IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.DOI

Semi-supervised Network for Detection of COVID-19 in Chest CT Scans

AHMED MOHAMMED*, CONGCONG WANG*, MENG ZHAO*, MOHIB ULLAH*, RABIA NASEEM*, HAO WANG, MARIUS PEDERSEN, FAOUZI ALAYA CHEIKH

*Authors with equal contribution, names are listed in alphabetical order
Norwegian University of Science and Technology (NTNU), Norway

Corresponding author: Mohib Ullah (e-mail: mohib.ullah@ntnu.no).

ABSTRACT Deep Learning-based chest Computed Tomography (CT) analysis has been proven to be effective and efficient for COVID-19 diagnosis. Existing deep learning approaches heavily rely on large labeled data sets, which are difficult to acquire in this pandemic situation. Therefore, semi-supervised approaches are in demand. In this paper, we propose an end-to-end semi-supervised COVID-19 detection approach, ResNext+, that only requires volume level data labels and can provide slice level prediction. The proposed approach incorporates a lung segmentation mask as well as spatial and channel attention to extract spatial features. Besides, Long Short Term Memory (LSTM) is utilized to acquire the axial dependency of the slices. Moreover, a slice attention module is applied before the final fully connected layer to generate the slice level prediction without additional supervision. An ablation study is conducted to show the efficiency of the attention blocks and the segmentation mask block. Experimental results, obtained from publicly available datasets, show a precision of 81.9% and F1 score of 81.4%. The closest state-of-the-art gives 76.7% precision and 78.8% F1 score. The 5% improvement in precision and 3% in the F1 score demonstrate the effectiveness of the proposed method. It is worth noticing that, applying image enhancement approaches do not boost the performance of the proposed method, sometimes even harm the scores, although the enhanced images illustrate better perceived visual image quality.

INDEX TERMS Attention, COVID-19, Computed Tomography, Detection, Enhancement, LSTM, Semi-supervised learning

I. INTRODUCTION

COVID-19 has proliferated to more than 213 countries and territories in the world. The total number of reported cases until the time of writing (August 18, 2020) surpassed 21.9 million. Besides claiming lives of more than 774299, the cases are surging every day from almost every territory of the world [1]. The exponential human-to-human spread of the virus instigated worldwide apprehension which consequently forced the nations to take extreme measures in quest of effective solutions. Among the current diagnosis solutions, the real-time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) test is the golden standard for COVID-19 confirmation. The rRT-PCR test is mainly done on respiratory samples obtained from people who have shown clinical symptoms [2]. However, the available rRT-PCR solutions have very high false-positive rates, which leads the suspected patients to be tested multiple times for achieving convincing diagnosis [3]. To efficiently utilize the scarce rRT-PCR resources as well as better accuracy in COVID-19 diagnosis, doctors are also relying on additional medical imaging technologies.

In the Computed Tomography (CT), COVID-19 manifests as a consolidated ground-glass opacity patch, scattered patches, and the thickening of interlobular septa on lung (CT) [4]. The lung lesions expand in size and density with the progression of disease [5]. COVID-19 infected area of lung appears more contrasted than its surroundings in the chest CT. Primarily, such visibility in the respiratory system makes the chest CT suitable for diagnosing suspected COVID-19 cases [6], [7]. A recent literature review, conducted in March 2020, shows that "chest radiographs are of little diagnostic value in early stages, whereas CT findings may be present even before symptom onset" [5]. However, the aggressively growing number of suspected cases and the limited availability of medical diagnostic methods and resources have been putting pressure on medical professionals all over the world.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

To significantly alleviate the diagnostic workload, computer vision researchers have recently proposed viable solutions for detecting COVID-19. Among the proposed solutions, Artificial Intelligence (AI)-based chest CT analysis is playing an important role in fighting the COVID-19 pandemic [8]. AI-based diagnosis [9] is a supplementary assistance tool for radiologists, who usually have to analyze a large number of CT scans for diagnosis on a daily basis.

The AI-based analysis of chest CT for probable COVID-19 prevalence involves multiple steps from image acquisition, image pre-processing, segmentation, and final diagnosis [10]. However, the recent deep learning-based approaches require labeled datasets to train models. Since the labeling process of CT scans requires expert knowledge (mainly from a radiologist) and a significant amount of time, most of the supervised learning-based models are trained on a limited amount of data. Fully supervised methods that are trained on insufficient data usually are limited in their performance [11]. Therefore, a semi-supervised approach for COVID-19 detection from weakly labeled data is indispensable. Following a similar approach to [12]–[14], we propose a semi-supervised approach to detect the pathology of COVID-19 in the individual slices of CT scan using only volume level data labels. More precisely, a Convolutional Neural Network named ResNext+ is proposed that integrates a lung segmentation mask with the corresponding CT volume and extract spatial features from the CT volume. Additionally, the spatial and channel attention module is Incorporated in Restnext+ architecture for refining the feature maps. Then bidirectional LSTM is exploited for the axial dependency of the input slices. Essentially, the bidirectional LSTM transforms the spatial features to spatial-axial features. After that, a slice attention module is introduced to weight the importance of each slice and finally, a fully connected layer is utilized for the final slice and volume level prediction. Furthermore, as a pre-processing, two enhancement approaches are exploited for improving the accuracy of the model. In a nutshell, the contributions of the proposed method are threefold:

- We designed an end-to-end framework that is capable of training on weakly labeled data. The network consists of a convolutional neural network named ResNext+ that takes a CT slice together with a binary mask of the lung section as input and fuse both pieces of information and gives spatial features of the CT volume. Channel and spatial attention modules are incorporated in an end-to-end fashion that helps refine the feature maps. Additionally, a bidirectional LSTM is exploited for transforming the spatial features into spatial-axial features.
- For enhancing the quality of slices, two types of enhancement algorithms namely stochastic sampling and tone mapping are exploited that specifically highlight the details inside the lung region for assisting network training and diagnosis.
- We introduced slice attention that helps the network to focus on the semantic slices for the final inference. In addition to volume level prediction, the slice attention network gives the slice level prediction which helps in localizing the infected region of the lung due to COVID 19.

II. RELATED WORK

Several imaging modalities including x-ray [15], [16], CT [6], [7], [17]–[21], and ultrasound [22], [23] have been employed for diagnosis of COVID-19. These imaging modalities can be used with the increasing number of deep learning-based COVID-19 detection methods. Usually, a pre-processing step like haze removal [24], enhancement [25], etc. are preceded by deep learning algorithms. In this work, the proposed approach is primarily intended for diagnoses of COVID-19 from chest CT images of suspected patients. The current state-of-the-art regarding deep learning-based diagnosis on CT scan images is mostly focused on the detection of lung nodules and COVID-19 diagnosis. Setio *et al.* [26] presented a nodule detection method where 2D patches of the candidate nodules in lung CT are extracted from multiple planes. The network contains various streams of 2D ConvNets for each patch from the lung volume. Later, their outputs are fused to obtain the final classification. Similarly, Xie *et al.* [27] introduced a nodule detection methodology improving the Fast Region-based Convolutional Network (R-CNN) network [28] through the introduction of two region proposal networks. The proposed network concatenates relevant information from the lower layer and their deconvolution layer to yield candidate nodules [29]. They used VGG16 [30] for feature extraction. Additionally, they incorporated the 3D input data contextual information that is generated by systematically training three separate models on three types of slices and finally fused the results. It is also mentioned that their model is trained again with the wrongly classified samples for improving the accuracy of the algorithm. To reduce the rate of fast positive, a novel architecture named ZNET [31] is introduced that uses two CNNs; one for obtaining candidate nodules and the other for reducing false positives [32]. The authors used UNet to generate a probability map based on which candidate nodules are acquired in axial slices. Subsequently, they generate candidate masks through thresholding. The LUNA16 challenge evaluation finally indicated that ZNET outperformed several other methods [32].

The success of such approaches leads to the successful deployment of several AI-based commercial CT platforms in combating COVID-19 [15], [33]. A comprehensive review of AI techniques in image data acquisition, segmentation, and diagnosis of COVID-19 is presented in [8]. The state-of-the-art AI-assisted diagnosis approaches can be partly grouped into the following three categories. A brief overview of each category is given in the following.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

A. CLASSIFICATION OF COVID-19 VERSUS NON-COVID-19

Some of the recent studies aim to discriminate COVID-19 patients from non-COVID-19 ones. Most of these methods are based on different variant of UNet [34]. For example, the algorithm proposed by Chen et al. [35] and Zheng et al. [36], [37] mainly utilize UNet, UNet++ [38], and U-Net+3D based model architectures. The architecture of UNet has been used in different networks for the region of interest extraction, predicting suspicious lung regions, segmentation, and other related tasks. Among the different tested segmentation models such as U-Net [34], V-Net [39], FCN-8s [40], and 3D U-Net++ [38], 3D U-Net++ is reported to yield the best performance for segmentation [37]. Also, the combination of 3D U-Net++ segmentation with ResNet-50 model [41] has shown to perform better classification compared to other models like DPN-92 [42], Inception-v3 [43] and Attention ResNet-50 [44]. Most CNN models proposed for lung segmentation and COVID-19 diagnosis are trained on slice [45] or volume level [36], thus can predict slice or volume level scores, respectively. In such approaches, after slice prediction blocks, the slice scores are mostly fused to come up with the case-level diagnosis.

