COVID-19 Detection from Chest X-Ray Images Using Detectron2 and Faster R-CNN

Ammar Alhaj Ali, Roman Jasek, Bronislav Chramco, Rasin Katta, and Said Krayem

***Abstract*— The COVID-19 outbreak has been causing immense damage to global health and has put the world under tremendous pressure since early 2020. The World Health Organization (WHO) has declared in March 2020 the novel coronavirus outbreak as a global pandemic. Testing of infected patients and early recognition of positive cases is considered a critical step in the fight against COVID-19 to avoid further spreading of this epidemic. As there are no fast and accurate tools available till now for the detection of COVID-19 positive cases, the need for supporting diagnostic tools has increased. Any technological method that can provide rapid and accurate detection will be very useful to medical professionals. However, there are several methods to detect COVID-19 positive cases that are typically performed based on chest X-Ray images that contain relevant information about the COVID-19 virus. This paper goal is to introduce a Detectron2 and Faster R-CNN to diagnose COVID-19 automatically from X-ray images. In addition, this study could support non-radiologists with better localization of the disease by visual bounding box.**

***Index Terms*— COVID-19, Deep learning, Detectron2, Object Detection, Faster R-CNN.**

I. Introduction

COVID-19 is a respiratory disease that was recognized in December 2019 and has posed critical challenges for the public health, medical communities, and research [1]. It has spread very rapidly around the world since early 2020.According to World Health Organization (until 15 June 2021), 175,847,347 COVID-19 cases were detected worldwide. 3.807.276 of these cases resulted in death [2]. Most patients with COVID-19 suffer mild to moderate symptoms, approximately 15% progress to serious pneumonia and about 5% eventually develop acute respiratory distress syndrome and/or multiple organ failure [3]. Rapid diagnosis and isolation are the keys to the prevention of transmission [4]. Hospitals screen patients suspected of having COVID-19 to keep infected patients isolated from healthy population and health workers [5]. Medical imaging is considered a significant technique to assist doctors to evaluate the disease and control measures. Clinically, X-ray is the most commonly used imaging mean in the diagnostic, due to its speed, low radiation, and low cost. Chest X-ray is one of the most popular used imaging test to visualize and quantify the consequences of thoracic diseases and provide high-resolution pictures of disease progression [6]. The use of chest X-Ray to detect COVID-19 infection in the patients exhibiting symptoms has many advantages over traditional diagnostic tests:

1. X-Ray is much widespread and cost-effective than other diagnostic tests.
2. Transfer of digital X-Ray images is easy and does not require any extra transportation techniques from the acquisition stage to the analysis stage, thus makes the diagnostic process extremely quick and effective.
3. Portable X-Ray machines enable testing within an isolation ward itself, unlike CT Scans, thereby reducing the infection risk [7].
4. In addition, health experts noticed changes in X-Ray images before the COVID-19 symptoms are visible [8].

In the last times, computer vision has provided an interesting method for accurate diagnosis of many diseases such as skin cancer classification, breast cancer detection, brain disease classification, and pneumonia detection in the chest X-Ray [9]. Deep learning has been applying for the detection and classification of pneumonia and other diseases by radiography, as deep learning models are capable of finding patterns in X-ray images of the chest, even sometimes those patterns are imperceptible to the human eye [10]. The application of deep learning techniques coupled with X-ray images could be a good choice for accurate detection of COVID-19 disease, also it can be supportive in overcoming the case of radiologists shortage in remote regions.

Detectron2 is an open-source framework developed by Facebook's AI research group, it implements state-of-the-art object detection algorithms with an accurate and fast model for object detection, and it is completely built under the PyTorch platform. One of the key features of Detectron2 is its modular design [11]. Detetron2 includes a variety of models like Mask R-CNN, Faster R-CNN, Panoptic FPN, RetinaNet, DensePose, Cascade R-CNN, and TensorMask. It supports many different computer vision tasks including object detection, instance segmentation, panoptic segmentation and human pose prediction [12].

The contributions of this study can be summarized as follows:

1. Using deep learning to develop an automated and accurate screening technique using chest X-ray for early warning of possible outbreak caused by COVID-19.
2. providing detailed experimental analysis in terms of Average precision (AP) and area under the curve (AUC) to measure the performance of the proposed model.
3. Identifying and localizing COVID-19 abnormalities on chest X-ray images.

The rest of the paper is organized as follows:

Section 2 deep learning based object detection. Section 3 methodology. Section 4 results and discussion. Finally, Section 5 summarizes the study and draws a conclusion.

II. Deep Learning Based Object Detection

In recent years, deep learning-based object discovery has attracted a lot of attention due to its widespread usage and applications. Object detection is a branch of computer vision that involves the classification and localization of objects in an image and video [13]. Image classification refers to assigning a class label to an image, whereas image localization involves identifying the location of one or more objects in the image and drawing a bounding box around them, Object detection combines these two tasks (classification and localization) as it draws a bounding box around each object of interest within the image and assigns them a class label. Despite all breakthroughs in deep learning-based object detection in recent years, it is still under intensive investigation in both academical and industrial sectors, such as robotic vision, monitoring security, face detection, video surveillance, and autonomous driving [14]. Object detection is the extension of object classification; the essential goal of object detection is to detect all the predefined classes (instances) in the image and provides its localization by drawing a bounding box around it. Recent techniques for object detection are generally related to two prominent families the first one is Region-Based Convolutional Neural Networks (R-CNNs) and the second one is You Only Look Once (YOLO) [14] [15].

