Toward Data-Model-Agnostic Autonomous Machine-Generated Data Labeling and Annotation Platform: COVID-19 Autoannotation Use Case

Ajay Arunachalam, Vinayakumar Ravi, Vasundhara Acharya, and Tuan D. Pham

***Abstract*—Quick, early, and precise detection is important for diagnosis to control the spread of COVID-19 infection. Artificial Intelligence (AI) technology could certainly be used as a modulating tool to ease the detection, and help with the preventive steps further. Convolutional neural networks (CNNs) have achieved state-of-the-art performance in many visual recognition tasks. Nevertheless, most of these state-of-the-art networks highly rely on the availability of a high amount of labeled data, being an essential step in supervised machine learning tasks. Conventionally, this manual, mundane, and time-consuming process of annotating images is done by humans. Learning to localize or detect COVID-19 infection masks in our specific case study typically requires the collection of CT scan data that has been labeled with bounding boxes or similar annotations, which generally is limited. A technique that could perform such learning with much less annotations, and transfer the learned proposals that are algorithm-driven to generate more synthetic annotated samples would be helpful & quite valuable. We present such a technique inspired by weakly trained mask region based convolutional neural networks (R-CNN) architecture for localization, in which the number of images with their pixel-level masks can be a small proportion of the total dataset, and then further improvise CNNs by inversely generating dense annotations on-the-go using an algorithmic-based computational approach. We focus on alleviating the bottleneck associated with deep learning models needing annotated data for training in an intuitive reverse engineering fashion through this work. Our proposed solution can certainly provide the prospect of automated labeling on-the-fly, thereby reducing much of the manual work. As a result, one can quickly train a precise COVID-19 infection detector with the leverage of autonomous frame-by-frame machine generated annotations. The model achieved mean precision accuracy (%) of 0.99, 0.931, and 0.8 for train, validation, and test set, respectively. The results demonstrate that the proposed method can be adopted in a clinical setting for assisting radiologists, and also our fully autonomous approach can be generalized to any detection/recognition tasks at ease.**

***Index Terms*—Artificial intelligence (AI), assisted annotations, annotations, autoannotation pipeline, autolabeling, CT scan, machine labels, mask region based convolutional neural networks (R-CNN), synthetic annotations.**

Manuscript received 6 April 2021; revised 26 May 2021; accepted 30 June 2021. Date of publication 28 July 2021; date of current version 16 June 2023. Review of this manuscript was arranged by Department Editor N. Gerdsri. *(Corresponding author: Vinayakumar Ravi.)*

Ajay Arunachalam is with Centre for Applied Autonomous Sensor Systems (AASS), Örebro University, 70281 Örebro, Sweden (e-mail: ajay.arunachalam@oru.se).

Vinayakumar Ravi and Tuan D. Pham are with the Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar 34754, Saudi Arabia (e-mail: vravi@pmu.edu.sa; tpham@pmu.edu.sa).

Vasundhara Acharya is with the Manipal Institute of Technology (MIT), Manipal Academy of Higher Education (MAHE), Manipal 576104, India (e-mail: vasundhara.acharya93@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TEM.2021.3094544.

Digital Object Identifier 10.1109/TEM.2021.3094544

I. Introduction

THE COVID-19 pandemic has had an irreversible effect on the world. In December 2019, an outbreak of a novel coronavirus emerged in Wuhan city in the Hubei province of China, and it evolved into a pandemic. Since, China reported its first case to the World Health Organization (WHO) in December 2019, thereafter 219 countries and territories have reported 2.5 million deaths, and about 113.4 million COVI9 cases [1]. It occurred due to infection from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and led to severe multisystem disease. The primary tool used for the diagnosis was the reverse transcription polymerase chain reaction (RT-PCR) [2]. Antigen tests are also being used, and developed but are not yet widely available and currently have variable reliability.

The medical industry has been looking for new technologies to control and monitor this global crisis. It has forced governments worldwide to accelerate the development and implementation of AI in four areas: Therapeutics, diagnosis, clinical decision-making, and public health. It has also encouraged many clinicians and policymakers to adopt AI in the field of medicine. Medical imaging techniques have proved to be vital in diagnosing diseases. Several works have been done to study the effect of social distancing to overcome the spread of the pandemic [3], [4]. Many governments have made it mandatory for people to wear masks in public to curb the spread of the disease. One of the studies [5] suggests that masks' impact is minimal when 50% or fewer people choose to wear face masks, and is significant when at least 80% of a population wears face masks. Identifying whether a distance is safe (complies with social distancing norms) and detecting masks has become a new area in the field of computer vision. Recent advancements advocate the need for intelligent systems to detect and capture human activities. These works involve tracking to segmentation, crowd monitoring, and GPS-based patient localization. There exist two stages to detect objects in motion: Object detection and object classification. Significant advancement has been made in object recognition and detection with the recent development of deep learning, especially deep CNNs. CNNs have shown excellent performance in image classification, object detection, and medical image analysis with their ability to obtain high accuracy and precise results with very limited preprocessing of

0018-9391 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

the input data. Object detection using CNNs can be classified into two categories: Single detectors and multistage detectors. In a multistage detector, the detection process is split into multiple steps, and in the case of a single-stage detector, the detection takes place in a single step. Several research works have been done to detect facial masks using multistage detectors [6], [7] and single-stage detectors [8], [9]. CNNs have also been employed to monitor the social distancing to curb the spread of COVID-19 [10], [11].

Due to the skepticism toward the sensitivity of RT-PCR test, medical image segmentation is highly desired as clinical decision support for quantitative assessment, and disease monitoring. Image segmentation tackles the problem of pixel-level predictions. Semantic segmentation classifies the semantic label for each pixel on a natural image [34]. Instance segmentation utilizes a pixel-level mask and a unique instance label to discriminate each semantic instance [35]. Radiology experts found that lung CT imaging can be used as an early diagnostic criteria as manifestations of COVID-19 cases as seen through CT imaging differed from those of the CT imaging manifestations of other viral pneumonia cases, such as influenza-A viral pneumonia (IAVP). They found that detecting and quantifying the infected lung regions would be quite advantageous in diagnosing COVID-19 [36]. Semantically segmenting CT scan images of COVID-19 patients was a crucial goal because it would help quantify the severity of illness and assist in disease diagnosis. But, as the number of suspected cases of COVID-19 increased, manual segmentation of CT images faced a serious challenge. It resulted in urgent requirements for accurate and automatic segmentation of COVID-19 infections. CNNs have shown a significant performance for many automatic image segmentation tasks. The works that employed CNN to segment the CT images to diagnose the disease are discussed under the related works section.

The application of machine learning (ML) ranges from image creation and reconstruction to diagnosis and outcome prediction. Medical image annotations result in significant accuracy in the early diagnosis and detection of the disease and understanding the noninfected normal cases. But, the biggest challenge is the availability of the data with annotations (labels). Sufficient annotated medical imaging data are mainly required to train accurate models. Medical imaging annotations have to be performed by clinical specialists, which is expensive, subjective, and labor-intensive. Large annotated medical imaging datasets for COVID-19 are currently unavailable publicly [37], [38]. This data's unavailability inhibits the development and application of medical image analysis to fight this global crisis.

The key motivation of the proposed idea is to deal with the lack of labeled training data especially observed in critical areas like healthcare. Largely, most of the state-of-the-art networks require large amount of labeled data as a prerequisite toward possibly minimizing the generalization error. Traditionally, the process of labeling data is done manually by humans. Through our algorithmic approach, we aim to alleviate the bottleneck of requiring high volumes of labeled data that is mainly associated with the deep learning models by generating synthetic machine-labeled data. Our proposed solution is targeted to provide the prospect of cheap labeled data instantly thereby minimizing the majority of the manual, mundane, and time-consuming human work. We also focus on providing a solution that is generic for any detection/recognition/localization/any other user-specific tasks that can be carried out at ease. There are tons of data annotation tools such as VGG Image Annotator (VIA), LabelMe, etc., and several collaborative platforms/services to help with the annotation process through human inputs manually. Also, there are many commercial platforms that focus on annotation, but none focuses on benchmarked-based machine-generated data validation and evaluations, and further, there is no robust mechanism to minimize generalization error leading to high-quality training data for any custom input data. The proposed approach provides intuitive facility one-step closer to tackle the real-world problem of scarcity of labeled training data.