B. CLASSIFICATION OF COVID-19 VERSUS OTHER VIRAL PNEUMONIA

On many occasions, the appearance of COVID-19 lung infections and those of other pneumonia cases are quite similar [46] in the CT image. Therefore, the discrimination between COVID-19 and other pneumonia cases would be of great importance in clinical practice [46], [47]. Thus, many researchers have recently looked into AI-based classification solutions. For example, Wang *et al.* [48] used transfer-learning by adapting the inception neural network for the classification of COVID-19 cases from the other viral pneumonia cases. They trained their model with a total of 217 pathogen-confirmed Region Of Interest (ROI) images that are extracted from 99 collected CT images (44 COVID-19 and 55 other viral pneumonia cases). With 236 ROI images as the testing set, the accuracy of 73.1%, a specificity of 67%, and sensitivity of 74% were achieved. Similarly, Ying *et al.* [49] developed a deep learning-based CT diagnosis and lesions localization system (named DeepPneumonia). Their proposed network contained a Details Relation Extraction neural Network (DRE-Net) to obtain image-level predictions and an aggregation step to get case-level labels. A total of 1990 CT images (obtained from 88 COVID-19 patients, 101 bacteria pneumonia patients, and 86 healthy cases) were used for training and testing. The network gave a result with an Area Under the Curve (AUC) of 0.92 in image-level and with an AUC of 0.95 and a recall of 0.96. Shi *et al.* [50] and Xu *et al.* [51] proposed screening models to distinguish COVID-19 out of community-acquired and Influenza-A viral pneumonia. Shi *et al.* [50] segmented CT images using VB-Net [52] (a modified version of V-net [39]) and extracted location-specific features from: volume, infected lesion number, histogram distribution, and surface area. Machine-learning methods were then applied to decide the best features and later predict COVID-19 patients from community-acquired pneumonia patients. Their results were based on 2685 CT images. Out of 2685, 1658 were confirmed COVID-19, and 1027 were pneumonia cases. They achieved a sensitivity of 0.907, the specificity of 0.833, and an accuracy of 0.879 under five-fold cross-validation. Xu *et al.* [51] employed multi CNN models with location-attention mechanism. A total of 618 CT samples (219 COVID-19, 224 Influenza-A viral pneumonia, and 175 healthy cases) were used to achieve an average F1-score of 0.856 for all the three categories.

C. SEVERITY ASSESSMENT OF COVID-19

Besides the identification of COVID-19 from other pneumonia cases, severity assessment has been a recent research focus. From a study on CT images of recovered COVID-19 patients, four stages of lung patterns were identified. The patterns, termed as early (0 to 4 days after the initial symptom), progressive (5 to 8 days), peak (9 to 13 days), and absorption stages (more than 14 days) [4] provide important evidence for the necessity of CT-based COVID-19 severity assessments. In this regard, Xiong *et al.* [53] analyzed 42 patients of COVID-19 with both the initial and follow-up CT images to assess the severity and progression of COVID-19. Correlations were evaluated among clinical, laboratory findings, and CT features. The linear regression analysis was used to identify the significant indicating variables for the severity progression of COVID-19. Additionally, another recent COVID-19 severity assessment based on the random forest method is proposed by Tang *et al.* [54]. The author's analysis and three-fold validation on their extracted 63 quantitative features of the chest CT images gave an accuracy of 0.875, a true positive rate of 0.933, and a true negative rate of 0.745.

Most of the diagnosis systems report pleasing results with high detection and classification accuracy. However, the majority of methods rely on fully supervised learning, both on volume level and slice level. Such supervised methods require time and resources of experts for data labeling. To address such issues, our proposed framework is based on a semi-supervised attention based network that performs slice level inference with only volume level data labels.

III. METHODOLOGY

The proposed framework is shown in Fig. 1. The approach is motivated by a semi-supervised capsule video endoscopy classification described in [12], [13] applied to CT volume. The framework processes the whole CT volume and performs four discrete steps in an end-to-end fashion. Initially, the individual slices of the CT scan are fed to the proposed ResNext+ network for extracting the spatial features. A brief description of the ResNext+ is given in Section III-B. Once the spatial features are extracted from the individual slice, the feature maps are given to the bidirectional LSTM. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

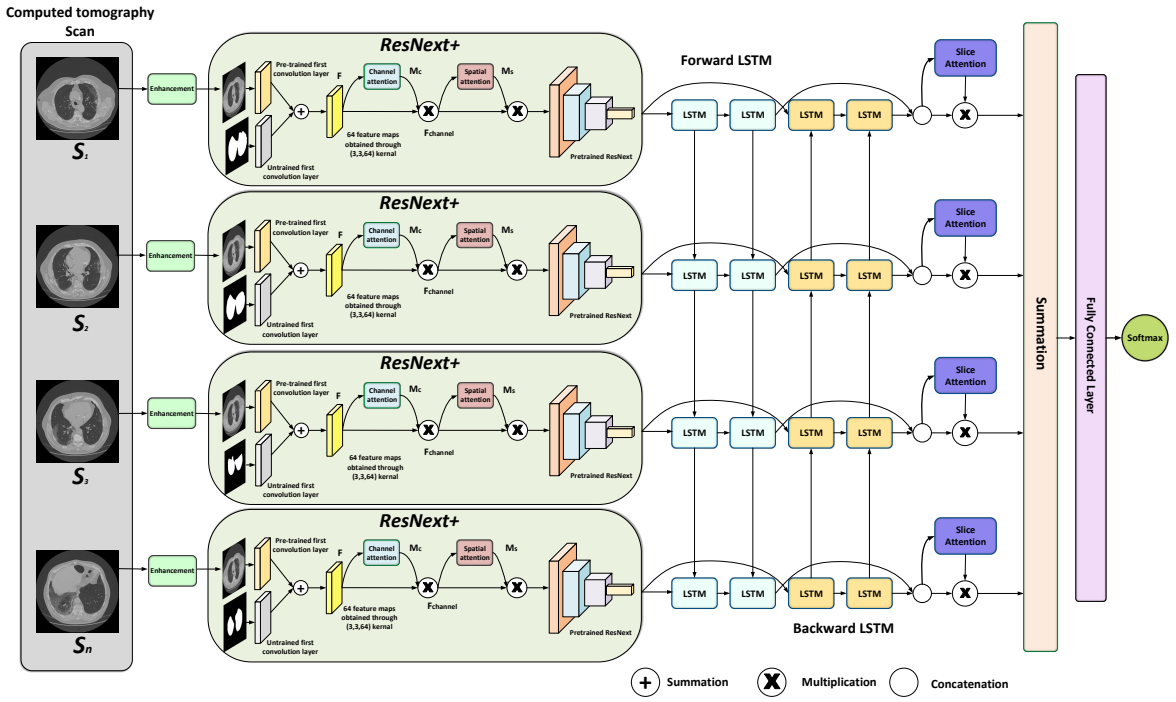


FIGURE 1: The CT slices are enhanced through the enhancement module and later spatial features are extracted through the ResNext+. The extracted spatial features are processed by bidirectional LSTM that transforms the spatial features to spatio-axial features and later refined by slice attention. The feature vector of each slice is summed and classified by a fully connected layer which is followed by a softmax that outputs underlying disease probabilities for the CT scan.

bidirectional LSTM exploits the axial dependency in the input slices and transforms the spatial features to spatio-axial features. A detailed description of the attention module is given in Section III-C. Each LSTM (forward and backward) gives a spatio-axial feature vector of dimension 1×512 which is concatenated and produces a spatio-axial feature vector of dimension 1×1024. As the feature maps are spatially refined by the channel and spatial attention in the ResNext+, the resulted spatio-axial features are refined by the slice's attention. Hence, in the third step of processing, the slice attention weights the importance of each slice for the inference. In the last step, a fully connected layer with 1024 hidden nodes and 2 output nodes classify the spatio-axial feature vector. To enhance the quality of the input slices, we adopted two enhancement techniques. A brief overview of the enhancement strategies is given in Section III-A.