A. R-CNN

Ross Girshick et al. proposed R-CNNs approach [16], it uses the selective search algorithm [17] to extract just 2000 regions from the image that are called region proposals [16]. Object detection system with R-CNN consists of three modules:
1. Module 1: Region proposal which Generates category-independent region proposals (bounding boxes) using computer vision techniques, these proposals define the set of candidate detection that are available to the used detector.
2. Module 2: Feature extractor, this module Uses the convolutional neural network to extract a fixed feature vector from each region.
3. Module 3: The Classifier is responsible to Predict the classes of the proposed candidates using extracted features. It is a set of class specific linear support victor machines SVMs. See Figure 1.

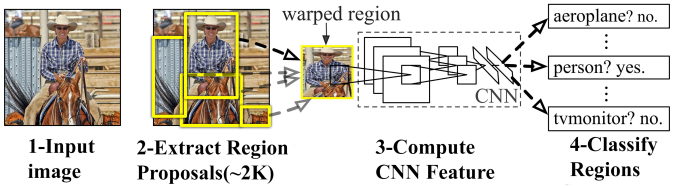

Fig. 1. R-CNN architecture.

R-CNNs take a huge amount of time to train the network because training is done in multiple stages. Beside to the training stage, the prediction stage also takes a long time and cannot be implemented in real-time as it takes almost 47 seconds for each image. The selective search algorithm is a fixed algorithm this could lead to the generation of bad candidate region proposals because there is no learning during region proposal phase. Ross Girshick [18] proposed another model to tackle these issues and draw-backs by offering a faster version of R-CNN which was called Fast R-CNN.

B. Fast R-CNN

A Fast R-CNN architecture takes an entire image and proposes candidate regions. The network first processes the whole image with several convolutional and max-pooling layers to extract a convolutional feature map. Then for each object proposal, it passes the region of interest (RoI) pooling layer to get a fixed-length feature vector from the feature map. Each feature vector is an input of the following classification and bounding box regression fully connected layers. The features are extracted from the entire input image once and are sent to CNN for classification and localization. See Figure 2. A major

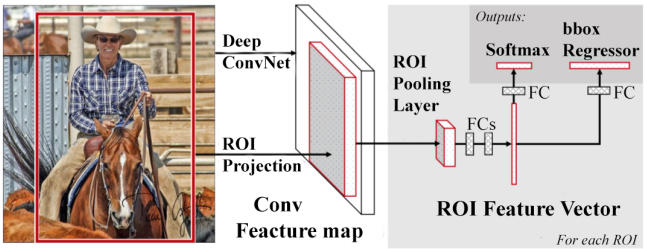

Fig. 2. Fast R-CNN architecture

improvement of Fast R-CNN is the model trains as a single model instead of three separate modules in R-CNN, another improvement of Fast R-CNN is that it uses a RoI pooling layer to extract a fixed-size feature map from the region proposals of different sizes [14]. The reason why Fast R-CNN is faster than R-CNN is that we do not have to feed 2,000 region proposals to the Convolutional Neural Network every time. Instead, the convolution only takes place once per image and a feature map is generated out of it.

Although Fast R-CNN improves the time of training and prediction, it still needs the region proposal as inputs, usually, the region proposals for each image are generated separately by using image processing techniques [15], these techniques are to be considered slow as they consume same running time as the detection network.

C. Faster R-CNN

Ren et al., [19] proposed Faster R-CNN to tackle drawbacks of Fast R-CNN. The essential improvement of Faster R-CNN is its ability to incorporate region proposals as a part of the model using Region Proposal Network (RPN). The architecture of final model contains two sub-networks, the first one is a Region Proposal Network (RPN) and the second one is the Fast R-CNN, both sub-networks are trained simultaneously for two different tasks the first task is region proposals and the

second task is bounding box classification and regression. See Figure 3.

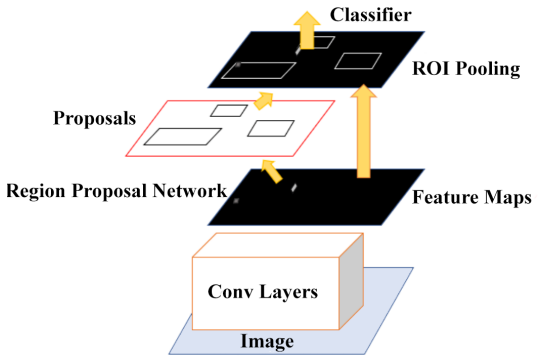

Fig. 3. Faster R-CNN architecture

D. YOLO

YOLO (you only look once) is another famous family of object detection, proposed by Redmon et al. after Faster R-CNN [20]. The main contribution of YOLO is real time detection. It has different versions such as YOLO, YOLOv2, YOLOv3, YOLOv4, and YOLOv5. YOLO versions might differ in architectures and techniques used. However, they contain a single neural network with the input as images, ground-truth boxes, and labels, the outputs of network are the bounding boxes and corresponding labels of the detected objects. YOLO pipeline first takes an image and split it into grids. Each grid has a dimension of n x n and is responsible to detect the objects that appear within it. While YOLO mainly advantage is speed the Faster R-CNN has better a accuracy.