The rest of the article is organized as follows. Related work on automatic labeling, and the summarization of the different ML and DL works done in general for COVID-19 research are discussed in Section II. The proposed autonomous autoannotation framework is described in Section III. The experiments and results are summarized in Section IV. A detailed discussion is provided in Section V. Finally, Section VI concludes this article.

II. Related Works

The various CT image datasets used to diagnose COVID-19 are tabulated in Table I. The COVID-19 positive cases are indicated as positive, healthy cases are indicated as normal, and pneumonia cases are indicated as pneumonia. AI and ML techniques have played a vital role in the fight against COVID-19 on multiple fronts. COVID-19 diagnosis by applying ML to CT scans and X-Rays has been proven to be extremely useful [41]–[49]. Availability of various sources of medical images such as X-Rays, CT images, and MRI empowers the ability of deep learning techniques to fight COVID-19 [49]. It has also been utilized in the analysis of COVID-19 case reports to identify earlier clinical symptoms [50], [51] and to analyze the social media to obtain sentiment and socioeconomic status in the current pandemic [52], [53]. Gozes *et al.* [41] used the U-Net model to segment the lung region from the CT images, and trained the ResNet model to classify the cases into normal and abnormal cases. Even though the model achieved an area under the curve score of 0.996, it used a manually labeled dataset that limited the training sample size. Li *et al.* [42] developed COVNet that had the ability to extract 2D and 3D features. U-Net was used for the segmentation of the lung regions. Even though the dataset is large, it is not publicly available, and it is not possible to determine what imaging features led the model to distinguish between COVID-19 and community-acquired pneumonia. Xu *et al.* [19] used a 3D CNN segmentation model to segment multiple candidate image cubes. The images were later classified into three types: COVID-19, IAVP, and irrelevant to infection. The main downside of their approach was that the manifestations of COVID-19 had some overlap with the manifestations of other pneumonia cases such as IAVP and other types of pneumonia. Ying *et al.* [43] designed a details relation extraction neural network with the feature pyramid network (FPN) to extract the top-K details in the CT images. An attention module was added to explore the importance of each detail. The model achieved high performance when compared to other pretrained models.

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COVID-19:CT SCAN DATASETS

Sr.No	Dataset	Number of images	Reference
1	COVID-19 Radiography Database	Positive:˙2191, Normal: 341 ,Pneumonia:1345	[12]
2	COVID-19 Chest X-Ray Dataset Initiative	Positive:55	[13]
3	SARS-CoV-2 CT-scan	Positive:1252 ,Normal:1230	[14]
4	COVID-CT-Dataset	Positive: 349,Normal: 397	[15]
5	COVID-Ctset	Positive : 15589, Normal: 48260	[16]
6	Pulmonary CT	Positive:1292 , Normal : 1325, Pneumonia:1735	[17]
7	COVID-19 Dataset	Positive :5427 , Normal:2628	[18]
8	CT images Dataset	Positive:219, Normal :175, Pneumonia:224	[19]
9	COVID-CS	Positive:195, Normal:258	[20]
10	COVID-19 CT segmentation dataset	Positive:100	[21]
11	Italian Dataset˙	Positive:100	[22]
12	Chest CT Dataset	Positive:72	[23]
13	MIDRC-RICORD dataset	Positive: 120, Normal :120	[24]
14	COVID-19 Dataset	Positive:924, Pneumonia:342	[25]
15	Chest CT Dataset	Positive:4400, Pneumonia: 9729	[26]
16	CT Dataset	Positive: 521 , Normal : 397, Pneumonia : 76	[27]
17	CT scans Dataset	Positive : 1262, Normal : 1230	[28]
18	COVID-19 chest CT images Dataset	Positive: 496, Normal : 260	[29]
19	Chest CT Dataset	Positive:230, Normal:130, Pneumonia:100	[30]
20	Chest CT Dataset	Positive:1495, Pneumonia:1027	[31]
21	Chest CT Dataset	Positive:1194, Normal :2239, Pneumonia:2239	[32]
22	Pulmonary CT Dataset	Positive:3854, Normal: 8566, Pneumonia:6871	[33]

Shan *et al.* [44] introduced a VB-Net-based deep learning model and human in the loop (HITL) that assisted the radiologists in the refinement of annotations. The HITL strategy reduced the manual delineation time. But, the datasets were collected from a single center which is not representative of COVID-19 patients in other areas, and the model is not able to quantify pneumonia. Yang *et al.* [15] created an open-source CT images dataset and utilized the DenseNet-169 model to classify the cases to COVID-19 and non-COVID-19. However, the size of the dataset is not enough to train a robust model. Anwar *et al.* [45] proposed EfficientNet deep learning architecture to detect COVID-19 from chest CT images. Even though their model achieved an accuracy of 89.7%, they did not use a separate validation set. Zheng *et al.* [46] developed a weakly supervised deep learning-based model to detect COVID-19 from 3D CT images. The model utilized U-Net to achieve the segmentation of the lung region. The model achieved an accuracy of 90.1%, but it was trained using imperfect ground-truth masks, and data were obtained from a single hospital. Mohammed *et al.* [47] developed ResNet+ that required only volume level data labels. Slice level prediction without additional supervision was achieved with the slice attention module. The model achieved a precision of 81.9, but the dataset had fewer COVID-19 positive cases when compared to negative cases, and did not include common and other viral pneumonia cases. Wang *et al.* [48] created a model by integrating two ResNet-based branches into one model framework. More discriminative representations were learned by transferring hierarchical attention information from the lesion region detection branch to the COVID-19 classification branch. The model achieved an accuracy of 93.3%, but the proposed classification model failed to diagnose CT scans with COVID-19 lesions at an early stage and misclassified normal scans as pneumonia. Chen *et al.* [54] employed UNet++ to extract valid areas and detect suspicious lesions in CT images of COVID-19 patients. The model achieved an accuracy of 95.24%. Gunraj *et al.* [55] designed COVIDNet-CT with machine driven approach to diagnose COVID-19 from CT images. The model achieved an accuracy of 99.1%. Silva *et al.* [56] proposed EfficientCovidNet with a voting-based approach, and a cross-dataset analysis to detect COVID-19. The model achieved an accuracy of 87.68%, but the cross-data analysis proved that the model's generalizability power was not suitable for its deployment in clinical settings. Bekhet *et al.* [57] proposed a fully automated CNN to diagnose COVID-19 from chest X-Ray images. The model attained an accuracy of 96%, but the comparison of the approach with other models was not demonstrated.

The major challenge faced by the deep learning techniques in analyzing the COVID-19 CT image analysis is the lack of publicly available datasets with well-defined labels. Various popular image annotation tools are available online to manually define regions in an image and create text-based descriptions of those regions [58]. Image annotation by utilizing features based on deep learning has shown better performance when compared to conventional visual features [59]. In the past few years, several deep learning based automatic image annotation methods have been developed [60]–[63]. Due to the COVID-19 pandemic, healthline frontiers are working at multiple fronts. Researchers working in medical image analysis are trying to assist them by developing algorithms for automatically scoring the pulmonary involvement in COVID-19 pneumonia. Liang *et al.* [39] utilized pseudocoloring methods to assist the physicians in grouping various tissues and isolating the relevant tissues. They also created a platform to annotate X-Ray and CT images to train the CNN. Zaffino *et al.* [40] developed a model to achieve the tissues' automated labeling from CT scans by employing a five-component Gaussian mixture model (GMM). The most probable component with a maximum *a posteriori* probability (MAP) estimate was then used to label the lung voxels. The correlation with clinical scores was used to measure the robustness of the model. Li *et al.* [17] visualized the vital regions in the pulmonary CT with gradient-weighted class activation mapping (Grad-CAM) method [64]. This mapping highlighted abnormal regions without the need for manual annotation. Analysis of closely related works are given in Tables II and III.

