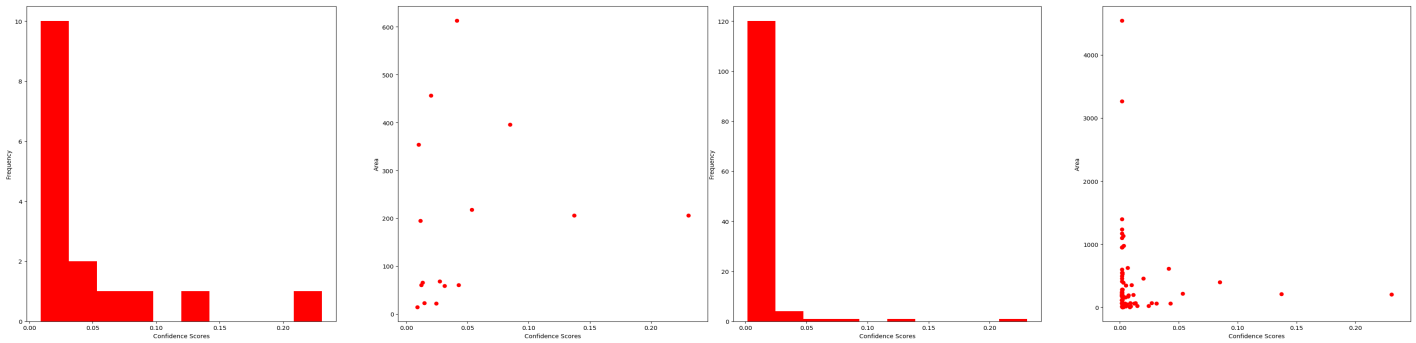
1. Replacement of SOFTMAX with SIGMOID activation function for the outputs of RoIs (segmentation model, test stage). Faster R-CNN implementation [RHGS15] uses SOFTMAX for scoring C outputs of each RoI (C:total number of classes, including background). The score of each non-background prediction is compared to the score threshold (RoI score$_\theta = 0.75$) to decide whether to keep this prediction or discard, so obviously it is very unlikely to get more than a single prediction out of each RoI. At the same time, even low-ranking predictions are tested for Non-max suppression (0.75 in all models). Replacing SOFTMAX with SIGMOID makes predictions independent of each other in each RoI, and hence have a higher chance of being accepted as a prediction. This approach did not yield a consistent improvement across all models, so we left it out of the final result.



(a) COVID-19



(b) Common Pneumonia



(c) Control (Negative)

Figure 3: Distribution of the confidence score and scatterplot of the area vs the confidence score in a CT slice with COVID-19 (Figure 3a), Common pneumonia (Figure 3b) and Control (Figure 3c). Columns 1 and 2: top 16 predictions in each image, Columns 3 and 4: all 128 predictions in each image. Best viewed in color.

2. Removal of empty boxes/replication of the predictions (classification model). Deletion of empty boxes (bounding boxes with the area of 0) improved the models' predictive power, but reduced the output size of the pre-defined RoI batch size (128), which is converted to a feature vector in the classification module **S**, and hence must remain fixed (see [TS20a] for details of batch to feature method). To resolve the problem, we applied a hack at this stage: the missing predictions (difference between the pre-defined RoI batch size and the current output) are sampled from the valid predictions *maintaining their ranking order*. What this means is that each sampled prediction is inserted in the batch between the box selected for replication and the next prediction. For example, if the predictions are $[3, 1, 2]$ and the first and the last ones are sampled for replication, the batch becomes $[3, 3, 1, 2, 2]$. This maintains the order of ranking of the predictions in the sample, which is what the classifier learns to predict the class of the input image.

3. Removal of small areas in the data (segmentation model). Most areas with GGO and C are small, see [ZZX+20,

ZYW+20, TS20a] for the detailed discussion of the distribution of lesions in chest CT scans. Training the segmentation model to predict small lesion areas leads both to lower precision at test stage, and lower COVID-19 sensitivity of the classification model. We decided to merge all GGO and C patches of less than 100 pixels with the background. As a result, the model's accuracy improved, as the predictions were not biased towards very small areas.