A. ENHANCEMENT

Accurate image-based disease diagnosis requires high-quality image data. CT images sometimes have low contrast that may hamper the visualization of critical structures. Moreover, it also affects the performance of deep learning algorithms as low contrast and suppressed details can make feature extraction difficult [55]. This motivates to apply image enhancement to the data before inputting it to the network. Enhancement of the medical images has a twofold impact on any automated disease detection framework. Firstly, enhancing the visual quality of input data improves visualization of significant pathologists; secondly, it improves the performance of feature extraction and segmentation algorithms [56]. Therefore, in this work, two image enhancement strategies (stochastic and tone mapping) were evaluated with the motivation to improve the feature learning of the proposed technique.

A brief description of the two enhancement approaches is given in the stochastic enhancement (Section III-A1) and tone mapping (Section III-A2) sections. Visual inspection (Fig. 7) of the enhancement results reveals that the methods lead to a well-contrasted lung area from the nearby bones and non-lung tissues. Moreover, the details of the infectious area are also improved as can be seen in the right lung (top row). Since the main focus of this work is not on visual quality, we have only investigated the impacts of the two enhancement methods on the performance of the proposed COVID-Attention-Net in terms of sensitivity, specificity, accuracy, precision, and recall. The detailed experimental analysis of COVID-Attention-Net is presented in Section V.

1) Stochastic enhancement

One of the evaluated enhancement methods is the stochastic sampling-based image enhancement algorithm proposed by Mohammed *et al.* [25] that helps in highlighting the lung tissues and the bronchioles in the CT images. The approach explores the local neighborhood of a pixel in two capacities.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

The algorithm analyzes the intensity similarity between the target pixel and the neighboring pixels which are characterized by the gradient between target pixel, the neighboring pixel, and their intensity difference. First, the image is decomposed into two layers D_1 and D_2. The local lightness and darkness contrast image D_1 is approximated with stochastic sampling and image local details D_2 are computed locally through the random walk. The enhanced CT image is given by:

$$\begin{aligned} I_{enh} &= KD + I_{base} \\ D &= \gamma D_1 + (1-\gamma) D_2 \end{aligned} \tag{1}$$

where γ is a mixing coefficient that controls the amount of local details against image contrast, K is a scalar constant and I_{base} is the base layer.

To compute the base layer of the image I_{base}, for each pixel x_0 in the image, neighboring pixels are sampled with M number of random walk. The random walk sampling is initialized at x_0 pass through random neighboring pixels $\mathbf{x_j} = \{x_0, x_1, \dots x_n\}$ on the j_{th} random walk. The similarity of the target pixel x_0, to the neighboring pixels $\mathbf{x_j}|j \in M$ is expressed by a weighting function $\mathbf{w_0}^j(x_0|\mathbf{x_j})$ expressed as:

$$\mathbf{w_0}^j(x_0|\mathbf{x_j}) = \exp\left(-\frac{\|x_0 - \mathbf{x_j}\|_1}{|\sigma_I|} - \frac{\|TV(\nabla I)\|_1}{|\sigma_g|}\right) \tag{2}$$

where x_0 is the target pixel and $\mathbf{x_j}$ corresponds to the set of intensity values of the neighboring pixels on j_{th} random walk. Similarly, σ_I and σ_g are the normalization constant. Hence, the first term of the exponential represents the l_1 norm of the intensity difference between the initial pixel x_0 and neighboring pixel $\mathbf{x_j}$ normalized by the constant σ_I. The second term represents the total variation of eigenvalues of the structural tensors at each pixel normalized by the constant σ_g. The total variation term measures whether the random walk has crossed edges or not. This can be formulated using eigenvalues of the structural tensors λ_+ and λ_- at each pixel. Using similar notation, the random walk gradient sampling is initialized with eigenvalues $(\lambda_+^0, \lambda_-^0)$ at the target pixel x_0 pass through random neighboring pixels $\mathbf{x_j} = \{(\lambda_+^0, \lambda_-^0), (\lambda_+^1, \lambda_-^1), \dots (\lambda_+^n, \lambda_-^n)\}$ on the j_{th} random walk. Mathematically, the total variation term is defined as a sequence $TV(\nabla I) = \{TV(\nabla I)_0, TV(\nabla I)_1, \dots TV(\nabla I)_n\}$ where $TV(\nabla I)_n$ is given by:

$$TV(\nabla I)_n = \sum_{i=0}^{\mathbf{n}} \left(\sqrt{(\lambda_+^{i+1} - \lambda_-^{i+1})} - \sqrt{(\lambda_+^i - \lambda_-^i)}\right) \tag{3}$$

where λ's are the eigenvalues of the structural tensors at each pixel which captures the dominant orientation of all neighboring pixel $\mathbf{x_j}$. Finally, each pixel in the base layer is computed as:

$$x_{den} = \frac{\sum_{j=0}^{M} w_0{}^j(x_0|\mathbf{x_j}) x_0}{\sum_{j=0}^{M} w_0{}^j(x_0|\mathbf{x_j})} \tag{4}$$

Eq. 1 and Eq. 4 summarize our enhancement. A graphical depiction of random walks is illustrated in Fig. 2. We applied this enhancement on all the slices of the CT scan as a preprocessing step for the deep network.

2) Tone Mapping

The second approach we consider for CT image enhancement is through tone mapping operators. In some imaging conditions, the linear transformation of raw CT images (usually in 16-bit and high dynamic range format) to some of the common 8-bit (low dynamic range) image formats leads to loss of important image details. In different imaging applications such as high dynamic range image reproduction, several tone mapping, and contrast stretching operations need to be applied to compress the images' dynamic range, while selectively preserving important image details [57]. Tone mapping operators have shown to be useful for CT images [58]. Therefore, we have tested a combination of global gamma and sigmoidal tone mapping operators for the preservation and enhancement of contrast around the lung regions of the CT scans, during image format conversion.

The CT images are stored as Digital Imaging and Communications in Medicine (DICOM) format 16-bit greyscale images with the pixel intensity proportional to tissue density represented in Hounsfield Unit (HU). A predefined threshold value of -600 HU is typically used to locate lung tissue [59]. Since most of the lung regions are represented by the lower mid of the intensity levels, we have applied inverse gamma followed by a sigmoid contrast enhancement function as given below:

$$I_{out} = \frac{1}{1 + e^{-a(I_{in})^{1/\gamma}}} \tag{5}$$

In Eq. 5, the inverse gamma is $\gamma = 1.5$ and $a = 0.35$. As it can be seen from the resulting images, shown in Fig. 7, the two operations globally scaled the lightness value of the images in such a way that the darker regions (mainly lungs) of the images remain enhanced while suppressing the brighter regions (bones and other related organs) [60].

B. RESNEXT+

The architecture of ResNext+ is inspired by the classical ResNext [61]. However, it is different from classical ResNext in two ways. First, ResNext+ is capable of fusing the original slice with the corresponding binary mask, in our case of the lung region, in the first layer of the network. For the fusion, we used both the pre-trained and the untrained convolution layer of classical ResNext. The choice is motivated by the fact that ResNext is originally trained on the Imagenet dataset which consists of natural images. So, it is logical to use an untrained layer for the binary mask and a pre-trained layer for the slice. The second key attribute of the ResNext+ comes from the introduction of the channel and spatial attention. Channel and spatial attention have shown substantial improvement in several vision problems [62]. The key idea of the attention module is to refine the feature map and to give

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

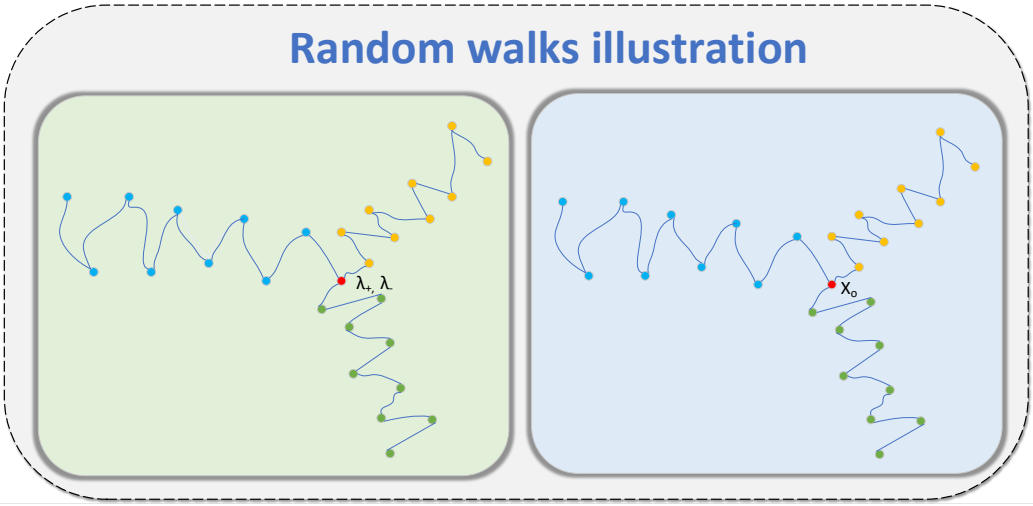


FIGURE 2: Illustration random walk: For each target pixel x_0, a random walk is initialized to compute intensity similarity and total variation of the gradient (λ_-, λ_+) along the random walk neighboring pixels $\mathbf{x_j}$. For clarity, in the figure, the number of iterations is $n = 3$, while the number of samples is $M = 9$.

consciousness to the network regarding the important regions in the slices for the inference. A detailed description of the attention module is given in Section III-C.