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

TABLE II
COMPARISON OF CLOSELY RELATED ANNOTATION WORKS TARGETED TOWARD COVID DIAGNOSIS

Ref	Method	Type	Model	Dataset Evaluated	Data Preprocessing	Labeled Data type	Manual Annotation
[17]	The U-Net-based segmentation method is used to extract the key regions of interest,followed by a deep neural network for prediction. The vital regions are highlighted with GRAD-CAM generating corresponding attention heatmaps.	Supervised	RestNet50 U-Net	custom collected 3D chest CT scans $>$ 1200 samples	Data augmentation	Segmentation mask generated by U-Net	N
[39]	The segmentation masks are generated from CT scans by employing a five-component Gaussian mixture model (GMM). The most probable component with a maximum a posteriori probability (MAP)estimate was used to label the lung voxels. State-of-the-art segmentation architectures are used for the three segmentation tasks.	Supervised	U-Net	3D lung-CT	Cropping, resizing, normalization, augmentation, etc.	Segmentation mask generated by U-Net	Y
[40]	Pseudo coloring is used to artificially assign colors to a grey scale which is then used to annotate the image done by radiologist & respiratory doctor. The work includes classification framework for the identification of COVID-19, the regression framework for the correlation analysis between the lesion areas and the clinical indicators. Grad-CAM is used to visualize the salient features	Supervised	CNN	X-Ray & CT images	pseudo -coloring, dimension normalization, etc	Bounding box	Y
Proposed Work	Mask R-CNN based detector is used $\rightarrow obtainthemask, boundingbox,$ and the class of the region of interest (ROI). The annotations of ROI by auto-annotation algorithm are utilized further to improve the pre-trained detector model without manual labeling from scratch.	Weakly Supervised	Mask-RCNN	CT images	None	generated by the Mask R-CNN based detector	Y

***-Represents the Best, Y/N - Yes or No.

TABLE III
COMPARISON OF CLOSELY RELATED ANNOTATION WORKS TARGETED TOWARD COVID DIAGNOSIS

Ref.	Annotation Type	Generate Synthetic Examples	Auto-Labeling	Ease Of Use	Generic/ ease of adaptablility to any detection/ localization tasks
[17]	Internal, attention heatmaps generated by using the gradient-weighted class activation	×	×	*	*
[39]	External, used existing lung-CT annotations to assist COVID-19 CT segmentation.	×	×	*	*
[40]	External, used pseudo-coloring to annotate the image done by radiologist & respiratory doctor.	×	×	*	*
Proposed Work	Internal, automatically machine generated annotations of infection masks is stored in JSON format	✓	✓	**	**

*** -Represents the Best, Y/N - Yes or No, ✓/× - Yes or No.

Our literature study shows that most of the works presented related to the COVID research are based on the semantic segmentation method, followed by binary classifier models. Further, most of the works are trained networks with large volumes of samples, be it a binary/multi classifier model or detection models. We primarily address the bottleneck largely associated with supervised deep models. Also, our focus, in comparison to other works is aimed at the pixel-level instance localization that is more significant in context to the nature and spread of the infectious disease, aimed to alleviate misdiagnosis or missed diagnosis, and quickly validate the clinical and laboratory findings for early treatment. Our localized COVID model can trace the severity degree with value-added information about size, shape, and count at instance-level to further help with follow-up and monitoring, and thereafter treating the patients with the precise remedial diagnosis. There are few key areas of possible improvisation that can be focused along the current research works. They are enlisted as follows.

1) No off-the-shelf system that does detection and localization with limited annotations ($<1\,\%$), and at the same time inversely generates more synthetic samples in the process.
2) Most of the works by large were kind of data-driven-hungry approach.
3) None focuses on benchmarked-based machine-generated samples and evaluations.
4) Generating more labeled training data for any custom user-input requirements, i.e., be it detection/recognition/localization/user-specific tasks, and any input dataset at ease.

In supplement to the existing works, and to address these research gaps, we develop a prototype envisioned toward computer-aided diagnosis systems backed with autonomous data labeling and machine-driven annotation solution.

The advantages of our proposed approach are as follows.

1) Localized autoannotations.

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

2) Automated machine generation of labels providing the pixel-level masks outputted with their corresponding annotation formats.
3) Learning with limited annotations, and inversely generating dense annotations on-the-go using an algorithmic-based computational approach.
4) Intuitive all-in-one pipeline to label objects and alleviate manual labeling process using AI-generated annotation proposals rendered with an autoannotation algorithm for enhancing and speedup the image annotation process.
5) Supports single/multiobject image autoannotations.
6) Fully automatic framework that can be generalized to any detection and localization tasks.

A. Motivation and Contributions

Labeling/annotations is one of the critical components of the supervised learning pipeline. Data labeling is one such task that requires much manual work. So, then can we get the bulk of such effort done by machines to minimize human tasks as the underlying research question of focus here. The trained model's inference depends on the labels we feed it with, which signifies the need for quality training data. The need of the hour is to alleviate such bottlenecks of having limited labeled data need for deep learning algorithms, and not go through the entire data annotation process by yourself. How intuitively can we synthetically generate annotated labels from a pretrained model trained on fewer samples? A more automatic way to annotate the dataset is significant for everyone who works on object detection, allowing us to focus on what matters instead of wasting time. We aim to facilitate this process by developing new algorithms and methods to avoid/reduce this mundane labor-intensive task.

The fundamental question that we aim to address here is: "How intuitively and quickly one can synthetically generate a large amount of reasonable quality machine-labeled datasets from a pretrained model trained on just a few samples?" This is just a first step, and we hope to arise with new methods and algorithms soon to facilitate further in this process.

The following are the key contributions.

1) A weakly supervised mask region based convolutional neural networks (R-CNN) detector trained with few samples for COVID infection detection from CT scan images.
2) Autoannotation algorithm.
3) Simple, flexible, and intuitive autoannotation framework for democratizing the power of AI.

III. Materials and Methods

A. Data Acquisition

The dataset was downloaded from the COVID-19 CT grand challenge [15], a set of over 750 PNG images. The challenge is to classify each CT image into COVID-19 positive case (the image has clinical findings of COVID-19) or COVID-19 negative case (the image does not have clinical findings of COVID-19). It is a binary classification problem based on the chest and lungs CT scan images.

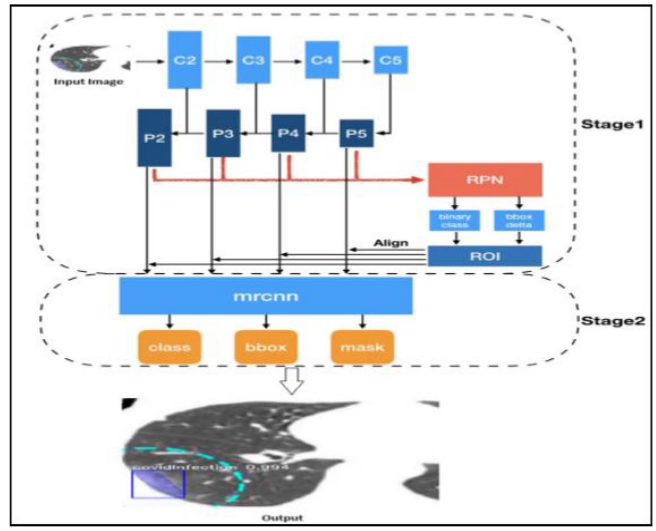


Fig. 1. Instance-based COVID-19 infection detector.

B. Instance-Based COVID-19 Infection Detector

The article presents a simple, flexible, and intuitive framework for COVID-19 infection instance segmentation. We use mask R-CNN [65] as a core of our framework due to its efficacy in detecting objects in an image while generating high-quality segmented masks for each instance. The other advantage being the solid baseline, that it outperforms all existing and single-model object detectors. The proposed instance-based real-time COVID detector is as shown in Fig. 1.

The detector predicts an object mask in parallel with the bounding box recognition. As seen in the figure, there are two stages of this instance-based detector. In the first stage, the proposals about the regions are generated for the possible objects. Next, based on stage one's proposals, the class of the object is predicted, the bounding box is refined, and a corresponding pixel-level mask is generated. The first stage consists of a neural network called region proposal network (RPN) that scans all FPN and proposes regions containing objects. At the second stage, another neural network takes the proposed regions from the first stage, assigns them to several specific areas of a feature map level, scans these areas, and generates objects classes, bounding boxes, and instance-level masks.

To better utilize each stage's features, mask R-CNN extends the backbone network to an FPN, which uses the inherent layering and multiscale properties of CNN to derive useful features for object detection. The goal of RPNs is to predict a set of region proposals efficiently. For RPNs training, the anchors with the largest intersection-over-union (IoU) overlapping with the ground-truth box are used as positive labels, and the anchors with IoU ratio below a certain threshold are used as negative labels. The representation of IoU is as seen in Fig. 2. If the prediction is cent percent accurate, then IoU = 1. Lower the IoU, the worse is the prediction result.

