ISSN 2563-7568

RESEARCH ARTICLE

Lightweight Model for the Prediction of COVID-19 Through the Detection and Segmentation of Lesions in Chest CT Scans

Aram Ter-Sarkisov

Department of Computer Science, City, University of London, London, United Kingdom

Abstract

We introduce a lightweight model that segments areas with the Ground Glass Opacity and Consolidation and predicts COVID-19 from chest CT scans. The model uses truncated ResNet18 and ResNet34 as a backbone net, and Mask R-CNN functionality for lesion segmentation. Without any class balancing and data 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, 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). The full source code, models and pre-trained weights are available on https://github.com/AlexTS1980/COVID-CT-Mask-Net.

Key Words: *Convolutional neural networks, COVID-19, Lesion segmentation, Lesion detection*

Corresponding Author: *Aram Ter-Sarkisov, Lecturer, Department of Computer Science, City, University of London, London EC1V 0HB, Tel: +44 7576286028; E-mail: alex.ter-sarkisov@city.ac.uk*

Received Date: *December 08, 2020,* **Accepted Date**: *January 18, 2021,* **Published Date**: *March 31, 2021*

Citation: *Ter-Sarkisov A. Lightweight Model for the Prediction of COVID-19 Through the Detection and Segmentation of Lesions in Chest CT Scans. Int J Auto AI Mach Learn. 2021;2(1):01-15.*

1. Introduction

COVID-19 pandemic that started in December 2019 is an ongoing global pandemic of a viral that, by early January 2021 led to more than 1.87M deaths and 86.5M confirmed cases. The pandemic has put an enormous pressure on the national healthcare systems, and caused a large number of negative economic consequences.

Since its onset, researchers in computer science and, especially in machine learning, have contributed greatly to battling the crisis in the form of developing datasets of computer tomography (CT) scans, radiology (x-rays) and ultrasound scans. Another substantial contribution of the researchers is the development of various methods of diagnosing COVID-19 from the data using Deep Learning methods.

These studies leverage the observation that COVID-19 has a number of lesion manifestations in chest CT scans that sets is apart from other types of pneumonia, such as Ground Glass Opacity, crazy paving pattern, Consolidation, and others [1]. Although observable, these differences may not be statistically significant, and differ in the specifics, e.g. the shape of a Ground Glass Opacity lesion in a specific area varies depending on the type of pneumonia, or similar looking lesions can be unilateral in common pneumonia (CP) and bilateral in COVID-19 patients. Deep Learning algorithms are known to pick up on such patterns and, when deployed in hospitals, help radiologists with the fast and correct diagnosing.

In this paper we present models using several variants of two different backbone models, ResNet18 and ReNet34 [2] 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 [3]. In the context of this paper, we refer to these models as lightweight models, implying that they use fewer weights and have simpler architecture than their full variants presented in the original ResNet paper. The same is valid for Feature Pyramid Network, which we truncated compared to the one used in [3].

The novelty and findings of our investigation can be summarized in the following way:

1 Truncation of ResNet18/34 backbone models by removing semantically rich top levels and Feature Pyramid Network layers substantially reduces the training time compared to deeper models like ResNet50 and improves the accuracy of the models, both in terms of lesion segmentation and COVID-19 prediction compared to full ResNet18/34 models.

2 The best segmentation model with the truncated ResNet34+FPN backbone (11.47M weights) achieved 44.76% mean average precision on the test segmentation split (main MS COCO criterion), which is at par with the top 25 results of MS COCO leaderboard (https://cocodataset.org/#detection-leaderboard) and the corresponding classification model

(11.74M weights) 92.89% overall accuracy on the classification test split and 91.76% COVID-19 sensitivity.

3 Segmentation model with the smallest truncated backbone (ResNet18+FPN, 4.02M weights) achieved the mean average precision of 37.59% and its classification variant (4.28M weights) the overall accuracy of 88.66% and COVID-19 sensitivity of 84.05%.

4 We introduced several improvements of the methodology introduced in [3], which includes the removal of small lesion areas in segmentation masks and empty Regions of Interest (RoI) areas from the RoI batch while maintaining its size.

5 We provide an in-depth analysis of the RoI batch distribution and its effect on the class prediction. By comparing the distribution of RoIs across each class pair using one-sided Kolmogorov-Smirnov test, we show that their differences are statistically significant except Common Pneumonia/Control.