III. Methodology

Our methodology starts with exploration of dataset to discover and understand the data. Then we move on by splitting the dataset into training and validation sets. The validation set enables us to evaluate our model. Regarding deep learning model architectures, we present some baseline models that are being used for object detection with detectron2. Then We proceed to explain how object detection models are evaluated. After that we move on with strategies to improve the base models, such as hyper-parameter tuning and data augmentations.

A. Dataset

The dataset in this study has obtained from the Kaggle competition, SIIM-FISABIO-RSNA COVID-19 Detection [21]. The Society for Imaging Informatics in Medicine (SIIM) is an international leading healthcare organization, its mission is to advance medical imaging informatics through education, research, and innovation [22]. The total number of chest X-ray image in the used dataset is 6666 of different sizes. There are four classes negative for pneumonia, typical, indeterminate, and atypical for COVID-19. Table1 shows classes chest X-rays findings of COVID-19 patients [5] [23].

Figure 4, Figure 5, and Figure 6 show samples of typical, indeterminate, and atypical classes respectively.

Figure 7. Distribution of all classes.

B. Baseline Models

In object detection, it is a popular practice where a pre-trained model is used as starting point on the object detection model, this technique is called transfer learning, where a model developed for a task is reused as starting point for another task [24] [25] [26]. Detectron2 provides many baseline models for Faster R-CNN [27] that are pre-trained on large dataset such as ImageNet [28], the baselines are based on three different backbone combinations:

1. FPN (Feature Pyramid Network) uses a ResNet and FPN backbone with standard convolution and FC (Fully connected) heads for mask and box prediction, respectively.
2. C4 use a ResNet conv4 backbone with conv5 head
3. DC5 (Dilated-C5) uses a ResNet conv5 backbone with dilations in conv5 and standard convolution and FC heads for mask and box prediction, respectively.

Table 2 shows the specifications of baseline models that we chose in our study. Box AP (Average precision) is the result of training the models on COCO 2017 dataset [29].

C. Evaluating Object Detection Models

In object detection, evaluation is not easy, because there are two separate tasks for the measurements of the classification and localization. AP (Average precision) is a popular metrics to evaluate object detection models like R-CNN and YOLO. AP is based on the precision-recall curve and is the precision averaged across all recall values between 0 and 1 at various Intersection over union IoU thresholds [30]. To digest AP, it is necessary first to offer a quick recap on precision, recall, and IoU. Recall refers to how many positives are correctly identified, in another words measures how good you find all the positives. The following equation is the mathematical definition of recall:

$$(Recall) = \frac{TP}{(TP + FN)}$$

Where TP is the true positives that are the number of correctly identified positive instances, and FN refers to the false negatives that are the number of incorrectly identified positive cases as negative.

Precision defines as the number of true positives divided by the sum of true positives and false positives numbers, in another words measures how accurate is the predictions. The following equation is the mathematical definition of precision:

$$Preicision = \frac{TP}{(TP + FP)}$$

Where FP is the number of false positives referring to the negative instances that are incorrectly classified as positive cases. IoU (Intersection Over union) is used to measure the overlap between two bounding boxes, also it is called Jaccard. We use it to measure how much the predicted bounding box intersects with the real object box (the ground truth). The following equation is the mathematical definition of IoU:

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

TABLE I
CATEGORIES OF CHEST X-RAY FINDINGS IN SUSPECTED COVID-19 PATIENTS.

Class Name	Rationale
Negative for pneumonia	No features of pneumonia
Typical appearance	Commonly reported imaging features of greater specificity for COVID-19 pneumonia
Indeterminate appearance	Nonspecific imaging features of COVID-19 pneumonia
Atypical appearance	Uncommonly or not reported features of COVID-19 pneumonia