During the model training process, the loss function of the mask R-CNN model for each proposal is calculated as

$$\text{Loss} = \text{Loss}_{\text{Classification}} + \text{Loss}_{\text{Localization}} + \text{Loss}_{\text{InstanceMask}} \tag{1}$$

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

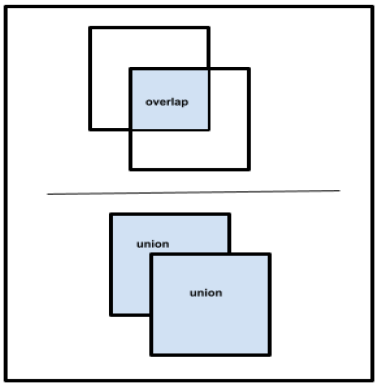


Fig. 2. IoU: Metric for measuring overlap between two bounding boxes or masks.

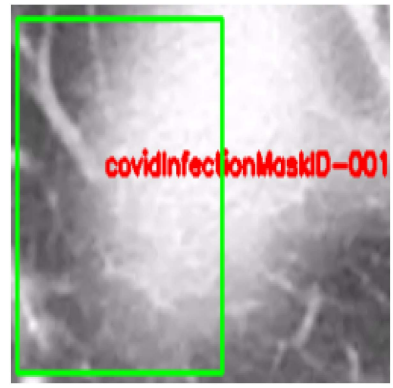


Fig. 3. Infection mask tracker.

where $\text{Loss}_{\text{Classification}}$ and $\text{Loss}_{\text{Localization}}$ represent classification loss and regression loss, respectively, and $\text{Loss}_{\text{InstanceMask}}$ represents segmentation loss; the classification and regression loss is computed as

$$\text{Loss}_{\text{Classification}} + \text{Loss}_{\text{Localization}} = \frac{1}{N}_{\text{Classifcation}} \left[\sum_i \text{Loss}_{\text{Classification}}(p_i, p_i^*)\right] + \frac{1}{N}_{\text{Localization}} \left[\sum_i \text{Loss}_{\text{Localization}}(t_i, t_i^*)\right] \quad (2)$$

where i represents the index of the anchor, p_i indicates the predicted probability of anchor i, t_i represents the four coordinate parameters of the box, and t_i^* represents the coordinate parameters of the ground-truth box corresponding to the positive anchor. If the anchor is positive, p_i^* is 1; otherwise, p_i^* is 0. By minimizing the loss function, the model is gradually optimized. The $\text{Loss}_{\text{InfectionMask}}$ is estimated as the average sigmoid cross-entropy loss as

$$\text{Loss}_{\text{InfectionMask}} = -\frac{1}{m^2} \sum_{i<=1, j<=m} [y_{ij} \log y_{ij}^k + (1 - y_{ij}) \log(1 - y_{ij}^k)] \quad (3)$$

where y_{ij} is the label of a cell(i, j) in the true mask for the region of size mXm; y_{ij}^k is the predicted value of the same cell in the mask learned for the ground-truth class k.

C. Multiple-Instances Tracking With Uniform Tagging of COVID Infections

To uniquely track the multiple infections, and label them continuously and uniformly throughout the whole input video stream, we used the centroid tracking algorithm. The bounding boxes are computed for each frame in the video. After the bounding boxes are assigned in the frame with their (x, y) coordinates, their centroids are calculated, and each bounding box is assigned a unique ID (see Fig. 3). For every subsequent frame bounding box, the object's centroid is estimated by associating with the centroid of the existing detected object by calculating the Euclidean distance between the detected and reference objects, using

$$d(x, y) = \sqrt{[(x_1 - y_1)^2 + (x_2 - y_2)^2]} \quad (4)$$

where x_1, y_1, x_2, y_2 are the bounding box coordinates for the detected object within the frame.

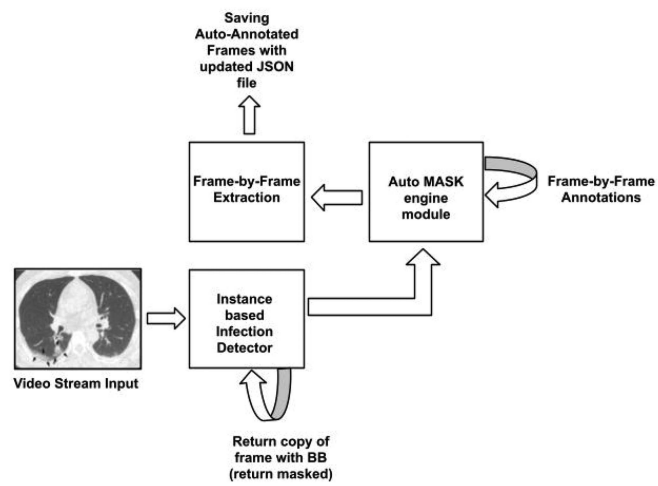


Fig. 4. Auto-annotation framework.

D. Proposed Frame-By-Frame Autoannotation Engine

Fig. 4 illustrates the pipeline of our AI-assisted-autoannotation framework. It consists of three main components: An infection detector (a proposal generator, a proposal classifier, an object localizer), an automask engine module, and a module for outputting frame-by-frame extracted annotations. The detector is used to obtain the mask, bounding box, and the class, which are localized.

The core of the framework is the proposed automask algorithm. It is responsible for autonomously annotating of the individual frames in the video input. The pseudocode of the same is given in Algorithm 1.

Once the image frame is segmented using mask R-CNN, the corresponding Region of Interest (ROIs) are generated. Further, the masks for each ROI is generated, followed by contour detection over the entire image. Then, the (x,y) coordinates are

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Frame-by-Frame Autoannotations.

Input: Video stream (New/unseen data)
Output: I_0, I_1,... I_n image sequence with autoannotated masked labels and annotation file

```
Create annotation file to save the automasking
Load annotation file to update the entries
Create region array list, create coordinate array list
for every image I in the sequence do
    result <- Instance-based detection of infection in the frame
    N <- result[generated region of interests]
    for i in N do
        mask <- masks[..., i]
        mask <- mask image with its binary mask
        contours <- do contour detection on I
        if length of contours is equal to 1 then
            contour <- reshape contour[0] to new shape(-1,2)
            Extract all the points (x coordinates) from contour[:,0]
            Extract all the points (y coordinates) from contour[:,1]
            Append the shape attributes and region attributes to region list
            Append the coordinate attributes to coordinates list
        end if
    end for
end for
Update & save annotation file
return Machine annotated labels
```

extracted from the contours. Finally, these shape, region, and coordinate attributes are outputted to the disk for each extracted frame. With this approach, the engine generates the annotations and its corresponding JSON file, which is responsible for annotating the images through the inference of a pretrained instance-based model that get the positions of the bounding boxes, and creating JSON metadata which can be then used in retraining. The machine-generated samples from this autoannotation module can be used further to improve the pretrained detector model without manual labeling from scratch. Once the detector fully learns, the inferences will be more accurate, and the autoannotation process implicitly will get better.

IV. Experimental Analysis and Results

The proposed AI-assisted autoannotation framework is validated by doing a quick proof-of-concept (POC) experiment. All the experiments were carried out with Intel Xeon (R) E-2176 G, GeForce RTX 2080 Ti graphics card, Ubuntu 18.04.02 LTS (Bionic) with 5.04 kernel that has 11 016 MB of next-gen, ultrafast GDDR6 memory. The Cuda version is 11.2, and the platform is developed using Python 3.6.5. We used the open-source dataset from the COVID-19 CT grand challenge, which is a set of over 750 PNG images of lung CT of which about half are COVID-19 positive. From the set of COVID-19 positive images, we randomly choose 18 images and split the data in train-validation-test set (8:5:5) ratio. The images were annotated using the VGG image annotator (VIA) tool [66]. The hyperparameters used are as given in Table IV.