The rest of the paper is structured as follows: Section 2 reviews the relevant literature, Section 3 introduces the data and the models, Section 4 discusses the experimental results and provides the analysis of the RoI batches, Section 5 concludes.

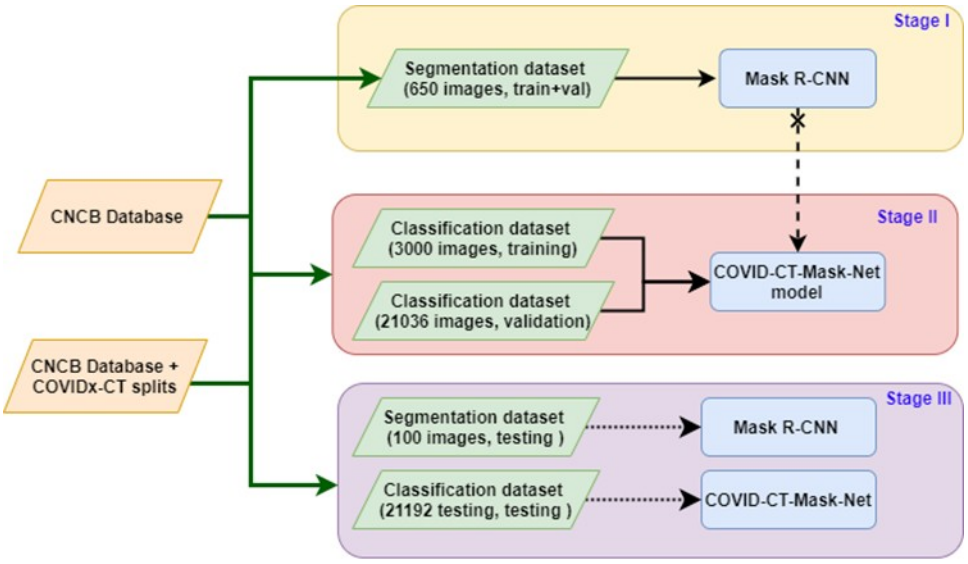
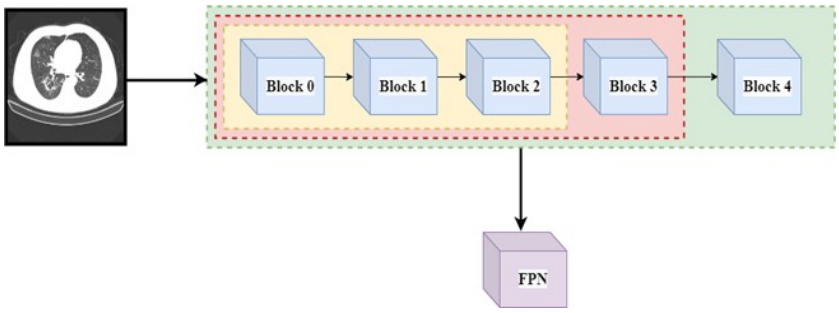
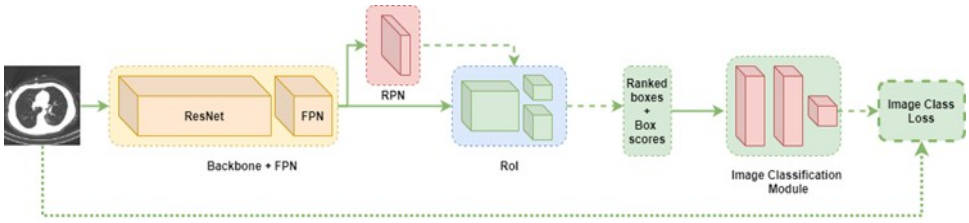


Figure 1: *The overall structure of the method. Stage I: training of the segmentation model, Stage II: training of the lightweight classification model, Stage III: testing of the segmentation and classification models. Normal green arrows: extraction of the raw and COVIDx-CT data into train/validation/test splits, normal black arrows: training data, dotted arrows: test data, broken arrow with a cross: weight copy.*



2(a): *ResNet18/34+FPN architecture. The original models have the same architecture as in [2]. 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 [4] (FPN) consists of one input and one output layer and is always connected to the last layer in the backbone net.*



2(b): *COVID-CT-Mask-Net with the ResNet backbone net. ResNet+FPN is presented in Figure 2a. RPN and RoI are connected to the FPN feature output only (one feature layer). Mask modules are deleted from the model. Normal arrows: tensors/ features, broken arrows: batches, dotted arrow: image label.*

Figure 2: *Architecture of the backbone nets Figure 2a and the lightweight COVID-CT-Mask-Net Figure 2b. Best viewed in color.*

2. Literature Review

In this Section we discuss the literature on which our results are based.