C. ATTENTION MODULE

Designing a deep network with high performance and few parameters is one of the goals of the researchers in the community. Primarily, the most intuitive ideas like increasing the depth [63], [64], and width [43], [65] of the network is a well-adopted trend. However, the focus is shifting to the cardinality [61], [66] and the attention mechanism. Attention is mainly inspired by the human visual system. It is a relatively new term that is applied to deep models for improving the representation capability of the network and also helped the network to focus on the most important features. In our work, we exploited the Convolutional Block Attention Module (CBAM) [62] for fusing the cross-channel and spatial information in a given slice. Unlike [62], in our proposed method, cross-channel and spatial attention is applied only after the fusion of the slice and the binary mask. The CBAM improves the information flow from the layers of the network which consequently helps in information accentuation or suppression and as a result, gives a better representation for the infection prediction. For a given set of 64 feature maps $F \in \mathbb{R}^{C\times H\times W}$, the attention module extracts a 1D channel attention map $M_c \in \mathbb{R}^{C\times 1\times 1}$ and a 2D spatial attention map $M_s \in \mathbb{R}^{1\times H\times W}$ as shown in Fig. 3 and Fig. 4, respectively. Mathematically, it can be represented as:

$$F_{channel} = M_c(F) \otimes F \tag{6}$$

$$F_{spatial} = M_s(F_{channel}) \otimes F_{channel} \tag{7}$$

where F is the set of feature maps obtained after applying the first convolution and fusion (1), $F_{channel}$ is the channel attention feature maps and $F_{spatial}$ is the refined spatial attention feature maps. $\otimes$ indicates the element-wise multiplication.

1) Channel attention

The basic idea of channel attention is to find out what are the most important feature maps in the input volume. For the channel attention, we followed a similar formulation to that of [62] and used average pooling and max-pooling for squeezing the spatial dimension of the input feature maps. The averaged pooled F^c_{avg} and max-pooled F^c_{max} features are forwarded to a fully connected Multi-Layer Perceptron (MLP) with one hidden layer that generate the channel attention map $M_c \in \mathbb{R}^{C\times 1\times 1}$. The channel attention mechanism can be summarized as:

$$M_c(F) = \sigma(MLP(Avg_{pool}(F)) + MLP(Max_{pool}(F))) \tag{8}$$

$F \in \mathbb{R}^{C\times H\times W}$ is the feature maps obtained through the CNN while Avg_{pool} and Max_{pool} are the average and max pooling operations, respectively.

$$M_c(F) = \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max}))) \tag{9}$$

The Sigmoid function σ is used as the main activation function for the channel attention module. $W_0 \in \mathbb{R}^{C/r\times C}$ and $W_1 \in \mathbb{R}^{C\times C/r}$ are the input to hidden layer and hidden layer to output weight parameter for the MLP. For keeping the parameters of the MLP small, the hidden layer activation size is set to $\mathbb{R}^{C/r\times 1\times 1}$ with r as the reduction ratio.

2) Spatial attention

Compared to channel attention, spatial attention aims to localize the most informative part of the feature maps that's complementary to channel attention. To calculate the spatial attention, first average pooling and max pooling operations

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

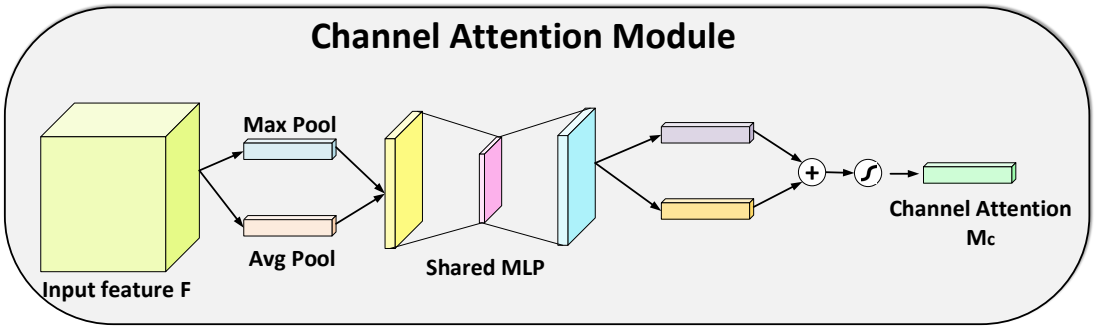


FIGURE 3: Illustration of Channel Attention Module

are applied to the feature maps and then the resulting feature maps are concatenated to get an efficient feature descriptor. On the resulting feature descriptor, a convolution layer is applied to generate the spatial attention map $M_s(F) \in \mathbb{R}^{H \times W}$. Mathematically, it can be defined as:

$$M_s(F) = \sigma(Conv^{9\times9}([Avg_{pool}(F); Max_{pool}(F)])) \quad (10)$$

$$M_s(F) = \sigma(Conv^{9\times9}([F^s_{avg}; F^s_{max}])) \quad (11)$$

where $F^s_{avg} \in \mathbb{R}^{1\times H\times W}$ and $F^s_{max} \in \mathbb{R}^{1\times H\times W}$ are the average and max pooling, respectively. The Sigmoid function σ is used as the main activation function, while $Conv^{9\times9}$ shows the convolution operation with a filter size of 9×9. The refined feature map is used as the slice descriptor and given to the bidirectional LSTM as the input.

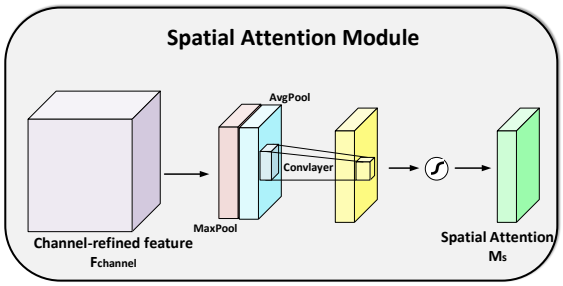


FIGURE 4: Illustration of Spatial Attention Module

IV. LONG-SHORT TERM MEMORY

A feed-forward neural network [67] output solely based on the input data. In contrast, a recurrent neural network [68] has an internal memory where it stores the results of the previous samples. Hence, in the recurrent network, the output at any time instant t not only depends on the input but also on the previous outputs of the network. Long-Short Term Memory (LSTM) is a special type of recurrent neural network that can retain past information for a longer period. The LSTM uses gates that can be seen as the information gateway that allows how much information can flow to the cell state through a sigmoid/hyperbolic tangent activation and a point-wise multiplication operation. The state of the LSTM cell is essentially the mechanism of storing the previous knowledge and a way of propagating only the useful knowledge to the next cells in the network. In an LSTM cell, the first two steps are related to calculating the information that needs to be kept in the cell state and the information that needs to be thrown away. Initially, the forget gate value is calculated as:

$$f_t = \sigma(W_f.[h_{t-1,x_t}] + b_f), \quad (12)$$

where W_f and b_f are the learnable parameters and x_t and h_{t-1} are the input and the output of the previous state. Similarly, the input gate value is calculated as:

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i), \quad (13)$$

with W_i and b_i the learnable parameter of the LSTM cell. Furthermore, an intermediate state value of the cell is also calculated from the current input x_t and previous output h_{t-1} and the learnable parameters W_c and b_c as:

$$\tilde{C}_t = \tanh(W_c.[h_{t-1}, x_t] + b_c), \quad (14)$$

$\tilde{C}_t$ can be seen a raw state value that would be refined through the forget gate f_t and the the input gate i_t values at time t as:

$$C_t = f_t * c_{t-1} + i_t * \tilde{C}_t, \quad (15)$$

Once the updated value of the state C at time t is calculated, the final output by the LSTM cell can be estimated in two steps. First an intermediate quantity o_t as:

$$o_t = \sigma(W_o[h_{h-1}, x_t] + b_o), \quad (16)$$

And based on o_t and C_t, the final output is determined as:

$$h_t = o_t * \tanh(C_t) \quad (17)$$

The h_t and C_t act like the previous output and previous state of the next LSTM cell in the LSTM network. Bidirectional LSTMs are an extension to the LSTM network that enhances performance by feeding CT slices to two independent LSTM networks in forward and backward direction along the axial

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

axis and concatenating the output features. Further analysis of LSTM is beyond the scope of the paper. For further details, the readers may refer to [69]–[72].