TABLE IV
Hyperparameters

Hyper-parameters	Values
Back Bone	ResNet101 + FPN
Train_ROIs_Per_Image	128
ROI_POSITIVE_RATIO	0.33
Max_GT_Instances	4
Detection_Min_Confidence	0.9
Image_Min_Dim and Image_Max_Dim	128 x 512
Loss Weights : rpn_class_loss	1
Loss Weights : rpn_bbox_loss	1
Loss Weights : mrcnn_class_loss	1
Loss Weights : mrcnn_bbox_loss	1
Loss Weights : mrcnn_mask_loss	2

First, the detector model was trained on the eight samples from the training set to get the corresponding infection model weights using the pretrained common objects in context (COCO) weights. The training was done for 130 epochs. We used a weight decay of 0.0001, a learning momentum of 0.9, and a learning rate of 0.001. We tuned the detector with increased weights for mask loss for the precise detection of the infection masks since the identification at pixel level was of critical importance.

The RPN anchor classifier loss (rpn_class_loss), RPN bounding box loss (rpn_bbox_loss), classifier head loss of maskrcnn (mrcnn_class_loss), bounding box refinement loss (mrcnn_bbox_loss), masks head loss (mrcnn_mask_loss), and the general loss metric for the training set are 0.024016667, 0.342656667, 0.057926667, 0.415266667, 0.467336667, and 1.307206667, respectively. The loss values obtained for the validation set are 0.30 095, 0.66 969, 0.088143333, 0.47 156, 0.426423333, and 1.95 677, respectively.

A. Evaluation Criterion

1) Detector and Automask Association: The output of the detector can be seen in Fig. 5. The network captures the interested ROIs, and the automask engine masks them autonomously with few misses, for instance, as seen in Fig. 6.

The evaluation of the detection network is calculated based on the mean average precision (mAP) or just referred to as AP, a popular metric used to measure the performance of models doing information retrieval and/or object detection tasks. The AP values are averaged to obtain the metric: mAP. Precisely, mAP is the average of AP values over all classes estimated as

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i. \tag{5}$$

This metric considers the matching between the ground-truth masks and their corresponding predictions, and gradually penalizes the resulting mask's correctness at a pixel level, from 0.5 to 0.95 of IoU. The calculation of IoU is

$$\text{Intersection over Union } = (\text{predicted box} \cap \text{ground truth box}) \\ /(\text{predicted box} \cup \text{ground truth box}) \tag{6}$$

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

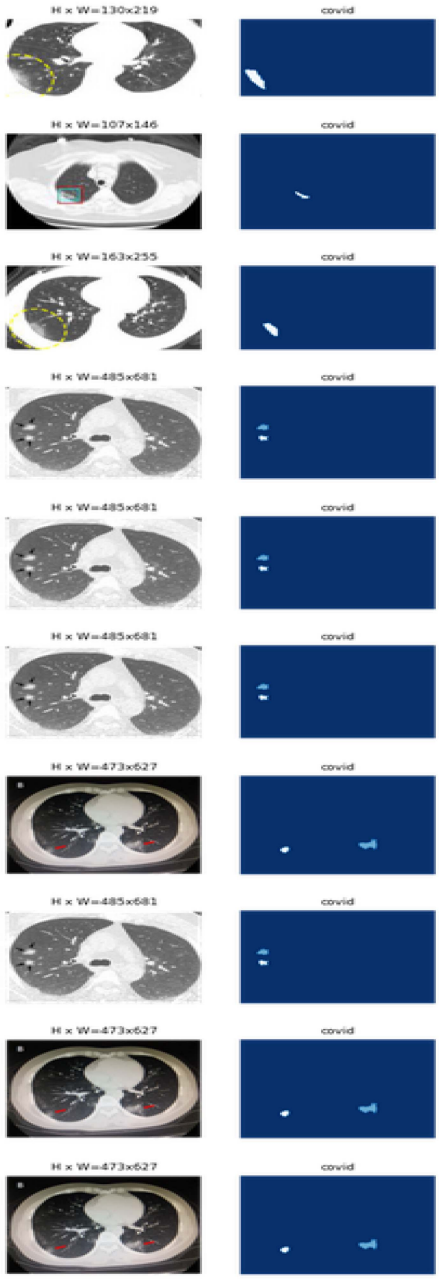


Fig. 5. Infection mask.

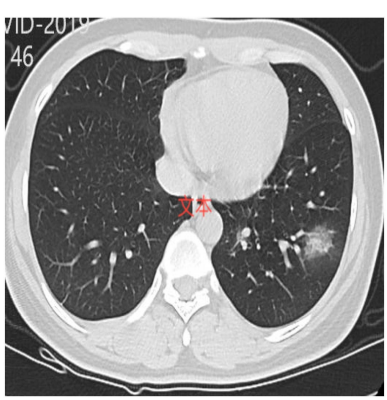


Fig. 6. COVID infection detector miss case.

TABLE V
DETECTOR PERFORMANCE

Instance-based COVID-19 Detector	Train	Validation	Test
Average Precision (AP)	0.989	0.903	0.761
	0.994	0.94	0.836
	0.991	0.981	0.782
	0.993	0.914	0.899
	0.975	0.919	0.723
	0.996		
	0.991		
	0.992		
Mean of all AP	**0.99**	**0.931**	**0.8**

TABLE VI
MASK ASSOCIATION RESULTS FOR AUTOANNOTATION ALGORITHM

Association Results/Quantitative metrics	Evaluation
New Unseen Video Stream Timeframe (Test)	18 sec
No. of instance detected frames generated from the video (0.5 FPS)	174
Avg. no of ROI per sequence	1 - 2
Total Misses	36
Avg. Misses per sequence	0.11
Success Rate for Auto-Annotation Algo	79.70%
Miss Rate for Auto-Annotation Algo	21.30%

where the predicted result indicates the predicted box, and the ground-truth results indicate the ground-truth box.

Table V shows the performance of the instance-based COVID detector on the Train-Val-Test set, respectively. To evaluate object detection models, the mean Average Precision (mAP) metric is widely used. The Average Precision (AP) finds the area under the precision-recall curve. The mAP is the average of AP, that is calculated for each class. The mAP is used to compare the ground-truth box to the predicted box, and returns a score. Higher the score, better is the detection accuracy. The average precision of the individual samples, and the mAP is reported. The train set reported 99% precision, while for the validation and test sets, the accuracy was lowered to 93% and 80%, respectively.

Next, to evaluate our automask algorithm's performance, we tested with a sample input unseen test video data compiled from the COVID infected images chosen randomly from the COVID-19 CT grand challenge dataset. The total duration length of the input video stream was 18 s. The autoannotation algorithm's performance is reported in Table VI. From the input video stream, around 175 frames were detected, generating that many machine-generated synthetic samples with their corresponding annotations of the infection masks which is saved with their information outputted to the VIA format JSON file. Evaluation of the machine-generated annotated data is done based on the Euclidean pixel distance between a human-labeled input mask, and the network outputted mask. Missed detections are calculated if the maximum probability out of that region is lower than 0.3, then it is considered a miss. The detection error is kept within a tolerance of 2 pixels with a 95% confidence region for a 512×512-pixel resolution image. The overall missed detection rate is <22 %, which will be lowered with the detector's correct inferences. We adhered to the stringent tolerance of 1–2 pixels as any deviation from the infection regions is critical to the developed virus infected regions observed in the lung CT scans.

The IoU is computed for the output prediction box in relation to the ground-truth box. We report the IoU for a randomly selected sample from each of the train-val-test set. The IoU score were 0.8814, 0.8045, and 0.6916, respectively.

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

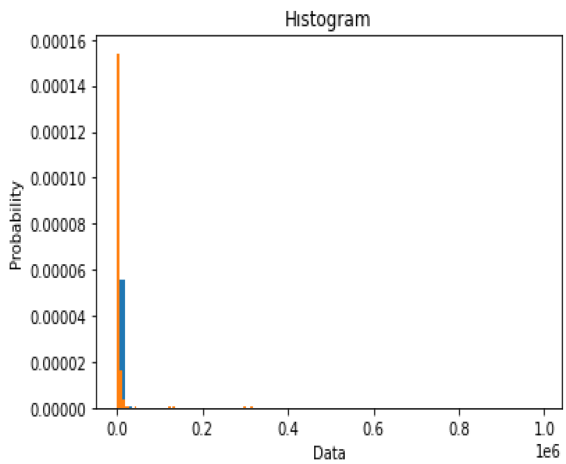


Fig. 7. Histogram overlay: Orange color bar depicts histogram from machine mask, and blue from the human-labeled mask.