Most publicly available chest CT scans slices datasets labelled at the image level contain observations for only two classes: COVID-19 and Common Pneumonia [5], COVID-19 and Control [6-8], COVID-19 and non-COVID-19 [5], or only COVID-19, such as the segmentation dataset: https://medicalsegmentation.com/covid19. The only large-scale open-source dataset that both contains a large number of observations for 3 classes and segmentation images is CNCB-NCOV [9]. There's a total of over 104K slices across three classes: COVID-19, CP and

Normal labelled at image level and 750 slices labelled at pixel level (normal lungs, Groung Glass Opacity and Consolidation classes) taken from 150 patients with COVID-19.

Most studies using both 2 and 3-class datasets, summarized, e.g. in [10-13] do not release open-source datasets, and are often limited in scope in terms of the number of available images, and therefore require additional investigation to determine their ability to generalize to larger datasets.

Most Deep Learning algorithms predicting COVID-19 from chest CT scans use one of the five approaches to classification: combination of convolutional neural net and a machine learning method, [14], general-purpose feature extractor such as ResNet or DenseNet, or a specialized one, like COVIDNet-CT mapping the input to the predicted class, [6,15-18], a combination of feature extraction and a semantic segmentation/image mask, [19-21] and a combination of regional instance extraction and global (image) classification, [3,22]. A number of fusions of convolutional neural nets and recurrent neural nets (long-short term memory network, LSTM) were also presented [23,24].

Each of these approaches has certain drawbacks regardless of the declared accuracy of the model. These drawbacks include a small size of the dataset [15], limited scope (e.g. only two classes are compared: COVID-19 vs CP[5], COVID-19 vs Control [6], COVID-19 vs non-COVID-19 [21], large training data requirement [16], large model size [17,3]. In [3] the drawback of using a large amount of data was addressed by training a Mask R-CNN [25] 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. [16] at the cost of the size of the model, which has 34.14M total parameters, of which 2.45M are trainable.

To the best of our knowledge, this is the first study that trains a range of lightweight models using small amounts of data, that generalize well to the unseen images, while maintaining a high level of explainability in the form of explicit lesion segmentation Section 4.1 and analysis of the training batch distribution Section 4.2.

3. Data and Models

In this Section we present the key details of the dataset and the methodology.

3.1. Data

Key information about the data and model sizes is presented in Table 1. The overall flowchart of the methodology is presented in Figure 1.

We use the same datasets and train/validation/test splits as in [3,22] for a fair comparison. The raw chest CT scan data is taken from CNCB-COVID repository [9], http://ncov-ai.big.ac.cn/download. For the segmentation problem, the train/validation split is 650, test split is 100. The train/validation/test splits for the classification model are taken from COVIDx-CT [16]: 3000 images (1000/class) 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.). Test data is relatively balanced: 25% COVID-19, 35% Common Pneumonia, 45% Normal/Control.

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 [22]. 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.

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 2a*. 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.*

Model	#Total parameters	#Trainable parameters	Training	Validation	Test	Ratio Test/Train
Mask R-CNN (ResNet50+FPN)	31.78M		650	-	100	0.15
Mask R-CNN (ResNet18+FPN)	14.52M					
Mask R-CNN (ResNet18T1+FPN)	6.12M					
Mask R-CNN (ResNet18T2+FPN)	4.02M					
Mask R-CNN (ResNet34+FPN)	24.63M		650	-	100	0.15
Mask R-CNN (ResNet34T1+FPN)	11.45M					
Mask R-CNN (ResNet34T2+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					
COVID-CT-Mask-Net (ResNet18T1+FPN)	6.35M					
COVID-CT-Mask-Net (ResNet18T2+FPN)	4.25M					
COVID-CT-Mask-Net (ResNet34+FPN)	24.86M	0.6M	3K	20.6K	21.1K	7.06
COVID-CT-Mask-Net (ResNet34T1 +FPN)	11.74M					
COVID-CT-Mask-Net (ResNet34T2+FPN)	4.92M					
COVIDNet-CT (best) [2]	1.8M	1.8M	60K	20.6K	21.1K	0.353
COVNet [7]	25.61M	25.61M	3K	370	438	0.129
ResNet18 [1]	11.69M	11.69M			90	0.17
				528		

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

3.2. Models

Methodology of the models in this paper follows the one earlier presented in [3,22] mainly in the conversion of a batch of ranked Regions of Interest to a feature vector.