3.1 Identification of areas critical for COVID-19 prediction.

Apart from the CT scan segmentation and classification, deep learning models can help explain factors associated with COVID-19, e.g. in the form of attention maps [YMK+20, YWR+20] or using specialized tools like GSInquire [GWW20] that identify critical factors in CT scans. The advantage of using instance segmentation models like Mask R-CNN is the detection, scoring and segmentation of classified isolated areas that contribute to the condition (class of the image). This is a more accurate and explicit approach than either feature maps in vanilla convnets, that merely indicate the strength of presence of nameless features, or full-image pixel-level score maps in FCNs, that do not distinguish between different instances of the class. Mask R-CNN naturally evolves separate instances of regional predictions that can overlap, both at bounding box and mask level.

This is illustrated in Figure 2 for the output of ResNet34$_{T1}$ model. Figures 2a and 2b are COVID-19 positive (column 3: lesion masks), Figures 2c and 2d are COVID-19 negative (Control), which is reflected in column 3 (column 3: no lesion mask). The first column is the input image overlaid with bounding box predictions for the lesion areas with a box confidence score and mask predictions for the object in the bounding box. Mask predictions are normalized using SIGMOID function, with a threshold of 0.5 that serves as a filter for the foreground (i.e. all pixels with scores exceeding the threshold are considered foreground/instance). For the (combined) mask score map in column 2 in Figure 2, we used raw (before SIGMOID normalization) scores from the mask prediction layer. Each prediction is done by Mask R-CNN *independently*, i.e. Mask R-CNN extracts the RoI from the FPN layer using RoIAlign ([HGDG17]) and predicted bounding box coordinates, filters it through the de-convolution layer to obtain a fixed-size (28×28) mask score map with pixel logits that is then resized to the size of the bounding box prediction. Looking at the combined mask score maps, it becomes clear how COVID-CT-Mask-Net learns to use the score information. Each combined score map for the negative images contains only one prediction with a very low confidence score (< 0.01), for which COVID-CT-Mask-Net outputs large logit values for Class 0 (Figure 2, column 4). Score maps for COVID-19 images contain a number of large high-scoring predictions. The total number of predictions in each image is the same due to the RoI score$_\theta = -0.01$, we plotted only a small number of highest-scoring RoIs to avoid image cluttering.

The analysis of the mask score maps in column 2, Figure 2 explains the effectiveness of the RoI batch to feature vector method, which is the main idea behind the transformation of Mask R-CNN into the classification model. Both the location (bounding box coordinates) and the importance (confidence score) of the areas critical to the COVID-19 diagnosis are output by Mask R-CNN and accepted by the classification module **S** *in the decreasing order/rank of their importance/confidence scores*. Since the RoI batch size is fixed regardless of the actual confidence scores (for the detailed discussion of the batch to feature method see [TS20a]), **S** can learn this ranking, and, eventually, associate a number of high-ranking RoIs located in the critical areas (see [ZZX+20, ZYW+20] for the analysis of COVID-19 vs Common Pneumonia chest CT scans) with the particular image class. To demonstrate this, we also plot the histograms of the confidence scores and the scatterplots of the confidence scores vs RoI area (bounding box size) in three difference CT scans (COVID-19, Common Pneumonia and Control) in Figure 3. Top 16 regions (columns 1-2) in Figure 3a is dominated by several mid-size ($\approx$800 pixels) high-scoring (≥ 0.95) critical areas, and the full batch (128 regions) in Columns 3-4 follows what seems to be a bathtub-shaped distribution: despite the fact that the majority of regions have a very low score (regardless of the size), there is a sufficient number of high-scoring regions in the batch for the model to learn the true class. Common Pneumonia distribution is presented in Figure 3b: there's a small number of very large (4000-6000 pixels) mid-to-high scoring regions with the scores between 0.1 and 0.6. The distribution of Control(Negative), Figure 3b is also distinct: the highest-scoring box (0.24) is very small ($\approx$200 pixels), and the rest of the batch have scores practically indistinguishable from 0 regardless of the size.