Algorithm 2: Quality Checks and Similarity Associations.

Input: Human-annotated and machine-annotated images
Output: Histogram comparison with rms difference, percent similarity

1: Compare histogram of both set of inputs.
2: Estimate rms value from histograms.
3: Estimate similarity using rms of the difference between each pair of corresponding pixels of both inputs divided by max distance threshold between two pixels.

2) Qualitative Assessment: To examine the machine-generated mask and the human-annotated mask, we preliminarily began with histogram comparisons (see Fig. 7), and further estimating the root-mean-square (rms) difference value from these histograms as

$$\text{RMSdel}_{\text{value}} = \sqrt{[\frac{1}{\text{len}(h_1)} \sum (h_1(x) - h_2(y)^2)]} \quad (7)$$

where h_1 is the histogram of the machine generated mask, and h_2 in the human-annotated mask. A histogram difference is an effective measure for detecting similarity in images.

The qualitative assessment of the machine generated and human-labeled data is done as given in Algorithm 2.

We also estimated the percent similarity between the both sets (i.e., machine generated labels, and manual human-annotated labels). The percent similarity was 92.3269%.

As seen in the Fig. 8, from the first image, it is possible to diagnose the COVID-19 positive case because the left lobe's cloudy white region is an example of ground-glass opacity (GGO), which is one of the critical characteristics of identifying COVID-19 in lung CT. The second image is COVID -ve, while the third image is the positive case. There is no trace of this ground glass appearance on the image to the right. Not all the lung CT images containing ground glass characteristics are COVID-19 positive and vice versa. Other conditions such as infectious diseases, interstitial lung diseases, and acute alveolar diseases also display GGO in lung CT scans. Further, from the last image, it can be inferred that not all of the images are easy to identify.

The output of the autoannotation algorithm can be seen in Fig. 9. The network captures most of the infections and predicts them correctly. It further provides a label mask with their score. The picture on the left shows the machine-generated auto mask, while that on the right shows the human-annotated ground-truth samples.

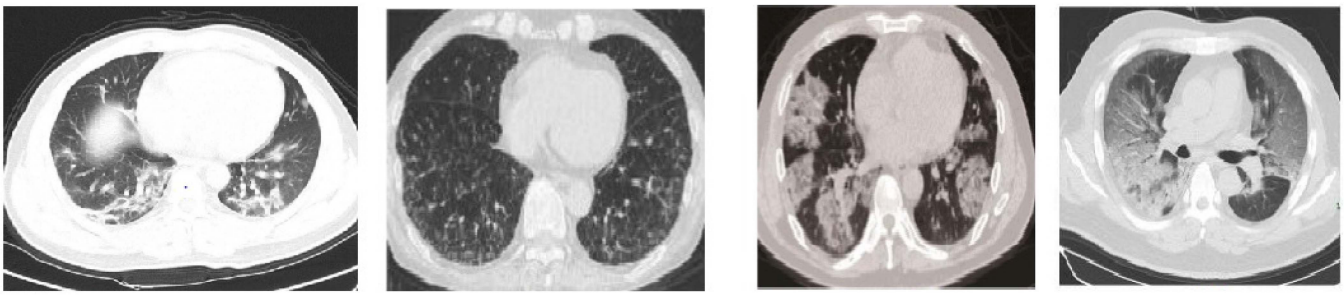
V. Discussion

This is the first work aimed toward pixel-level infection detection of the COVID-19 virus infected regions, and corresponding automated machine generation of the pixel-level masks to the best of our knowledge. We currently validated our POC prototype, which was tested with only limited annotated ground-truth data for both stages, i.e., the detector performance, and the consequent machine-generated labels evaluation. Compared to the massive amount of data that are generally required for training the deep learning algorithms, we begin with a very weakly supervised model. As long as we train the detector model upfront with large amount of labeled data, the inferences are likely to be more accurate, and precise, thereby reducing the missed detection's, minimizing the point pixel deviation of the machine-generated masks thereby needing no human confirmation or no correction altogether. But, said so, we do not intend to proceed along this lines, as this brings to the same fundamental core problem of requiring more manual human-annotated data initially which in itself is overhead. Instead, we focus on introducing a penalized-based scheme following the gradients to self-improvise the detector.

Presently, the COVID infection detector's inference speed of detecting the infection masks is at 0.5–0.6 frames per second (FPS). The total duration length of the new/unseen test video stream was 18 s, that produced around 175 machine-generated synthetic samples. To autogenerate the labels for these samples, it took around 419 s (≈7 min). With a higher inference speed, this time will further be lowered in combination with the computing device in use (CPU/GPU). We display few random machine-generated, and their corresponding human-annotated ground-truth label of the extracted frames in Fig. 9. Also, the output of the autoannotation module as seen opened with the VIA image annotation tool for a particular extracted frame within the new unseen test video input stream is as given in Figs. 10 and 11. At present, the autonomous machine-generated masks have some point pixel deviation from the human-labeled ground-truth masks.

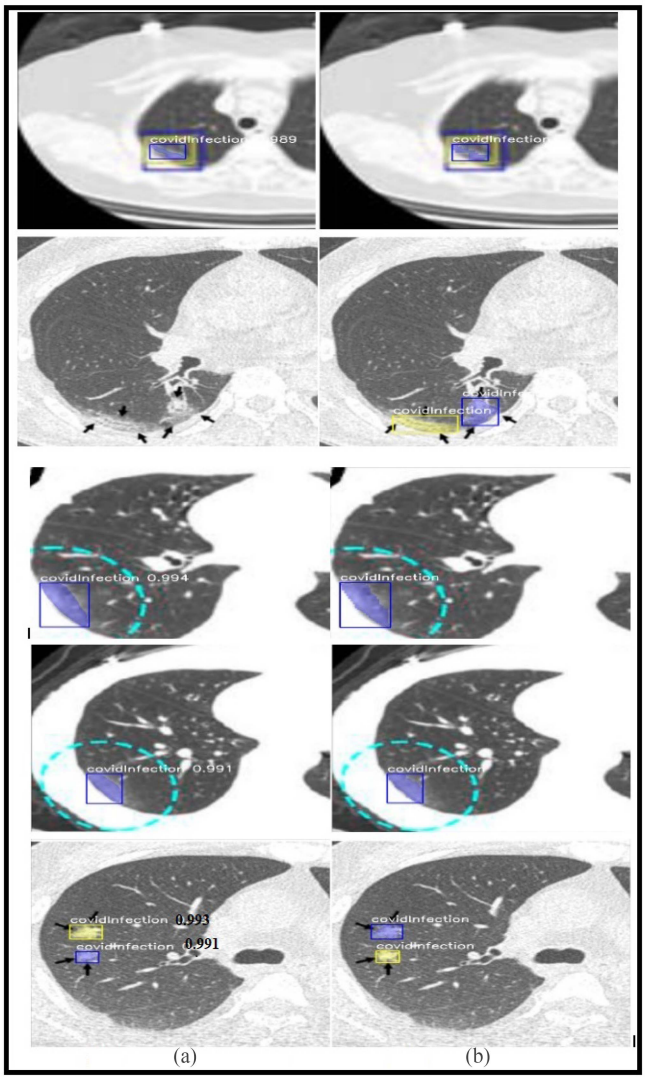
Using mask R-CNN architecture, our framework presents an all-in-one model design that output individual frame-by-frame infection masks, aiming to improvise the complex instance segmentation tasks, and intrinsically leading to real-time object detector while generating more synthetic pixel-level samples along the detection process. The proposed mask R-CNN-based autoannotation framework is robust as it can accurately detect and segment infections for the varying CT scans accurately, which were also of very low quality. The detector was very weakly trained with less than 10 samples, which is nothing compared to the massive training data required for the deep

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c) (d)

Fig. 8. (a) COVID-19 positive mainly because the cloudy white region of the left lobe is an example of GGO. (b) COVID-19 negative. (c) COVID-19 Positive. (d) Another positive case of COVID-19.



(a) (b)

Fig. 9. (a) Machine auto-annotation. (b) Human annotation.