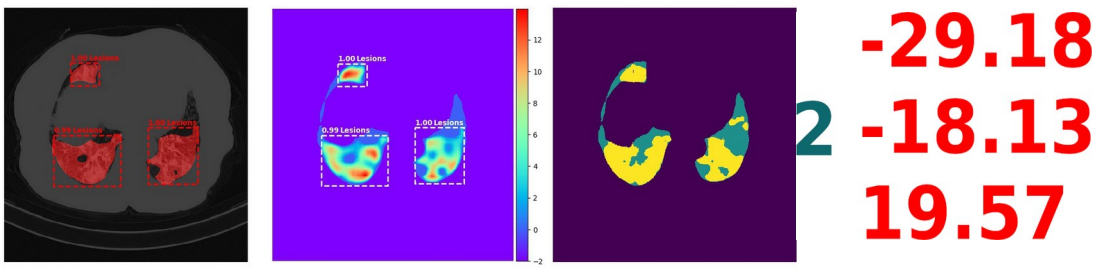
The main contribution of this paper is the training of the lightweight segmentation and classification models with ResNet18+FPN and ResNet34+FPN backbones to produce results that beat or approach those of the full-sized ResNet50+FPN models with 4 FPN layers for both tasks. In all backbone nets the last (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 2a 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.

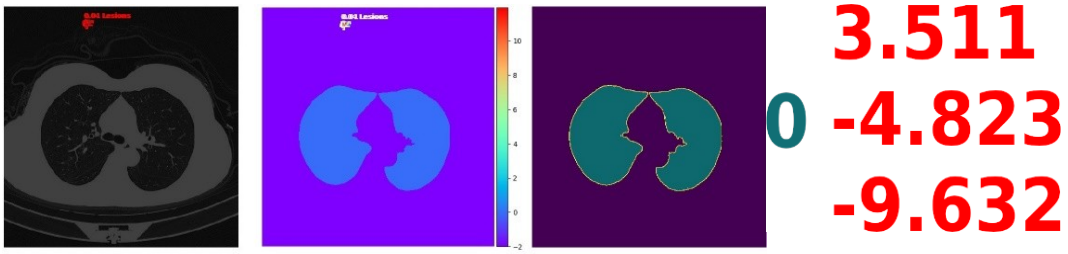
2 ResNet 18/34$_{T1}$: the first truncated model. The last ResNet block is removed, FPN is connected to Block 3, and FPN has the same number of maps (256) as the last block in ResNet.

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.

Feature Pyramid Network, introduced in [4] rescales high-level features using convolution kernels, which helps with finding objects of small size in images. In the original Torchvision implementation (https://pytorch.org/docs/stable/torchvisionmodels.html#object-detection-instance-segmentation-and-person-keypoint-detection) used in [3,22] the last layer in each of 4 blocks of ResNet50 are connected to the FPN layer, therefore Faster and Mask R-CNN use all 4 FPN blocks, increasing the total weight count by 3.34M parameters. In our implementation we delete not just the ResNet blocks, but all except the last FPN layer, reducing the related number of parameters to 0.65M in each model.



3(a)



3(b)

Figure 3: *Segmentation results of ResNet34$_{T1}$ model for two CT scan slices (different levels of the lungs). Images in each Figure 3a-3b pertain to the same scan slice. Figure 3a is COVID-19 positive, Figure 3b is 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.*

Table 2: *Average precision of segmentation models and training time (in minutes). Best lightweight results in bold.*

Model	AP@0.5IoU	AP@0.75IoU	AP@[0.5:0.95]	Training
ResNet18+FPN	0.3699	0.2898	0.3137	128.75
ResNet18T1+FPN	0.4995	0.3778	0.3932	108.8
ResNet18T2+FPN	0.4993	0.3852	0.3759	126.9
ResNet34+FPN	0.5357	0.3333	0.3465	119.23
ResNet34T1+FPN	0.5988	0.4506	0.4476	134.01
ResNet34T2+FPN	0.4491	0.3118	0.3404	129.2
ResNet50+FPN[22](merged masks)	0.6192	0.4522	0.4468	137.38
ResNet50+FPN[22](separate masks)	0.5020	0.4198	0.3871	145.54