1) Slice attention

Slice attention is the mechanism of enabling the network to focus mainly on the semantic slices to assist in modeling the axial dependencies in the data. The slice attention principle is similar to softmax function where the values are normalized and the sum is equal to 1. However, the slice attention of the individual slices shows the probability of slice having COVID-19. From the architectural point of view, slice attention is modeled as a two-layer fully connected neural network that takes the output of the bidirectional LSTM and gives the slice attention score. In our setting, the output of both the forward and backward LSTM is 1×512 feature vectors which yield a feature vector of size 1×1024 after concatenation. Mathematically, the slice attention is expressed as:

$$\mathcal{S} = \sum_{n=1}^{N} \alpha_n f_n \tag{18}$$

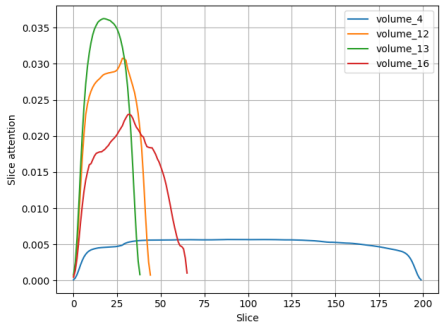


FIGURE 5: Response of slice attention on different CT volumes.

where α_n is the attention response, f_n is the input feature vector and N is the number of slices considered for the inference. Given that the slice attention is a two-layer fully connected neural network, the attention response is obtained as:

$$\alpha_n = \frac{\exp(w^T \tanh(bf_n^T))}{\sum_{n=1}^{N} \exp(w^T \tanh(bf_n^T))} \tag{19}$$

In Eq. 19, w and b are the parameters of the two-layer network and N is the total number of slices. During the training, given the $\tanh$ activation function works for both the negative and positive values, the gradient of the cost function is back-propagated efficiently. Hence, it can be seen in Fig. 5 that slice attention is producing an effective response by computing an adaptive weighted average of the bidirectional LSTM features. The accumulated slice attention response $\mathcal{S}$ is passed through a fully connected layer with 1024 hidden nodes and two output nodes that yields v_i; the volume response, that is consequentially given to a 2-way softmax function for the final class probability inference.

V. EXPERIMENTS

A set of experiments has been performed to evaluate the performance of the proposed network without applying any kind of enhancement as well as after applying enhancement as described in the prior section. The experimental details, dataset, and underlying pre-processing steps are presented as follows.

A. DATASET

To train and test the proposed COVID-Attention-Net, we used a total of 302 CT volumes (20 with confirmed COVID-19 patients) consisting of a total 3520 positive and 19,353 negative cases slices. The positive and negative CT data is acquired from Joseph [1] *et al.* [73] (collected from few Chinese, Iranian and Italian hospitals) and Tianchi Lung diseases diagnosis competition CT images [2],[3], respectively. Two radiologists assisted in the manual annotation of 20 positive cases both at volume and slice level, though the slice level annotations are only used for performance evaluation purposes. The dataset is finally split into training and testing set with the random 80/20 ratio and the difficulty level of both sets are confirmed to be balanced by the radiologist. The CT scans, originally existing in mhd or nifty formats, are linearly transformed to the standard grayscale intensity range. The positive and negative data were acquired from different sources and both had different dimensions, therefore all the 2D slices were resized to 256×256 to ensure the same spatial dimension for the whole input data.

B. LUNG MASK EXTRACTION

Most of the time, chest CT images contain lung and non-lung tissues such as bones and fat. Since COVID-19 effects can only be viewed in the lung region, we incorporated the lung mask in the first stages of the proposed network, as shown in Section III. Hence, the lung mask is extracted by binarizing the CT slices with a threshold of -600 HU, adapted from Liao *et al.* [74]. Then convex hull of the mask is computed, followed by a dilation operation to include the outer wall of the lung. The original CT slice and the corresponding estimated mask are shown in the first and second row of Fig. 6. For some challenging volumes, such as the lung part containing severe pathologies, the binarizing method may fail to segment the lung part. Therefore, those severe and wrongly segmented cases are manually checked and removed from the training dataset.

C. CLASS REBALANCING AND DATA AUGMENTATION

As described earlier, the number of positive (20) and negative (282) CT volumes are not balanced. Such type of long-tailed data distribution is a frequently encountered issue in

[1]https://academictorrents.com/details/136ffddd0959108becb2b3a86630bec049fcb0ff (Accessed: 10 July 2020)

[2]https://tianchi.aliyun.com/competition/entrance/231724/introduction (Accessed: 10 July 2020)

[3]https://aistudio.baidu.com/aistudio/datasetdetail/8689 (Accessed: 10 July 2020)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

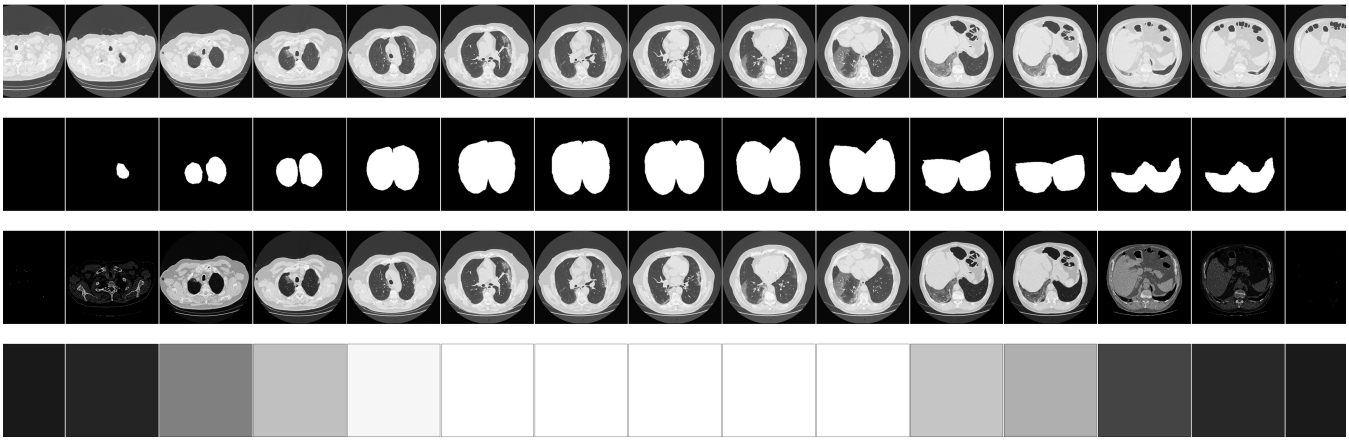


FIGURE 6: The top two rows show the original slices with their corresponding mask for Volume 12 in the test set. The third row shows the output of the slice attention combined with the original slice encoded as $(I^{0.001+\frac{1}{\alpha_n}})$ for better visualization, where α_n is given by Eq. (19). The last row shows grayscale coded slice with black the irrelevant slices and white the relevant slices. Attention slices vary from Blank (Black) to the original frame corresponding to low and high value of attention weights, respectively. As it can be seen the network is able to localize slices containing COVID-19 slices (from the fourth column to the penultimate one).

several classification problems [75]. It adversely affects the sensitivity of the detection algorithm and consequently, the network either wrongly identifies majority true positives as false positive or true negative as a false negative. Class rebalancing strategies including resampling, oversampling, and cost-sensitive reweighting have been incorporated in several classification methods to resolve the class imbalance challenge and considerably improve the performance of an algorithm [76]. Resampling operates on data level and modifies the class distribution of training data. Considering the extreme imbalance between the quantity of negative and positive instances, we incorporate resampling strategy in case of positive instances. To further minimize the skew in data distribution, we later apply data augmentation by introducing invariance in the positive samples. Intensity transformations including contrast stretching, the addition of Gaussian noise, blur, and spatial transformations such as zooming, scaling, rotation, and elastic deformation is applied to augment positive sample count. In this way, we increase the intra and interclass disparity in our dataset.

D. IMPLEMENTATION DETAILS

Our method is implemented with PyTorch library [77] and trained on a single NVIDIA TITAN RTX GPU with 24GB graphic memory. For the stochastic image enhancement, the number of iterations n are selected to be 20 while the number number of samples M in each iteration is fixed at 250. The normalization constant σ_I and σ_g are chosen empirically and fixed at 0.5 and 0.3, respectively. For all our experimental scenarios, we used pre-trained ResNext [61] convolutional layers to extract features from slices. The slices and their corresponding masks are resized to 224×224. We applied a 7×7 convolution to the slices and their corresponding masks before summation as shown in Fig. 1. The network is trained end-to-end with binary cross-entropy Eq. 20 and batch-size of 1, as each CT volume contains a variable number of slices. The Adam optimizer [78] with cyclic learning rate scheduling technique of $lr_{min} = 1e^{-5}$ and $lr_{max} = 1e^{-4}$ values is used for all our training [79]. The LSTM blocks are initialized from a normal distribution with 0 mean and 0.01 variance. We have also disabled all batch normalization layers running estimates and trained each experimental case for a total of 40 epochs.

$$l(g, p) = -g \log(p) - (1 - g) \log(1 - p) \tag{20}$$

where $p = \frac{\exp(-v_i)}{\sum_{i=0}^{1} \exp(-v_i)}$ and v_i is the output of the last fully connected layer and g is the ground truth label.