4 Conclusions

We presented several variants of lightweight segmentation and classification models based on Mask R-CNN with ResNet18+FPN and ResNet34+FPN backbone networks. With as few as 11.74M total and 600K trainable parameters, COVID-CT-Mask-Net classification model with ResNet34$_{T1}$+FPN backbone achieves a 91.76% COVID sensitivity and 92.89% overall accuracy across three classes (COVID-19, Common Pneumonia, Control). The model with ResNet18$_{T1}$+FPN backbone with 6.35M parameters achieves the COVID-19 sensitivity of 91.35% and overall accuracy of 93.95%. The smallest model with ResNet18$_{T2}$+FPN backbone with just 4.25M parameters achieves a 84.05% COVID-19 sensitivity and 88.66% overall accuracy. We also presented an in-depth analysis of the mask score maps across all three image classes and the distribution of the features of the predicted critical areas (confidence score, size). We demonstrated the ability of Mask R-CNN to explicitly detect and segment areas critical for the accurate prediction of COVID-19 and other classes.

References

[BGCB20] Charmaine Butt, Jagpal Gill, David Chun, and Benson A Babu. Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence*, pages 1–7, 2020.

[GWW20] Hayden Gunraj, Linda Wang, and Alexander Wong. Covidnet-ct: A tailored deep convolutional neural network design for detection of covid-19 cases from chest ct images. *arXiv preprint arXiv:2009.05383*, 2020.

[HGDG17] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[JWX+20] Shuo Jin, Bo Wang, Haibo Xu, Chuan Luo, Lai Wei, Wei Zhao, Xuexue Hou, Wenshuo Ma, Zhengqing Xu, Zhuozhao Zheng, et al. Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks. *medRxiv*, 2020.

[LMB+14] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[LQX+20] Lin Li, Lixin Qin, Zeguo Xu, Youbing Yin, Xin Wang, Bin Kong, Junjie Bai, Yi Lu, Zhenghan Fang, Qi Song, et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology*, 2020.

[RHGS15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[SZL+] Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Huiying Zhao, Yusheng Jie, Ruixuan Wang, Yutian Chong, Jun Shen, Yunfei Zha, and Yuedong Yang. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *medRxiv*.

[TS20a] Aram Ter-Sarkisov. Covid-ct-mask-net: Prediction of covid-19 from ct scans using regional features. *medRxiv*, 2020.

[TS20b] Aram Ter-Sarkisov. Detection and segmentation of lesion areas in chest ct scans for the prediction of covid-19. *medRxiv*, 2020.

[WGM+20] Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Chao-Wei Zhao, and Ming-Ming Cheng. Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. *arXiv preprint arXiv:2004.07054*, 2020.

[YMK+20] Shakib Yazdani, Shervin Minaee, Rahele Kafieh, Narges Saeedizadeh, and Milan Sonka. Covid ct-net: Predicting covid-19 from chest ct images using attentional convolutional network. *arXiv preprint arXiv:2009.05096*, 2020.

[YWR+20] Tao Yan, Pak Kin Wong, Hao Ren, Huaqiao Wang, Jiangtao Wang, and Yang Li. Automatic distinction between covid-19 and common pneumonia using multi-scale convolutional neural network on chest ct scans. *Chaos, Solitons & Fractals*, 140:110153, 2020.

[ZLS+20] Kang Zhang, Xiaohong Liu, Jun Shen, Zhihuan Li, Ye Sang, Xingwang Wu, Yunfei Zha, Wenhua Liang, Chengdi Wang, Ke Wang, et al. Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography. *Cell*, 2020.

[ZYW+20] Dahai Zhao, Feifei Yao, Lijie Wang, Ling Zheng, Yongjun Gao, Jun Ye, Feng Guo, Hui Zhao, and Rongbao Gao. A comparative study on the clinical features of covid-19 pneumonia to other pneumonias. *Clinical Infectious Diseases*, 2020.

[ZZHX20] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: a ct scan dataset about covid-19. *arXiv preprint arXiv:2003.13865*, 2020.

[ZZX+20] Wei Zhao, Zheng Zhong, Xingzhi Xie, Qizhi Yu, and Jun Liu. Ct scans of patients with 2019 novel coronavirus (covid-19) pneumonia. *Theranostics*, 10(10):4606, 2020.