Table 3: *Class sensitivity and overall accuracy results on COVIDx-CT test data (21192 images) and the training time (in minutes) and several reference models. Best lightweight results in bold. Results for CovNet iCTCF and COVID CT-Net are taken from the respective publications, results for COVIDNet-CT were obtained by running the provided opensource code, which may not fully match the results in the paper. The sizes of test datasets are specified in Table 1. For [17] mean AUC across 3 classes instead of overall accuracy is provided. For [1] only AUC for COVID-19 prediction is provided as the article was retracted and is no longer publicly available. F1 score is class-adjusted (COVID-19: 20%, Common Pneumonia: 35%, Normal: 45%).*

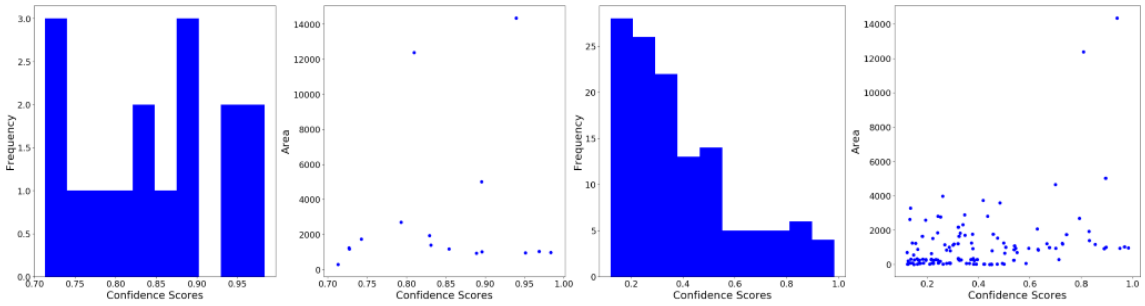
Model	COVID-19	Pneumonia	Normal	Overall	F1 score	Training
ResNet18+FPN	87.51%	77.31%	74.57%	78.18%	79.11%	151.2
ResNet18T1+FPN	91.35%	91.63%	**96.98%**	**93.95%**	94.01%	106.60
ResNet18T2+FPN	84.05%	85.81%	93.01%	88.66%	88.69%	113.66
ResNet34+FPN	86.98%	**94.27%**	71.12%	82.45%	83.02%	163.81
ResNet34T1+FPN	**91.76%**	91.70%	94.36%	92.89%	92.92%	144.73
ResNet34T2+FPN	89.25%	93.32%	92.11%	91.99%	92.00%	112.60
COVID-CT-Mask-Net [22](merged masks)	92.68%	96.69%	97.74%	96.63%	96.69%	430.00
COVID-CT-Mask-Net [22](separate masks)	93.88%	95.06%	96.91%	95.64%	95.71%	460.00
COVIDNet-CT [2]	92.69%	97.00%	98.00%	97.10%	N/A	N/A
CovNet [7]	90.00%	87.00%	94.00%	95.00%	N/A	N/A
iCTCF[28]	80.00%	N/A	51.00%	85.00%	N/A	N/A
COVID CT-Net[26]	90.00%	N/A	93.55%	92.00%	N/A	N/A

4. Experimental Results

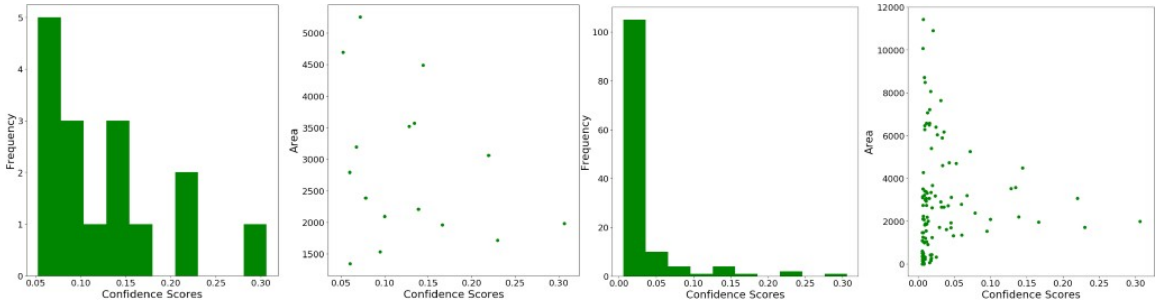
For the explanation of the accuracy metrics and comparison, see [27], as we adapt MS COCO's average precision (AP) at two Intersect over Union (IoU) threshold values and mean AP across 10 IoU thresholds between 0.5 and 0.95 with at 0.05 step. To test the models we used RoI and RPN NMS threshold of 0.75 and confidence score threshold of 0.75. The hyperparameters of the classification model are the same as in the best model in [22], 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 [22].