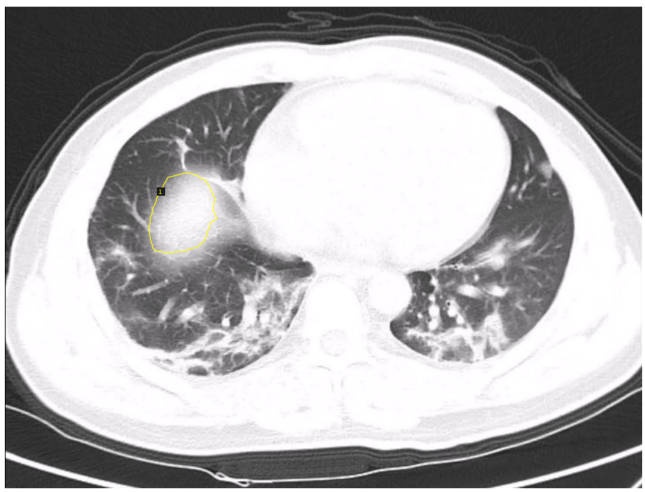


Fig. 10. VIA inference of machine-labeled image.

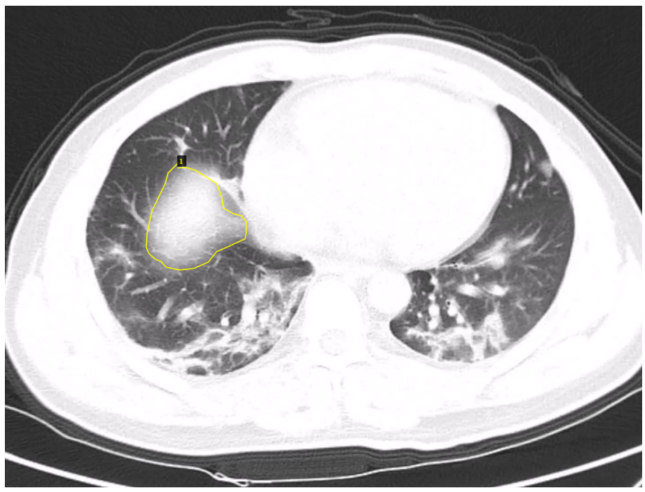


Fig. 11. VIA inference of human-labeled image.

models to perform well. Limitations mainly occurred once the cloudy white regions were very faint and had a light background, but then with such cases even the human annotators will also face difficulties to identify infected regions correctly. In specific situations, the autoannotation algorithm fails to automask the frames, which resulted in the maximum number of misses. Our deep learning based instance-level COVID model presented is trained on frames of 128 × 512 pixels. At present, we have not done any preprocessing and postprocessing steps, and the images were used as it is without any transformation. The inclusion

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

of the preprocessing may help to improve the detector model further. Next, our long-term focus is to self-improvise the current model either by integrating reinforcement learning or using transfer learning to propagate labels with an encoder-decoder network given whatever the limited trained data used initially. But, then at the same time the system should be able to generate more enhanced quality of synthetic annotated samples quickly that can help humans ease their work, and alleviate the mundane manual data labeling process. The prototyped detector missed detecting a few infected cases entirely from the tested input samples, as seen in Fig. 9 (second from top). Toward this, we initially plan to verify the detector by using ResNeXt-101-FPN architecture as the backbone. Currently, our designed framework is based on mask R-CNN and the VIA annotation format. We also aim to generalize our prototype to include the different detectors, specifically you only live once (YOLO) and the corresponding YOLO compatible annotation format. Further, we also plan to integrate the COCO annotation format. It would be worthy of integrating all the different image annotations as part of our framework while providing the facility with provision of different libraries, i.e., Torch, TensorFlow, etc. We also plan to inspect the COVID 3D CT SCAN dataset [37] and then extend our work to autoannotate the 3D image.

VI. Conclusion

Chest CT has been profoundly used for clinical trials of COVID-19 infection identification. AI technology can be leveraged to aid in all the associated tasks with COVID data. With the rapid spread of infections, one certainly needs quick findings to support the healthcare solutions offering potential alternatives. With our this effort, we aim to address one key aspect toward detection of COVID-19 pneumonia that is based on data-driven approach requiring large amounts of labeled data for the success of supervised deep architectures. Labeling data is a nontrivial task that requires enormous manual effort. Further, annotation methods mainly rely on the bulk action of human labelers. Can we get the bulk of this effort done by machines to complement human tasks? This was the key motivation and underlying research question that we were keenly interested in. The contributions of the article are threefold: 1) Proposed frame-by-frame autoannotations architecture for generating machine labeled data, 2) a fully automated intuitive generic framework aimed toward democratizing the power of AI to generate large machine labeled datasets in any object recognition context quickly, and 3) a full weakly trained supervised mask R-CNN COVID infection detector is developed. The detections were obtained with very little human-inputted-annotated data to generate large annotated synthetic data automatically on the fly. The mAP values of 0.99, 0.931, and 0.8 was achieved for train, validation, and test set, respectively, demonstrating the ability of detection network which is promising. This performance can be further excelled with the detector's enhancement for providing a higher quality of annotations to eliminate HITL interventions completely. As a part of our extended work, we are working on developing an intuitive UX platform for generating machine annotations by providing a simple point-and-click user interface for the users to use our facility at ease, while given their any custom user-input data, supporting different network architectures, different annotation formats, and service for different data formats - point-cloud, 3D, text, audio, etc., along with inclusion of different deep learning frameworks, and different image annotation types. In the near future, we also intend to include the CT scans obtained from normal patients to demonstrate the robustness of the model in differentiating COVID positive cases from the cases that do not show clinical findings of COVID-19 infected strains.

Acknowledgment

The authors would like to thank all the reviewers and editors for their valuable suggestions to enhance the depth and quality of the manuscript.

References

[1] "World Covid-19 tracker: Latest cases and deaths by country." Accessed: May 31, 2021, https://www.cnn.com/interactive/2020/health/coronavirus-maps-and-cases/

[2] V. M. Corman *et al.*, "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR," *Eurosurveillance*, vol. 25, no. 3, 2020, Art. no. 2000045.

[3] K. Prem *et al.*, "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study," *Lancet Public Health*, vol. 5, no. 5, pp. e 261-e270, 2020.

[4] N. Kahalé, "On the economic impact of social distancing measures," *Frontiers Sociol.*, vol. 5, pp. 1–13, 2020.

[5] D. Kai, G.-P. Goldstein, A. Morgunov, V. Nangalia, and A. Rotkirch, "Universal masking is urgent in the COVID-19 pandemic: Seir and agent based models, empirical validation, policy recommendations," 2020, *arXiv:2004.13553*.

[6] A. Chavda, J. Dsouza, S. Badgujar, and A. Damani, "Multi-stage CNN architecture for face mask detection," *6th Inter. Conf. Converg. Technol. (I2CT), IEEE*, 2021, pp. 1–8.

[7] M. S. Islam, E. H. Moon, M. A. Shaikat, and M. J. Alam, "A novel approach to detect face mask using CNN," in *Proc. 3rd Int. Conf. Intell. Sustain. Syst.*, 2020, pp. 800–806.

[8] C. Li, R. Wang, J. Li, and L. Fei, "Face detection based on YOLOv3," in *Recent Trends in Intelligent Computing, Communication and Devices*. Singapore: Springer, 2020, pp. 277–284.

[9] T. Q. Vinh and N. T. N. Anh, "Real-time face mask detector using YOLOv3 algorithm and Haar cascade classifier," in *Proc. Int. Conf. Adv. Comput. Appl.*, 2020, pp. 146–149.

[10] S. Saponara, A. Elhanashi, and A. Gagliardi, "Implementing a real-time, AI-based, people detection and social distancing measuring system for COVID-19," *J. Real-Time Image Process.*, pp. 1–11, 2021.

[11] I. Ahmed, M. Ahmad, J. J. Rodrigues, G. Jeon, and S. Din, "A deep learning-based social distance monitoring framework for COVID-19," *Sustain. Cities Soc.*, vol. 65, 2021, Art. no. 102571.

[12] T. Rahman, "Covid-19 radiography database," 2021. [Online]. Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database/

[13] GitHub. 2021. agchung/Figure1-COVID-chestxray-dataset. Accessed 15 May 2021. [Online] Available: https://github.com/agchung/Figure1-COVID-chestxray-dataset/

[14] P. Angelov and E. A. Soares, "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification," *medRxiv*, 2020.