Since the number of slices in each volume of our dataset ranges from 30 to 350 and due to the limited availability of single GPU memory, the number of input slices is restricted to a maximum of 50. Therefore, random sampling is done in our implementation to input 50 slices from a particular volume to feed the network.

E. EXPERIMENTAL RESULTS

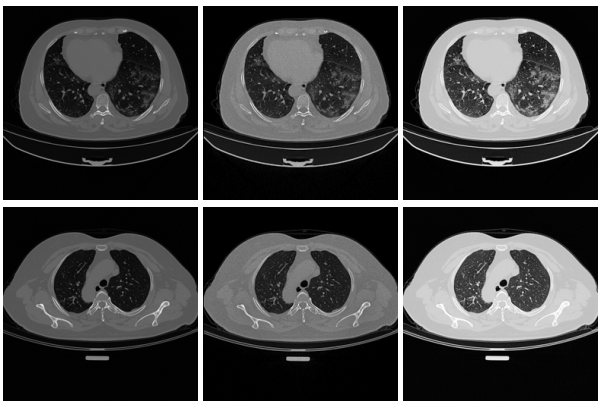
To evaluate the performance of the proposed detection method, six commonly used evaluation metrics (Accuracy (ACC), Precision (PRE), F1-score (F1), Sensitivity (SEN), and Specificity (SPE)) computed from the confusion matrix between the ground-truth labels and the predicted labels are used. The performance is evaluated both on the volume and slice level. In order to analyze the role of each block of the proposed method, the following comparisons are done. First, the method is evaluated by modifying the enhancement block. In this regard, three experiments are implemented to analyze the performance of the network in terms of assessment metrics mentioned above by inputting data

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

without applying any enhancement, after applying stochastic enhancement [25] and after applying tone mapping. Then, an ablation study of the network is conducted to evaluate the improvement in the performance of the network. The detail of the experiments and results are discussed and analyzed in the following section.

F. ANALYSIS OF IMAGE ENHANCEMENT

Stochastic enhancement and tone mapping, described in Section III-A, are applied to the original images to highlight details inside the lung area and therefore provide the network with more information. As shown in Fig. 7, after enhancement, the area of interest in the image can be seen more clearly without the introduction of any additional artifacts, especially in the COVID-19 infected lung area (first row).



(a) (b) (c)

FIGURE 7: Qualitative results of the enhancement methods: (a) Original image. (b)Image enhanced by stochastic enhancement method [25]. (c) Image enhanced by tone mapping.

Table 1 shows the detection results on the original and enhanced CT images, both on volume level and slice level. On volume level evaluation, images with or without enhancement all achieve a 100% performance in all the assessment metrics, which means that all positive cases are correctly detected. However, on the slice level, the sensitivity after stochastic enhancement improves by 2.8%, while the values of other evaluation metrics are all dropped compared with the original images. There could be two possibilities; our model parameters are fine-tuned only on the original images instead of fine-tuning it separately for the original and enhanced images as input. The second could be that the enhancement is applied on whole slices instead of the lung area (the region of interest (ROI) in our case), which may enhance the non-lung areas, leading the incorrect image regions to receive more attention.

G. ABLATION STUDY OF THE NETWORK

To illustrate the effectiveness of each module included in the network, an ablation study is performed by removing some modules from the network while retaining the rest. We experimented with three configurations in order to analyse the impact of different components on the performance of the proposed network:

a). excluding spatial and channel attention (N-SCA)
b). excluding slice attention (N-SLA)
c). excluding the segmentation mask (N-MSCA).

The spatial and channel attention modules of the network are first removed from the original model and the results indicate that the network can precisely predict volume level COVID-19 with the proposed model and without spatial and channel attention (N-SCA). The spatial and channel attention modules can boost the slice level performance of the network by 2.17%, 5.2%, 2.6%, and 6.0% on ACC, PRE, F1, and SPE respectively. The scores of SEN also shows a similar trend. This can be observed by comparing the results of proposed and N-SCA configurations in Table 2.

When the lung segmentation mask is further removed from the proposed network in case of N-MSCA configuration (i.e. without segmentation mask, spatial and channel attention), the network performance drops to 93.3% for ACC, 0 for PRE, F1 and, SEN which indicates that the network classifies all the positive volumes as negatives. In other words, N-MSCA can not detect the COVID-19 in the overall volume. It is worth pointing out that the N-MSCA obtains a comparably better result for the slice level by locating the important slices. However, even the ability of the network to focus better on the slice level does not contribute to the correct final prediction on the volume level. This experiment demonstrates the importance of the segmentation mask. Moreover, it emphasizes the significance of the spatial and channel attention modules in the proposed architecture as well.

In order to investigate the influence of the slice attention module on volume level, we conduct the third experiment (N-SLA configuration), where the slice attention of the network is removed from the proposed setup. The network obtained the same performance for volume level evaluation. However, without the slice attention module, it is not possible to localize the slices that contain COVID-19, which reduces the explainability of such approaches.

VI. DISCUSSION

In this work, we proposed a deep learning-based end-to-end framework that not only gives a volume level detection, but is also capable of classifying slices containing COVID-19 infection. Furthermore, we present the first study on COVID-19 detection employing semi-supervised network using volume level labels to achieve slice level prediction.

A. EFFECTIVENESS AND APPLICATION

We demonstrate the effectiveness of our method on volume level and slice level prediction. Considering volume level diagnosis, we attain 100% performance on all evaluation metrics (with or without enhancement), implying that the proposed method can correctly detect all the COVID-19 cases on our test data. For slice level attention, the labeled

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3018498, IEEE Access

TABLE 1: Detection results of different enhancement methods.

		ACC	PRE	REC	F1	SEN	SPE
Volume Level Evaluation	Orig.	1.0	1.0	1.0	1.0	1.0	1.0
	Stoc. Enh. [25]	1.0	1.0	1.0	1.0	1.0	1.0
	Tone Enh.	1.0	1.0	1.0	1.0	1.0	1.0
Slice Level Evaluation	Orig.	0.776	0.819	0.854	0.814	0.855	0.793
	Stoc. Enh. [25]	0.762	0.792	0.882	0.809	0.883	0.729
	Tone Enh.	0.729	0.788	0.773	0.761	0.773	0.772

TABLE 2: Results of ablation study of the proposed approach. N-SCA: without spatial and channel attention; N-SLA: without slice attention; N-MSCA: without segmentation mask, spatial and channel attention.

		ACC	PRE	F1	SEN	SPE
Volume Level Evaluation	N-SCA	1.0	1.0	1.0	1.0	1.0
	N-MSCA	0.933	0.00	0.00	0.00	1.0
	N-SLA	1.0	1.0	1.0	1.0	1.0
	Proposed	1.0	1.0	1.0	1.0	1.0
Slice Level Evaluation	N-SCA	0.755	0.767	0.788	0.856	0.733
	N-MSCA	**0.788**	0.763	**0.834**	**0.975**	0.626
	N-SLA	NA	NA	NA	NA	NA
	Proposed	0.776	**0.819**	0.814	0.855	**0.793**

data is used just for validation, which means that the model is unsupervised for making slice level prediction. Nevertheless, promising results are obtained with the proposed model, indicating that the attention modules can help to locate the more suspicious slices.

Our main goal is to assist the doctors in the diagnosis of COVID-19, so the application can be beneficial from two perspectives: First, at the volume level, the proposed network can give a pre-diagnosis for the doctors to identify the individual/overall suspected COVID-19 cases. Second, at the slice level, the slice attention can allow the doctors to only focus on the sensitive slices that are candidates of containing the COVID-19 infection instead of examining the entire volume.

B. LIMITATIONS AND FUTURE WORK

While the proposed approach shows an encouraging performance, there are several limitations regarding our dataset and methodology.

- Dataset limitations: The dataset in our study does not include common and other viral pneumonia, which is also important for COVID-19 detection. There are fewer COVID-19 positive cases compared to the negative cases which lead the dataset to class imbalance issues. We conducted re-sampling on the slice level to weaken the imbalance, but still, this problem introduces challenges on training and evaluation of the network.
- Methodology limitations: COVID-19 detection is a new emerging research field. Therefore, there is no standard dataset publicly available. Thus, the comparison of the proposed technique with state-of-the-art is not currently feasible. The evaluated stochastic enhancement is also applied to the gray-scale images, which may introduce information loss in the quantization step. Additionally, our network parameters are only fine-tuned on original images, which may not fully demonstrate the gain in network performance realized by including enhancement methods.