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 achieves mAP of 44.76% and marginally outperforms the best model in [22], ResNet50+FPN with merged masks, 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%. High segmentation performance does not immediately translate into the equally strong advantage in classification, but overall the models that did best for the segmentation task also achieved the highest accuracy in COVID-19 sensitivity, overall accuracy and true negative rate.

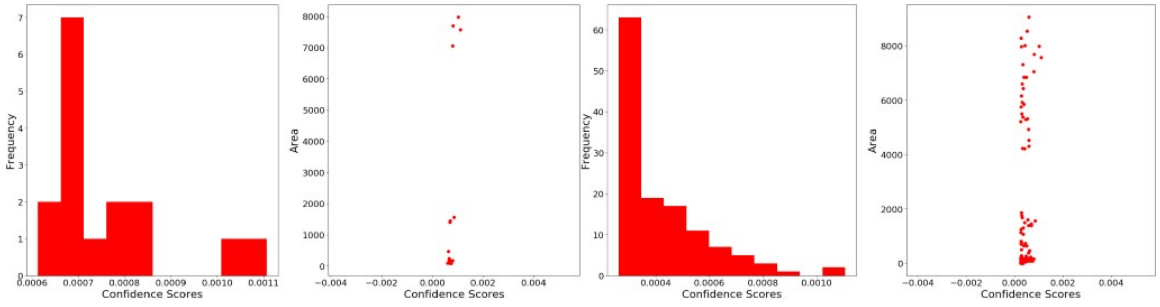
We experimented with a number of additional hacks for each model:



4(a) COVID-19



4(b) Common Pneumonia



4(c) Control (Negative)

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

1 Replacement of softmax with sigmoid activation function for the outputs of RoIs (segmentation model, test stage). Faster R-CNN implementation [28] 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 $\text{score}_{\theta} = 0.75$) 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.

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 [3] 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 and the first and the last ones are sampled for replication, the batch becomes [25,15,16]. 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 [3,29,30] for the detailed discussion of the distribution of lesions in chest CT scans. Training the seg- mentation 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.

4.1. 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 [26,14] or using specialized tools like GSInquire [15] 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 isolated areas (instances) 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 objects belonging to the same class. Mask R-CNN independently evolves separate instances of regional predictions that can overlap, both at bounding box and mask level.

This is illustrated in Figure 3 for the output of $ResNet34_{T1}$ model. Figure 3a is COVID-19 positive, Figure 3b is COVID-19 negative (Control, no lesions at slice level), 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 usually 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), but for the (combined) mask score map in column

2 in Figure 3, we used raw (before SIGMOID normalization) scores from Mask R-CNN. Each prediction is done by Mask R-CNN independently, i.e. the full path of extracting the RoI from the FPN layer using RoIAlign [25], predicting bounding box coordinates, filtering it through the de-convolution layer to obtain a fixed-size (28x28) mask score map with pixel logits that is then resized to the size of the bounding box prediction is done independently for each object. Looking at the combined mask score maps, it becomes clear how COVID-CT-Mask-Net learns to use the score information. Each 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 3, 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 $score_0$=0.01, we plotted only a small number of the highest-scoring RoIs to avoid image cluttering.