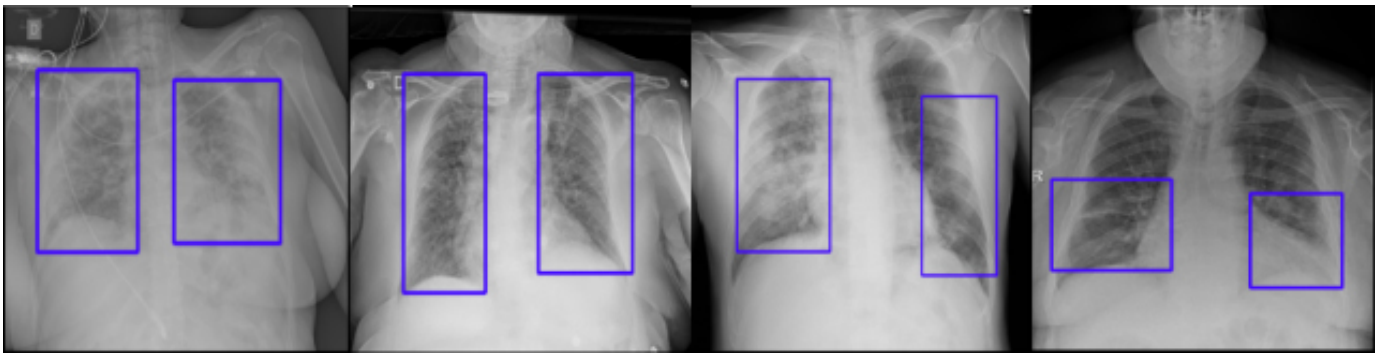

Fig. 4. Samples of the typical class

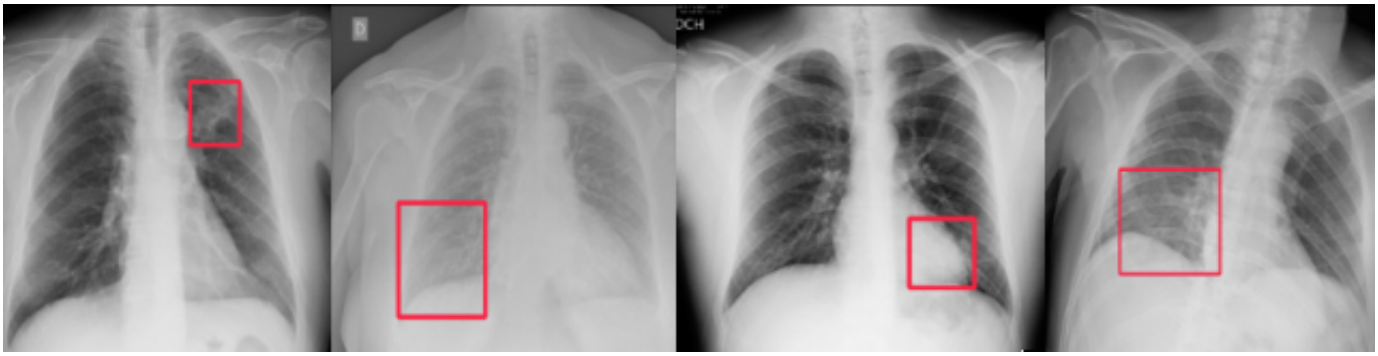

Fig. 5. Samples of the indeterminate class

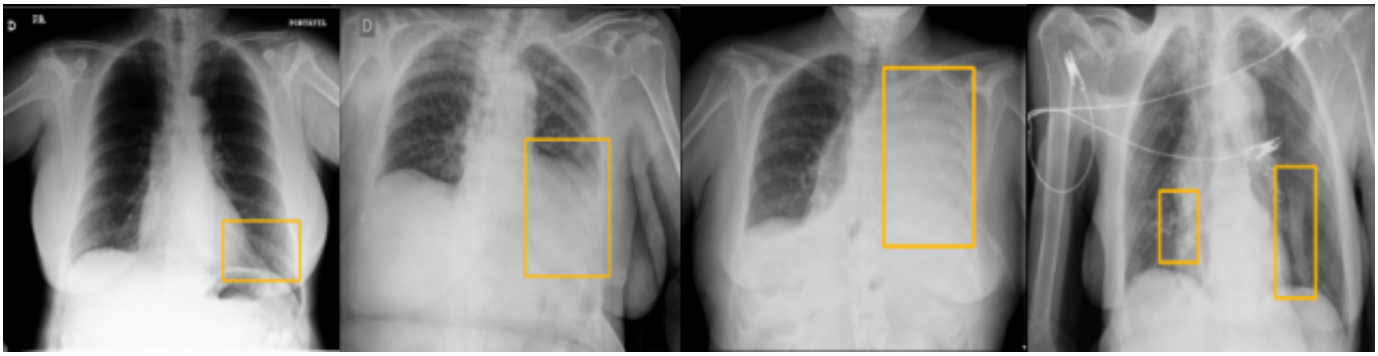

Fig. 6. Samples of the atypical class

TABLE II
SPECIFICATIONS OF SELECTED BASELINE MODELS.

Model Name	Train time (s/iter)	Inference time (s/im)	Train mem (GB)	Box AP
R101-C4	0.619	0.139	5.9	41.1
R101-DC5	0.452	0.086	6.1	40.6
X101-FPN	0.638	0.098	6.7	43.0

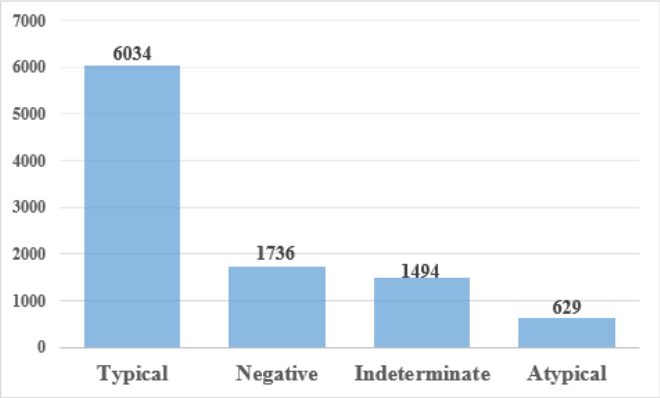

Fig. 7. Distribution of all classes.