Considering such limitations of this study; we plan to improve our model in two ways in the future.

- Data acquisition and labeling: More COVID-19 and other pneumonia cases will be labeled and added to the dataset to demonstrate the robustness of our model and improve the data imbalance issue. Moreover, having a larger dataset will improve the attention. For example, narrower slice attention will be achieved with a larger dataset compared to the Gaussian like slice attention that is achieved from the current dataset. Narrower slice attention will help the doctor to pinpoint the slices with COVID-19 infection.
- Methodology improvement: Further investigation will be done regarding the application of enhancement on original images instead of re-scaled ones. A Comparison with the state-of-the-art will also be conducted once a standard public dataset is available.
- In the current study, the masks are annotated manually for training the network. However, with the availability of standard public datasets, the segmentation task can be learned and trained in an end-to-end fashion.

VII. CONCLUSION

A semi-supervised deep learning-based framework for COVID-19 detection is proposed in this paper. The proposed framework use combination of lung segmentation mask, attention aware mechanism, and LSTM for extracting the spatial, axial, and temporal features from the CT volume. Initially, resampling accompanied by data augmentation techniques is applied to address the scarcity and imbalance of binary class data distribution. As a pre-processing step, stochastic and tone-mapping based image enhancement methods were evaluated for performance improvement of the model. Finally, the performance evaluation of the proposed framework is conducted using several module configurations.

The ablation study shows that the combination of all the attention modules and the segmentation mask yields the best performance. On volume level prediction, the proposed method achieved a 100% performance on all evaluation metrics and experimental cases. For slice level prediction, however, a different performance was observed in different experimental cases. In general, the integration of slice attention enables radiologists to concentrate only on the salient areas of the whole CT volume. From clinical perspectives, the proposed framework can facilitate the prognosis of COVID-19 by radiologists. Moreover, it paves the way for future research targeted at COVID-19 detection from limited and weakly labeled data.

REFERENCES

[1] Johns Hopkins University, "Coronavirus COVID-19 global cases by the center for systems science and engineering (CSSE)," https://coronavirus.jhu.edu/map.html, 2020, Online. Visited April 2, 2020.

[2] Centers for Disease Control and Prevention, "Testing for COVID-19," https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html, 2020, Online. Visited April 2, 2020.

[3] T Liang, "Handbook of COVID-19 prevention and treatment," 2020.

[4] Feng Pan, Tianhe Ye, Peng Sun, Shan Gui, Bo Liang, Lingli Li, Dandan Zheng, Jiazheng Wang, Richard L Hesketh, Lian Yang, et al., "Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia," Radiology, p. 200370, 2020.

[5] Sana Salehi, Aidin Abedi, Sudheer Balakrishnan, and Ali Gholamrezanezhad, "Coronavirus disease 2019 (COVID-19): A systematic review of imaging findings in 919 patients," American Journal of Roentgenology, pp. 1–7, 2020.

[6] Ying-Hui Jin, Lin Cai, Zhen-Shun Cheng, Hong Cheng, Tong Deng, Yi-Pin Fan, Cheng Fang, Di Huang, Lu-Qi Huang, Qiao Huang, et al., "A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version)," Military Medical Research, vol. 7, no. 1, pp. 4, 2020.

[7] Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang, and Wenbin Ji, "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," Radiology, p. 200432, 2020.

[8] Feng Shi, Jun Wang, Jun Shi, Ziyan Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, and Dinggang Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," arXiv preprint arXiv:2004.02731, 2020.

[9] Rongheng Lin, Zezhou Ye, Hao Wang, and Budan Wu, "Chronic diseases and health monitoring big data: A survey," IEEE reviews in biomedical engineering, vol. 11, pp. 275–288, 2018.

[10] Joseph Bullock, Katherine Hoffmann Pham, Cynthia Sin Nga Lam, Miguel Luengo-Oroz, et al., "Mapping the landscape of artificial intelligence applications against COVID-19," arXiv preprint arXiv:2003.11336, 2020.

[11] Veronika Cheplygina, Marleen de Bruijne, and Josien PW Pluim, "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," Medical image analysis, vol. 54, pp. 280–296, 2019.

[12] Ahmed Mohammed, Ivar Farup, Marius Pedersen, Sule Yildirim, and Øistein Hovde, "Ps-devcem: Pathology-sensitive deep learning model for video capsule endoscopy based on weakly labeled data," under review, 2020.

[13] Ahmed Kedir Mohammed, Computational Techniques for Pathology Detection and Quality Enhancement with emphasis on Colonic Capsule Endoscopy, Ph.D. thesis, 2019.

[14] Ahmed Kedir, Mohib Ullah, and Jacob Renzo Bauer, "Spectranet: A deep model for skin oxygenation measurement from multi-spectral data," Electronic Imaging, 2020.

[15] United Imaging, "United imaging sends out more than 100 CT scanners and X-ray machines to aid diagnosis of the coronavirus," https://www.itnonline.com/content, 2020, Online. Visited April 8, 2020.

[16] Ioannis D Apostolopoulos and Tzani A Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," Physical and Engineering Sciences in Medicine, p. 1, 2020.

[17] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct scans," arXiv preprint arXiv:2004.14133, 2020.

[18] Xi Ouyang, Jiayu Huo, Liming Xia, Fei Shan, Jun Liu, Zhanhao Mo, Fuhua Yan, Zhongxiang Ding, Qi Yang, Bin Song, et al., "Dual-sampling attention network for diagnosis of covid-19 from community acquired pneumonia," arXiv preprint arXiv:2005.02690, 2020.

[19] Jun Wang, Yiming Bao, Yaofeng Wen, Hongbing Lu, Hu Luo, Yunfei Xiang, Xiaoming Li, Chen Liu, and Dahong Qian, "Prior-attention residual learning for more discriminative covid-19 screening in ct images," IEEE Transactions on Medical Imaging, 2020.

[20] Kelei He, Wei Zhao, Xingzhi Xie, Wen Ji, Mingxia Liu, Zhenyu Tang, Feng Shi, Yang Gao, Jun Liu, Junfeng Zhang, et al., "Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of covid-19 in ct images," arXiv preprint arXiv:2005.03832, 2020.

[21] Shaoping Hu, Yuan Gao, Zhangming Niu, Yinghui Jiang, Lao Li, Xianglu Xiao, Minhao Wang, Evandro Fei Fang, Wade Menpes-Smith, Jun Xia, et al., "Weakly supervised deep learning for covid-19 infection detection and classification from ct images," arXiv preprint arXiv:2004.06689, 2020.

[22] D Buonsenso, A Piano, F Raffaelli, N Bonadia, K De Gaetano Donati, and F Franceschi, "Novel coronavirus disease-19 pnemoniae: a case report and potential applications during covid-19 outbreak," European Review for Medical and Pharmacological Sciences, vol. 24, pp. 2776–2780, 2020.

[23] Subhankar Roy, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, Iris Huijben, Nishith Chennakeshava, Federico Mento, Alessandro Sentelli, et al., "Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound," IEEE Transactions on Medical Imaging, 2020.

[24] Shiba Kuanar, KR Rao, Dwarikanath Mahapatra, and Monalisa Bilas, "Night time haze and glow removal using deep dilated convolutional network," arXiv preprint arXiv:1902.00855, 2019.

[25] Ahmed Mohammed, Ivar Farup, Marius Pedersen, Øistein Hovde, and Sule Yildirim Yayilgan, "Stochastic capsule endoscopy image enhancement," Journal of Imaging, vol. 4, no. 6, pp. 75, 2018.

[26] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J Van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I Sánchez, and Bram van Ginneken, "Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1160–1169, 2016.

[27] Hongtao Xie, Dongbao Yang, Nannan Sun, Zhineng Chen, and Yongdong Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," Pattern Recognition, vol. 85, pp. 109–119, 2019.

[28] Ross Girshick, "Fast R-CNN," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.

[29] Sarah E Gerard, Taylor J Patton, Gary E Christensen, John E Bayouth, and Joseph M Reinhardt, "FissureNet: A deep learning approach for pulmonary fissure detection in CT images," IEEE transactions on medical imaging, vol. 38, no. 1, pp. 156–166, 2018.

[30] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[31] Yue Zhang, Jiong Wu, Wanli Chen, Yifan Chen, and Xiaoying Tang, "Prostate segmentation using z-net," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, pp. 11–14.

[32] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas De Bel, Moira SN Berens, Cas van den Bogaard, Piergiorgio Cerello, Hao Chen, Qi Dou, Maria Evelina Fantacci, Bram Geurts, et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge," Medical image analysis, vol. 42, pp. 1–13, 2017.