4.2. Distribution of observations in the RoI output batch

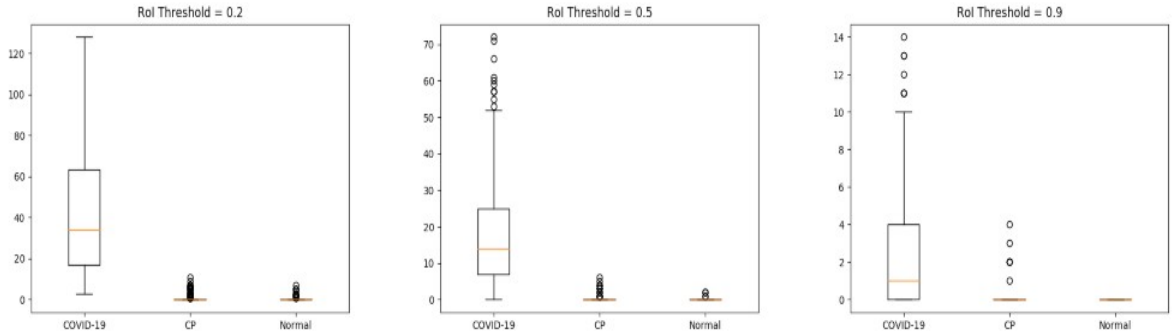
The analysis of the mask score maps in column 2, Figure 3 illustrates the effectiveness of the RoI batch to feature vector method, which is the main idea behind the transformation of Mask R-CNN into the classification model. Both the location (bounding box coordinates) and the importance (confidence score) of the areas critical to the COVID-19 diagnosis are output by RoI 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 RoIs' confidence scores, scan learn this ranking, and, eventually, associate a number of RoIs located in the critical areas (see [21,29] for the analysis of COVID-19 vs Common Pneumonia chest CT scans) with the particular image class.

To demonstrate this, we plot the histograms of the confidence scores and the scatterplots of the confidence scores vs RoI area (bounding box size) in three difference CT scan slices, one for each class in Figure 4. Top 16 regions (columns 1-2) in Figure 4a are dominated by several mid-size (1000 pixels) high-scoring (0.95) critical areas, and the full batch (128 regions) in Columns 3-4 follows what seems to be a Exponential distribution. Therefore, 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 4b: there's a small number of mid to large (2000-4000 pixels) low to mid scoring regions with the scores between 0.1 and 0.3, but the majority of RoIs have a score close to 0. The distribution of Control(Negative), Figure 4b is also distinct: the highest-scoring box (0.001) is very large (8000 pixels), and the rest of the batch have scores practically indistinguishable from 0 regardless of the size.

For a rigorous validation of the method, we also analyze the statistical distribution of the RoIs in the test sample (n=1425) extracted from the test split, in which all 3 classes are equally represented (475 slices per class). We consider 3 RoI $score_0$:0.2, 0.5, 0.9. Each image is fed through COVID-CT-Mask-Net, and at the batch construction stage we extract the number of RoIs with confidence score exceeding these thresholds. Boxplots in Figure 5 and Table 4 present the mean and 95% confidence intervals (CI) for each class/threshold and Table 5 presents the results of one-sided Kolmogorov-Smirnov test comparing each pair of distribution at significance level α=1%. This tests the hypothesis if the first distribution is less than the second and rejects it for p-values

less than α. The only result that is not statistically significant is CP vs Normal, which explains a larger number of confusions and requires further investigation.

Figure 5: *Boxplots of the number of RoIs for each class with scores>RoI score$_\theta$. Sample size:1425.*



The value of this result is that, contrary to [21,23], who showed that the differences in many COVID-19 and CP correlates are not statistically significant, the differences in the ranks of RoIs are mostly statistically significant across all 3 classes. Results in Figures 4 and 5 and Tables 4 and 5 were obtained with ResNet34$_{T\ 1}$+FPN.

Table 4: *Mean number of RoIs exceeding the threshold for each class + 0.95% CI.*

Class	RoI scoreθ=0.2	RoI scoreθ=0.5	RoI scoreθ=0.9
COVID-19	42.71 ± 61.06	17.87 ± 27.39	2.4 ± 5.55
CP	0.65 ± 3.15	0.15 ± 1.25	0.03 ± 0.30
Normal	0.11 ± 1.25	0.016 ± 0.32	0 ± 0

Table 5: *One-sided Kolmogorov-Smirnov Test Results. Score/p-value. The result not statistically significant at α=1% in bold.*

Class	RoI scoreθ=0.2	RoI scoreθ=0.5	RoI scoreθ=0.9
COVID-19/CP	0.96/1e-10	0.92/1e-10	0.63/1e-10
COVID-19/Normal	0.98/1e-10	0.94/1e-10	N/A
CP/Normal	0.18/2e-8	**0.067/0.091**	N/A

5. Conclusions

The main objective of this paper was the investigation of the ability of ResNet18/34-based models with deleted layers (lightweight truncated models) to train fast to segment lesions and predict COVID-19 from chest CT scans while maintaining the previously achieved levels of segmentation precision and classification accuracy.