[15] J. Zhao, Y. Zhang, X. He, and P. Xie, "COVID-CT-dataset: A CT scan dataset about COVID-19," vol. 490, 2020, *arXiv:2003.13865*.

[16] M. Rahimzadeh, A. Attar, and S. M. Sakhaei, "A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset," *Biomedical Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102588.

[17] L. Li *et al.*, "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy," *Radiology*, vol. 296, no. 2, pp. E 65-E71, 2020.

[18] W. El-Shafai, "Extensive COVID-19 X-Ray and CT chest images dataset," 2020. [Online]. Available: http://dx.doi.org/10.17632/8h65ywd2jr.3

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.

[19] X. Xu *et al.*, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," *European Radiol.*, vol. 31, no. 8, pp. 6096–6104, 2021, doi: 10.1007/s00330-021-07715-1.

[20] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)," *Eur. Radiol.*, 2021.

[21] "COVID-19 - medical segmentation," Accessed 21 May 2021. [Online]. Available: http://medicalsegmentation.com/covid19/

[22] "Covid-19 - SIRM. 2021." Accessed 23 May 2021. [Online]. Available: https://www.sirm.org/category/senza-categoria/covid-19/

[23] J. Pu *et al.*, "Automated quantification of COVID-19 severity and progression using chest CT images," *Eur. Radiol.*, vol. 31, no. 1, pp. 436–446, 2021.

[24] E. B. Tsai *et al.*, "The RSNA international COVID-19 open annotated radiology database (RICORD)," *Radiology*, vol. 299, no. 1, 2021, Art. no. 203957.

[25] S. Wang *et al.*, "A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis," *Eur. Respiratory J.*, vol. 56, no. 2, 2020, Art. no. 2000775.

[26] Q. Zhang *et al.*, "Artificial intelligence clinicians can use chest computed tomography technology to automatically diagnose coronavirus disease 2019 (COVID-19) pneumonia and enhance low-quality images," *Infection Drug Resistance*, vol. 14, pp. 671–687, 2021.

[27] R. Hu, G. Ruan, S. Xiang, M. Huang, Q. Liang, and J. Li, "Automated diagnosis of COVID-19 using deep learning and data augmentation on chest CT," *medRxiv*, 2020, doi: 10.1101/2020.04.24.20078998.

[28] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning," *J. Biomol. Struct. Dyn.*, pp. 1–8, 2020.

[29] D. Singh, V. Kumar, and M. Kaur, "Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks," *Eur. J. Clin. Microbiol. Infectious Disease*, vol. 39, no. 7, pp. 1379–1389, 2020.

[30] Z. Han *et al.*, "Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2584–2594, Aug. 2020.

[31] H. Kang *et al.*, "Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2606–2614, Aug. 2020.

[32] H. Ko *et al.*, "COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: Model development and validation," *J. Med. Internet Res.*, vol. 22, no. 6, 2020, Art. no. e 19569.

[33] Q. Ni *et al.*, "A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images," *Eur. Radiol.*, vol. 30, no. 12, pp. 6517–6527, 2020.

[34] X. Liu, Z. Deng, and Y. Yang, "Recent progress in semantic image segmentation," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 1089–1106, 2019.

[35] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: State of the art," *Int. J. Multimedia Inf. Retrieval*, no. 9, pp. 171–189, 2020.

[36] A. Saood and I. Hatem, "COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet," *BMC Med. Imag.*, vol. 21, no. 1, 2021, Art. no. 19.

[37] J. Ma *et al.*, "Towards data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation," *Med. Phys.*, vol. 48, no. 3, pp. 1197–1210, 2020.

[38] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 4–15, Apr. 2021.

[39] S. Liang *et al.*, "Fast automated detection of COVID-19 from medical images using convolutional neural networks," *Commun. Biol.*, vol. 4, no. 1, 2021, Art. no. 35.

[40] P. Zaffino *et al.*, "An open-source COVID-19 CT dataset with automatic lung tissue classification for radiomics," *Bioengineering*, vol. 8, no. 2, 2021, Art. no. 26.

[41] O. Gozes *et al.*, "Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," pp. 1–19, 2020. [Online]. Available: https://arxiv.org/abs/2003.05037v1

[42] L. Li *et al.*, "Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT," *Radiology*, vol. 296, no. 2, 2020, Art. no. 200905.

[43] Y. Song *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT Images," *IEEE/ACM Trans. Computat. Bio. Bioinformat.*, pp. 1–8, 2021, doi: 10.1109/TCBB.2021.3065361.

[44] F. Shan *et al.*, "Lung infection quantification of COVID-19 in CT images with deep learning," pp. 1–19, 2020, [Online]. Available: https://arxiv.org/abs/2003.04655v2

[45] T. Anwar and S. Zakir, "Deep learning based diagnosis of COVID-19 using chest CT-scan images," in *Proc. IEEE 23rd Int. Multitopic Conf.*, 2020, pp. 1–5.

[46] C. Zheng *et al.*, "Deep learning-based detection for COVID-19 From chest CT using weak label," pp. 1–13, 2020, [Online]. Available: https://europepmc.org/article/PPR/PPR117705.

[47] A. Mohammed *et al.*, "Weakly-supervised network for detection of COVID-19 in chest CT scans," *IEEE Access*, vol. 8, pp. 155987–156000, 2020.

[48] J. Wang *et al.*, "Prior-attention residual learning for more discriminative COVID-19 screening in CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2572–2583, Aug. 2020.

[49] S. Bhattacharya *et al.* "Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey," *Sustain. Cities Soc.*, vol. 65, 2021, Art. no. 102589.

[50] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. M. U. Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *Int. J. Inf. Technol.*, vol. 12, no. 3, pp. 731–739, 2020.

[51] A. Vaid *et al.*, "Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: Machine learning approach," *JMIR Med. Inform.*, vol. 9, no. 1, 2021, Art. no. e 24207.

[52] J. Samuel *et al.*, "COVID-19 public sentiment insights and machine learning for tweets classification," *Information*, vol. 11, no. 6, 2020, Art. no. 314.

[53] D. Dimitrov *et al.*, "TweetsCOV19-A knowledge base of semantically annotated tweets about the COVID-19 pandemic," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 2991–2998.

[54] J. Chen *et al.*, "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography," *Sci. Rep.*, vol. 10, no. 1, 2020, Art. no. 19196.

[55] H. Gunraj, L. Wang, and A. Wong, "COVIDNet-CT: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images," *Frontiers Med.*, vol. 7, 2020, Art. no. 608525.

[56] P. Silva *et al.*, "COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis," *Inform. Med. Unlocked*, vol. 20, 2020, Art. no. 100427.

[57] S. Bekhet, M. Hassaballah, M. A. Kenk, and M. A. Hameed, "An artificial intelligence based technique for COVID-19 diagnosis from chest X-ray," in *Proc. IEEE 2nd Novel Intell. Leading Emerg. Sci. Conf.*, 2020, pp. 191–195.

[58] "24 best image annotation tools for computer vision," 2020. [Online]. Available: https://lionbridge.ai/articles/image-annotation-tools-for-computer-vision/

[59] S. Zhu, Z. Shi, C. Sun, and S. Shen, "Deep neural network based image annotation," *Pattern Recognit. Lett.*, vol. 65, pp. 103–108, 2015.

[60] V. N. Murthy, S. Maji, and R. Manmatha, "Automatic image annotation using deep learning representations," in *Proc. 5th ACM Int. Conf. Multimedia Retrieval*, 2015, pp. 603–606.

[61] Y. Gong, Y. Jia, T. Leung, A. Toshev, and S. Ioffe, "Deep convolutional ranking for multilabel image annotation," *CoRR*, pp. 1–9, 2014.

[62] M. B. Mayhew, B. Chen, and K. S. Ni, "Assessing semantic information in convolutional neural network representations of images via image annotation," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 2266–2270.

[63] H. Hu, G.-T. Zhou, Z. Deng, Z. Liao, and G. Mori, "Learning structured inference neural networks with label relations," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 2960–2968.

[64] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 618–626.

[65] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 2961–2969.

[66] "VGG Image Annotator (VIA)." [Online]. Available: https://www.robots.ox.ac.uk/vgg/software/via/

Authorized licensed use limited to: Wilfrid Laurier University. Downloaded on November 10,2024 at 13:28:13 UTC from IEEE Xplore. Restrictions apply.