Where A and B are the predicted bounding box and the ground-truth box. And ∩ , ∪ denote the intersection and union operator, respectively. IoU ratio (say 0.5) is used as a threshold for determining whether a predicted score is a True Positive (TP) or a False Positive (FP). Prediction is TP if the IoU of a predicted bounding box of class 'A' and the ground truth box of the same class 'A' is above or equal threshold of 0.5. Otherwise, it is FP. Practically, the TP and FP are sufficient to calculate precision and recall, as recall could be calculated directly using following formula:

$$(Recall) = \frac{TP}{len(GroundTruth)}$$

By setting the threshold at different levels, we get different pairs of precision and recall. We can draw a precision-recall curve with recall on the x-axis and precision on the y-axis, See Figure 8.

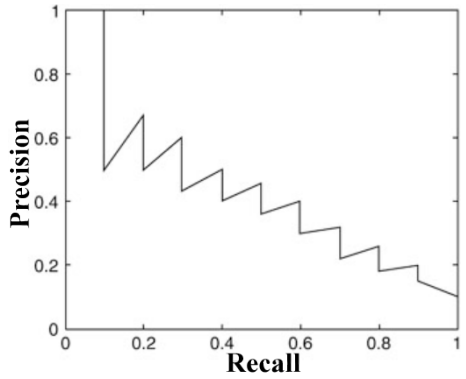

Fig. 8. Precision-recall curve.

the precision-recall curve shows the trade-off between the precision and recall values for different thresholds. It helps to select the best threshold to maximize both metrics.

AP summarizes the precision-recall curve into a single value representing the average of all precisions. It can be defined as the area under the precision-recall curve (AUC). Often, the precision-recall curve shows a zigzag pattern, before calculating AP we often smooth out the zigzag pattern first to reduce the impact of the wiggles in the curve. first, we interpolate the precision at multiple recall levels. The interpolated precision P_interp at a certain recall level r is defined as the highest precision found for any recall level $r1 \geq r$

$$P_{interp}(r) = maxP(_{r1}) \tag{1}$$

Then AP can be defined as the area under the interpolated precision-recall curve that can be calculated using the following formula:

$$AP = \sum_{i=1}^{0} (r_{(i+1)} - r_i) * p_{interp}(r_{(i+1)}) \tag{2}$$

Where n is number of recall levels and r_1, r_2 …, r_n is recall levels. Precision and recall fall within 0 and 1. Therefore, AP is always between 0 and 1. The mAP is average of the APs for all the classes, it is calculated according to the next equation

$$mAP = [\frac{1}{n}] \sum_{k=1}^{n} AP_k \tag{3}$$

Where AP_k is the AP of the class k, and n is the number of classes. it is important to note this, some papers use mAP and AP interchangeably.

D. Training Process for Different Models

In order to increase efficiency, in this study, we resized the images to 640x640 pixel,and split the dataset into two sets: a training set and a validation set with a ratio of 80:20 respectively. Both training and inference are carried out at same resolution. To improve the performance of models we use the Data Augmentation (DA) technique that increases the amount of data by adding slightly modified copies of already existing data. DA acts as a regularizer and helps to reduce overfitting. Used data augmentation techniques were: Brightness, vertical flip, and horizontal flip. We selected three pre-trained models that have good box AP for training. We tuned many hyper-parameters, and set image_per_batch = 2, Base_learning_rate = 0.0025, batch_size_per_image = 512, , batch_size_per_image = 512, num_workers = 4, and num_classes = 4. All other configurations are kept as default values.

IV. Results and Discussion

The training of the models was done in 20000 and 30000 iterations. We chose three ways to calculate AP on our models, the AP(0.50:0.95) value is the average AP for IoU from 0.5 to 0.95, with a step size of 0.05. AP50 is the AP with IoU = 0.5, and AP75 is the AP when IoU = 0.75. The results are shown in the Table 3 for all three trained models, the AP values have slight variations with the increase in training iterations from 20,000 to 30,000. When the IoU threshold becomes larger, the AP value drops down dramatically for all models.

When using the faster R-CNN network there are four losses, two losses at RPN and two at fast R-CNN network. The losses for the Region Proposal Network are loss_rpn_loc and loss_rpn_cls. Loss_rpn_loc is localization loss or the loss of the bounding box regressor for the RPN. Loss_rpn_cls is the loss of the classifier that classifies if a bounding box is an interesting object or background. See Figure 9. The losses

TABLE III
THE COMPARISON OF RESULTS WITH DIFFERENT MODELS AND ITERATIONS.

No.	Model	iterations	AP(0.50:0.95)	AP50	AP75
1	X101-FPN	20000	22.46	31.39	19.43
		30000	23.68	32.45	20.78
2	R101-DC5	20000	22.56	31.30	19.77
		30000	22.58	30.18	20.29
3	R101-C4	20000	21.54	30.02	18.94
		30000	22.17	30.25	20.14

for the fast R-CNN network are loss_classes and loss_box_reg. Loss_classes is loss for the classification of detected objects into various classes. Loss_box_reg is localization loss or the loss of the bounding box regressor. See Figure 10.