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 (last block of ResNet34 deleted) achieved 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 (last block of ResNet18 deleted) with 6.35M parameters achieved the COVID-19 sensitivity of 91.35% and overall accuracy of 93.95%. The smallest model with $ResNet18_{T2}$+FPN backbone (last two blocks deleted) with just 4.25M parameters achieved a 84.05% COVID-19 sensitivity and 88.66% overall accuracy. Additionally, we presented an in-depth analysis of the effect of mask score maps and the distribution of the Regions of Interest in the batches sampled on the prediction outcome. Using one-sided KS test we showed that the difference in the scores of RoIs in the batches is statistically significant for all class pairs, except Common Pneumonia and Negative. Various statistical analyses showed that the distribution of RoIs are different across all three classes.

Conflict of interests: The authors declare that there is no conflict of interest.

References

1. Bai HX, Hsieh B, Xiong Z, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. Radiology. 2020;296:E46-E54.
2. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. 2016;770-8.
3. https://www.medrxiv.org/content/10.1101/2020.10.11.20211052v3
4. Lin T-Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR). 2017;2117-25.
5. https://www.medrxiv.org/content/10.1101/2020.04.24.20078584v3
6. https://www.medrxiv.org/content/10.1101/2020.02.23.20026930v1
7. Ning W, Lei S, Yang J, et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. Nat Biomed Eng. 2020;4:1197-207.
8. https://www.medrxiv.org/content/10.1101/2020.06.08.20121541v3
9. Zhang K, Liu X, Shen J, et al. Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography. Cell. 2020;181:1423-1433.e11.
10. https://arxiv.org/abs/2008.04815
11. Muhammad LJ, Islam MM, Usman SS, et al. Predictive Data mining models for novel coronavirus (COVID-19) infected patients recovery. Sn Comput Sci. 2020;1:206.
12. Asraf A, Islam MZ, Haque MR, et al. Deep learning applications to combat novel coronavirus (COVID-19) pandemic. Sn Comput Sci. 2020;1:363.
13. Rahman MM, Islam MM, Manik MMH, et al. Machine learning approaches for tackling novel coronavirus (COVID-19) pandemic. Health Policy and Technology. Elsevier. 2021.
14. Saha P, Sadi MS, Islam MM. EMCNet: Automated COVID-19 diagnosis from x-ray images using convolutional neural network and ensemble of machine learning classifiers. Inform Med Unlocked. 2021;22:100505.

15. Butt C, Gill J, Chun D, et al. Deep learning system to screen coronavirus disease 2019 pneumonia. Appl Intell. 2020;22:1-7.
16. https://arxiv.org/abs/2009.05383
17. Li L, Qin L, Xu Z, et al. Using artificial intelligence to detect covid-19 and community-acquired pneumonia based on pulmonary ct: evaluation of the diagnostic accuracy. Radiology. 2020;296:E65-E71.
18. Yan T, Wong PK, Ren H, et al. Automatic distinction between covid-19 and common pneumonia using multi-scale convolutional neural network on chest ct scans. Chaos Soliton Fract. 2020;140:110153.
19. https://www.medrxiv.org/content/10.1101/2020.03.19.20039354v1
20. https://arxiv.org/abs/2004.07054
21. https://arxiv.org/abs/2003.13865
22. https://www.medrxiv.org/content/10.1101/2020.10.23.20218461v4
23. Islam MZ, Islam MM, Asraf A, A combined deep CNN-LSTM network for the detection of novel coronavirus (covid-19) using x-ray images. Inform Med Unlocked. 2020;20:100412.
24. https://www.medrxiv.org/content/10.1101/2020.08.24.20181339v1
25. He K, Gkioxari G, Dollar P, et al. Mask R-CNN. in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy. 2017;2980-8.
26. https://arxiv.org/abs/2009.05096
27. Lin T-Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context. In European conference on computer vision. 2014;740-55.
28. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems. 2015;1:91-9.
29. Zhao D, Yao F, Wang L, et al. A comparative study on the clinical features of covid-19 pneumonia to other pneumonias. Clin Infect Dis. 2020;71:756-61.
30. Zhao W, Zhong Z, Xie X. Ct scans of patients with 2019 novel coronavirus (covid-19) pneumonia. Theranostics. 2020;10:4606-13.