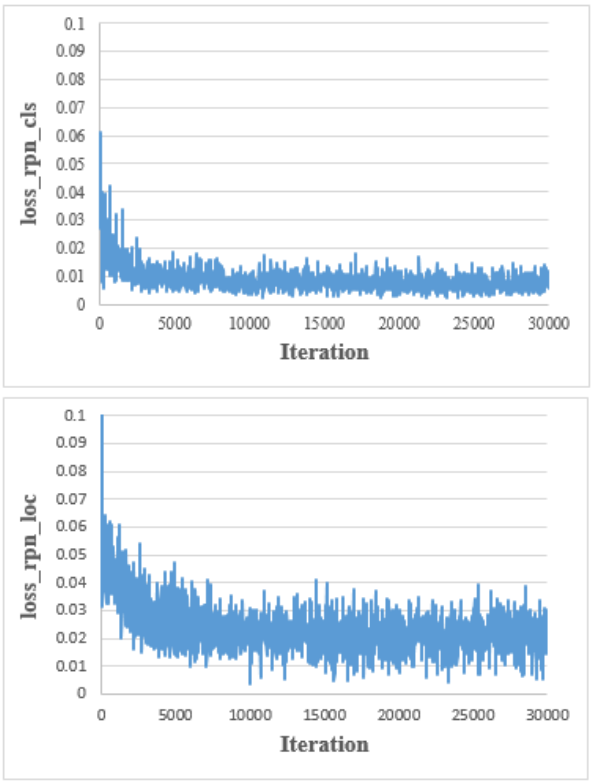

Fig. 9. Losses for Regional Proposal Network.

Figure 11 shows the ground truth and the prediction of Chest X-ray images for COVID-19 patients taken from the validation data set. And Figure 12 shows the Chest X-ray images of negative cases taken from the validation data set were the prediction also produce negative results.

V. CONCLUSIONS

As discussed in this paper, the early detection and diagnosis of COVID-19 using automated low cost methods are considered an essential step in preventing the disease spread and the progression of the pandemic. In this study, we exploit Detectron2 with faster R-CNN for detecting COVID-19 through chest radiography images. Experiments were conducted to evaluate some baseline models in Detectron2 on

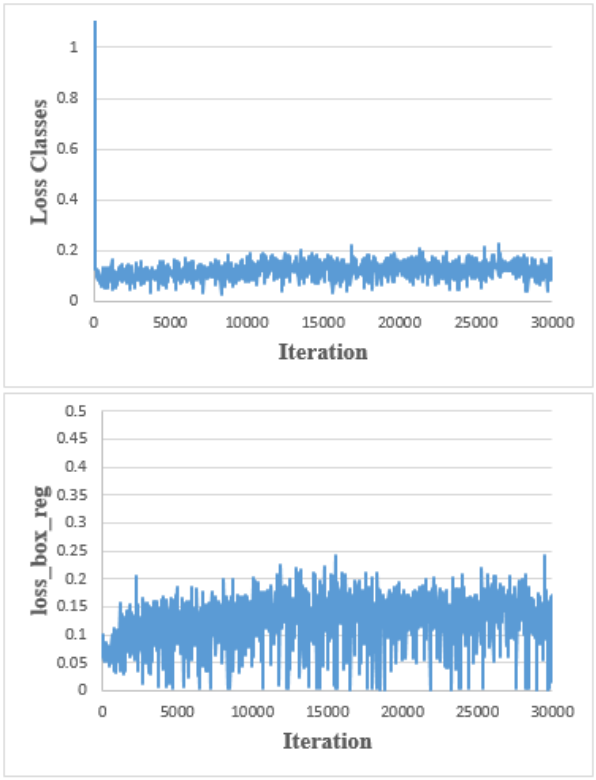

Fig. 10. Losses for the fast R-CNN.

the COVID-19 dataset, using three different backbones R101-C4, R101-DC5, and X101-FPN. The use of these techniques in the diagnostic of COVID-19 can be a powerful tool for radiologists to avoid human error and can assist them to make correct decisions in critical conditions. This research supports the idea that deep learning is paving the way for optimizing healthcare systems and improving the results of diagnostic procedures. Despite deep learning has been become one of the most important and powerful computing tools in the diagnosis of COVID-19, but developers must be so careful to avoid overfitting of COVID-19 diagnostic models, and as COVID-19 is considered relatively a new disease, these models must be trained on large, heterogeneous datasets to cover all kinds of chest x-ray images for patients infected by different kinds of pneumonia that are related with COVID-19.

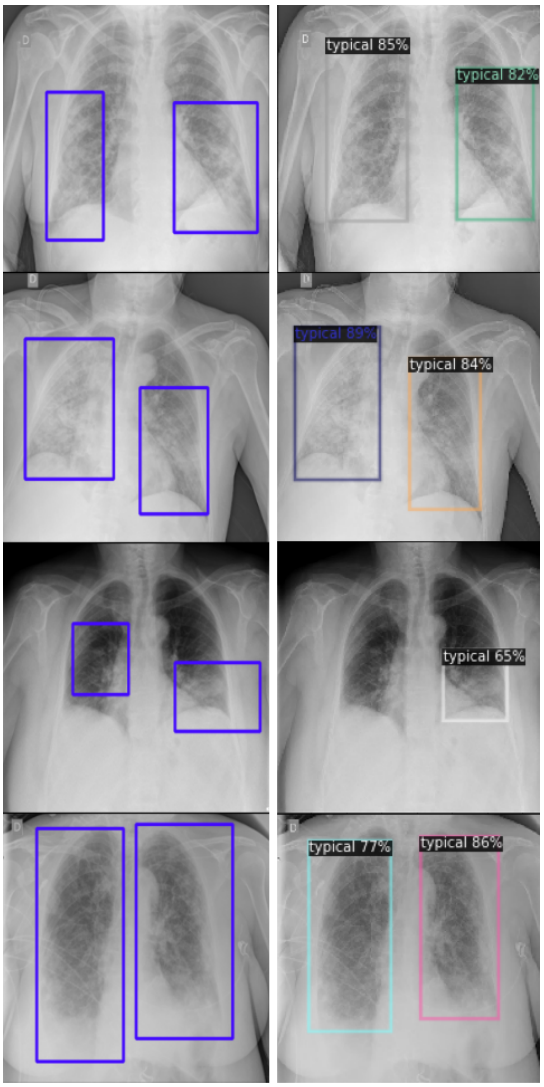

Fig. 11. Chest X-ray images of COVID-19 patients.(a) ground-truth.(b) Prediction

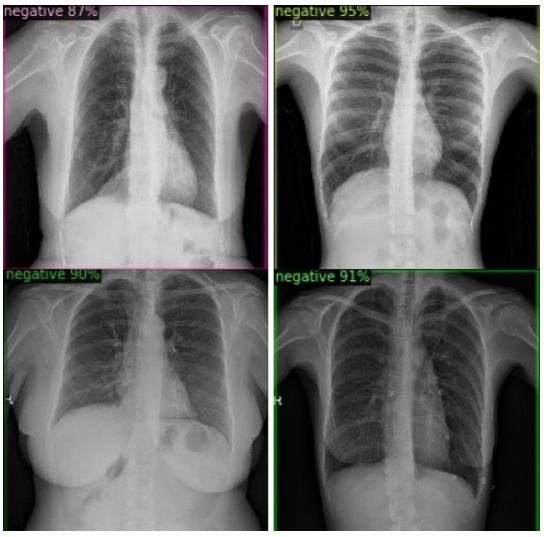

Fig. 12. Chest X-ray images of negative cases.

REFERENCES

[1] A. S. Fauci, H. C. Lane, and R. R. Redfield, "Covid-19 — navigating the unchart-ed," N. Engl. J. Med., vol. 382, no. 13, pp. 1268–1269, 2020.

[2] "Coronavirus disease (COVID-19) – World Health Organization," Who.int. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019. [Ac-cessed: 05-Dec-2021].

[3] X. Cao, "COVID-19: immunopathology and its implications for ther-apy," Nat. Rev. Immunol., vol. 20, no. 5, pp. 269–270, 2020.

[4] K. R. Peck, "Early diagnosis and rapid isolation: response to COVID-19 outbreak in Korea," Clin. Microbiol. Infect., vol. 26, no. 7, pp. 805–807, 2020.

[5] T. M. H. de Jaegere, J. Krdzalic, B. A. C. M. Fasen, R. M. Kwee, and COVID-19 CT Investigators South-East Netherlands (CISEN) study group, "Radiological Society of North America chest CT classification system for reporting COVID-19 pneumonia: In-terobserver variability and correlation with reverse-transcription polymerase chain reac-tion," Radiol Cardiothorac Imaging, vol. 2, no. 3, p. e200213, 2020.

[6] J. Zhang et al., "Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection," IEEE Trans. Med. Imaging, vol. 40, no. 3, pp. 879–890, 2021.

[7] A. Mangal et al., "CovidAID: COVID-19 detection using chest X-Ray," arXiv [eess.IV], 2020.

[8] K. H. Shibly, S. K. Dey, M. T.-U. Islam, and M. M. Rahman, "COVID faster R-CNN: A novel framework to Diagnose Novel Coronavirus Disease (COVID-19) in X-Ray images," Inform. Med. Unlocked, vol. 20, no. 100405, p. 100405, 2020.

[9] A. K. Das, S. Kalam, C. Kumar, and D. Sinha, "TLCoV- An automated Covid-19 screening model using Transfer Learning from chest X-ray images," Chaos Solitons Fractals, vol. 144, no. 110713, p. 110713, 2021.

[10] E. Luz, P. L. Silva, R. Silva, L. Silva, G. Moreira, and D. Menotti, "Towards an effec-tive and efficient deep learning model for COVID-19 patterns detection in X-ray imag-es," arXiv [eess.IV], 2020.

[11] R. Divya and J. D. Peter, "Smart healthcare system-a brain-like comput-ing approach for analyzing the performance of detectron2 and PoseNet models for anomalous action detection in aged people with movement impairments," Complex intell. syst., 2021

[12] U. S. Ahmed, "Training an Object Detection Model in a few min-utes using Detectron2," Red Buffer, 29-Mar-2021. [Online]. Avail-able: https://medium.com/red-buffer/training-an-object-detection-model-in-a-few-minutes-using-detectron2-5bd0aa5550d4. [Accessed: 05-Dec-2021].

[13] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," arXiv [cs.CV], 2021.

[14] L. Jiao et al., "A survey of deep learning-based object detection," IEEE Access, vol. 7, pp. 128837–128868, 2019.

[15] V. Pham, C. Pham, and T. Dang, "Road damage detection and classifi-cation with De-tectron2 and faster R-CNN," in 2020 IEEE International Conference on Big Data (Big Data), 2020.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[17] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selec-tive search for object recognition," Int. J. Comput. Vis., vol. 104, no. 2, pp. 154–171, 2013.

[18] R. Girshick, "Fast R-CNN," in 2015 IEEE International Conference on Computer Vision (ICCV), 2015.

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," arXiv [cs.CV], 2015.

[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, re-al-time object detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[21] "SIIM-FISABIO-RSNA COVID-19 Detection," Kaggle.com. [Online]. Available: https://www.kaggle.com/c/siim-covid19-detection/overview. [Accessed: 05-Dec-2021].

[22] "SIIM," Siim.org. [Online]. Available: https://siim.org. [Accessed: 05-Dec-2021].

[23] E. Martínez Chamorro, A. Díez Tascón, L. Ibáñez Sanz, S. Ossaba Vélez, and S. Borruel Nacenta, "Diagnóstico radiológico del paciente con COVID-19," Radiol. (Engl. Ed.), vol. 63, no. 1, pp. 56–73, 2021.

[24] A. A. Ali, B. Chramcov, R. Jasek, R. Katta, and S. Krayem, "Classifica-tion of plant diseases using convolutional neural networks," in Artificial

Intelligence in Intelligent Sys-tems, Cham: Springer International Publishing, 2021, pp. 268–275.
[25] A. A. Ali, B. Chramcov, R. Jasek, R. Katta, S. Krayem, and E. Awwama, "Tomato leaf diseases detection using deep learning," in Lecture Notes in Networks and Systems, Cham: Springer International Publishing, 2021, pp. 199–208.
[26] A. A. Ali, B. Chramcov, R. Jasek, R. Katta, S. Krayem, and M. Kadi, "Detection of steel surface defects using U-net with pre-trained encoder," in Software Engineering Application in Informatics, Cham: Springer International Publishing, 2021, pp. 185–196.
[27] "Detectron2 Model Zoo and Baselines." [Online]. Available: https://github.com/facebookresearch/detectron2/blob/main/MODEL-ZOO.md. [Ac-cessed: 05-Dec-2021].
[28] "ImageNet," Image-net.org. [Online]. Available: https://www.image-net.org/. [Ac-cessed: 05-Dec-2021].
[29] "COCO - common objects in context," Cocodataset.org. [Online]. Available: https://cocodataset.org/. [Accessed: 05-Dec-2021].
[30] N. Zeng, "An introduction to evaluation metrics for object detection," Zenggyu.com, 16-Dec-2018. [Online]. Available: https://blog.zenggyu.com/en/post/2018-12-16/an-introduction-to-evaluation-metrics-for-object-detection/. [Accessed: 05-Dec-2021].

Ammar Alhaj Ali Ammar received Ph.D. from Tomas Bata University in Zlin, and Bachelor's degree in computer engineering from Aleppo University. He has 19 years of experience as a software engineer,but last few years, He has shifted towards Artificial Intelligence and Deep learning, his research interests are in the fields of deep learning and computer vision.

Roman Jasek Roman is Professor of System Engineering and Computer Science of the Department of Informatics and Artificial Intelligence at Tomas Bata University in Zlín. It deals with the protection of processes and data and the use of artificial intelligence in cybersecurity. He is the director and expert guarantor of the penetration testing laboratory and the head of the research teams.

Bronislav Chramcov Bronislav studied Automatization and Control Technology at the Faculty of Technology in Zlin of the University of Technology in Brno and he took his degree in 1998. In 2006 he graduated his doctoral degree from the Faculty of Applied Informatics of Tomas Bata University in Zlin. He is working now as a associate professor at the Faculty of Applied Informatics of Tomas Bata University in Zlin. He is Vice-Dean for research, development and doctoral study. His research activities are focused on discrete event systems simulation.

Rasin Katta Rasin is a Ph.D. student at Tomas Bata University in Zlin, he holds MBA in Green Energy and sustainable business from Bologna Business School in Italy, and a Bachelor's degree in computer engineering from, Aleppo University. He has 25 years of experience in different industries on C-Level and has in-depth knowledge and experience in Quality. Last year, He has started to explore Machine Learning and Deep learning, his research interests are in the fields of deep learning applications in medical, farming and industry.

Said Krayem Said is a computer engineering professor. He is currently a close associate of the Department of Computer Science and Artificial Intelligence at Tomas Bata University in Zlín. It deals with fault tolerance in cybernetic systems. For a long time, he was a collaborator of the Academy of Sciences of the Czech Republic.