[33] Alibaba Cloud, "CT image analytics for COVID-19," https://www.alibabacloud.com/zh/solutions/ct-image-analytics, 2020, Online. Visited April 8, 2020.

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[35] Jun Chen, Lianlian Wu, Jun Zhang, Liang Zhang, Dexin Gong, Yilin Zhao, Shan Hu, Yonggui Wang, Xiao Hu, Biqing Zheng, et al., "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study," medRxiv, 2020.

[36] Chuansheng Zheng, Xianbo Deng, Qing Fu, Qiang Zhou, Jiapei Feng, Hui Ma, Wenyu Liu, and Xinggang Wang, "Deep learning-based detection for COVID-19 from chest CT using weak label," medRxiv, 2020.

[37] Shuo Jin, Bo Wang, Haibo Xu, Chuan Luo, Lai Wei, Wei Zhao, Xuexue Hou, Wenshuo Ma, Zhengqing Xu, Zhuozhao Zheng, et al., "AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical ai system in four weeks," medRxiv, 2020.

[38] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11. Springer, 2018.

[39] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 Fourth International Conference on 3D Vision. IEEE, 2016, pp. 565–571.

[40] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

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

[42] Yunpeng Chen, Jianan Li, Huaxin Xiao, Xiaojie Jin, Shuicheng Yan, and Jiashi Feng, "Dual path networks," in Advances in neural information processing systems, 2017, pp. 4467–4475.

[43] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

[44] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang, "Residual attention network for image classification," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2017, pp. 3156–3164.

[45] Cheng Jin, Weixiang Chen, Yukun Cao, Zhanwei Xu, Xin Zhang, Lei Deng, Chuansheng Zheng, Jie Zhou, Heshui Shi, and Jianjiang Feng, "Development and evaluation of an AI system for COVID-19 diagnosis," medRxiv, 2020.

[46] Heshui Shi, Xiaoyu Han, Nanchuan Jiang, Yukun Cao, Osamah Alwalid, Jin Gu, Yanqing Fan, and Chuansheng Zheng, "Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study," The Lancet Infectious Diseases, 2020.

[47] Wei Zhao, Zheng Zhong, Xingzhi Xie, Qizhi Yu, and Jun Liu, "Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study," American Journal of Roentgenology, pp. 1–6, 2020.

[48] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, et al., "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)," medRxiv, 2020.

[49] Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Huiying Zhao, Yusheng Jie, Ruixuan Wang, et al., "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," medRxiv, 2020.

[50] Feng Shi, Liming Xia, Fei Shan, Dijia Wu, Ying Wei, Huan Yuan, Huiting Jiang, Yaozong Gao, He Sui, and Dinggang Shen, "Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification," arXiv preprint arXiv:2003.09860, 2020.

[51] Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, Shuangzhi Lv, Liang Yu, Yanfei Chen, Junwei Su, Guanjing Lang, et al., "Deep learning system to screen coronavirus disease 2019 pneumonia," arXiv preprint arXiv:2002.09334, 2020.

[52] Fei Shan, Yaozong Gao, Jun Wang, Weiya Shi, Nannan Shi, Miaofei Han, Zhong Xue, Dinggang Shen, and Yuxin Shi, "Lung infection quantification of COVID-19 in CT images with deep learning," arXiv preprint arXiv:2003.04655, 2020.

[53] Ying Xiong, Dong Sun, Yao Liu, Yanqing Fan, Lingyun Zhao, Xiaoming Li, and Wenzhen Zhu, "Clinical and high-resolution ct features of the covid-19 infection: Comparison of the initial and follow-up changes," Investigative radiology, 2020.

[54] Zhenyu Tang, Wei Zhao, Xingzhi Xie, Zheng Zhong, Feng Shi, Jun Liu, and Dinggang Shen, "Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest ct images," arXiv preprint arXiv:2003.11988, 2020.

[55] Samuel Dodge and Lina Karam, "Understanding how image quality affects deep neural networks," in 2016 eighth international conference on quality of multimedia experience (QoMEX). IEEE, 2016, pp. 1–6.

[56] Nitin Satpute, Rabia Naseem, Egidijus Pelanis, Juan Gómez-Luna, Faouzi Alaya Cheikh, Ole Jakob Elle, and Joaquín Olivares, "GPU acceleration of liver enhancement for tumor segmentation," Computer Methods and Programs in Biomedicine, vol. 184, pp. 105285, 2020.

[57] Erik Reinhard, Greg Ward, Sumanta Pattanaik, and Paul Debevec, High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics), Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.

[58] David Völgyes, Anne Catrine Trægde Martinsen, Arne Stray-Pedersen, Dag Waaler, and Marius Pedersen, "A weighted histogram-based tone mapping algorithm for ct images," Algorithms, vol. 11, no. 8, pp. 111, 2018.

[59] Michael A Campos and Alejandro A Diaz, "The role of computed tomography for the evaluation of lung disease in alpha-1 antitrypsin deficiency," Chest, vol. 153, no. 5, pp. 1240–1248, 2018.

[60] Ruud Janssen, Computational Image Quality, vol. PM101, SPIE PRESS, 2001.

[61] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He, "Aggregated residual transformations for deep neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1492–1500.

[62] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "CBAM: Convolutional block attention module," in Proceedings of the european conference on computer vision, 2018, pp. 3–19.

[63] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[64] Dongyoon Han, Jiwhan Kim, and Junmo Kim, "Deep pyramidal residual networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5927–5935.

[65] Sergey Zagoruyko and Nikos Komodakis, "Wide residual networks," arXiv preprint arXiv:1605.07146, 2016.

[66] François Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.

[67] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio, Deep learning, vol. 1, MIT press Cambridge, 2016.

[68] Simon Haykin, Neural networks, vol. 2, Prentice hall New York, 1994.

[69] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber, "Lstm: A search space odyssey," IEEE transactions on neural networks and learning systems, vol. 28, no. 10, pp. 2222–2232, 2017.

[70] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever, "An empirical exploration of recurrent network architectures," in International Conference on Machine Learning, 2015, pp. 2342–2350.

[71] Colah, "Understanding lstm networks," 2015, http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[72] Andrej Karpathy, Justin Johnson, and Li Fei-Fei, "Visualizing and understanding recurrent networks," arXiv preprint arXiv:1506.02078, 2015.

[73] Joseph Paul Cohen, Paul Morrison, and Lan Dao, "COVID-19 image data collection," arXiv 2003.11597, 2020.

[74] Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu, and Sen Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3484–3495, 2019.

[75] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," arXiv preprint arXiv:1912.02413, 2019.

[76] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Neural Networks, vol. 106, pp. 249–259, 2018.

[77] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al., "Pytorch: An imperative style, high-performance deep learning library," in Advances in Neural Information Processing Systems, 2019, pp. 8024–8035.

[78] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[79] Leslie N Smith, "Cyclical learning rates for training neural networks," in 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017, pp. 464–472.

AHMED MOHAMMED He is a post doctoral research fellow at NTNU. He received his Ph.D. degree in computer science from the Norwegian University of Science and Technology (NTNU), Norway, in 2020. He received Master's degree in Electronics and Information Engineering from Chonbuk National University, South Korea in 2014. His research interests are broadly in artificial intelligence, with emphasis on medical imaging and computer vision for data efficient and explainable anomaly detection and diagnosis.

CONGCONG WANG is a researcher at NTNU. She received her Ph.D. degree in computer science from the Norwegian University of Science and Technology (NTNU), Norway, in 2020. She received a Bachelor degree in electronic engineering from Shandong University in 2011 and a master degree in Erasmus Mundus master programme CIMET (Colour in Informatics and Media Technology) in 2014. Her research interests include image processing, medical image analysis, computer vision and applied deep learning with emphasis on video processing and 3D reconstruction and understanding for computer/machine vision.

RABIA NASEEM is a PhD candidate in the Department of Computer Science at Norwegian University of Science and Technology, Norway. She holds Masters degree in Software Engineering from University of Engineering and Technology, Taxila, Pakistan. She is working under EU project 'High Performance Soft Tissue Navigation' and her research is focused on medical image processing particularly medical image enhancement.

MENG ZHAO received the B.S. degree in Automation from Tianjin University, China, in 2010 and the M.S and Ph.D. degree in Control Science and Engineering from Tianjin University, China, in 2016. She is now a Lecturer in School of Computer Science and Engineering, Tianjin University of Technology, China. Recently, she has successfully finished a research stay as a postdoc in NTNU, Norway, supported by the European Research Consortium for Informatics and Mathematics (ERCIM) "Alain Bensoussan Fellowship Programme". Her research interest includes medical image processing, medical/biomedical engineering and machine learning/deep learning in medical informatics.

MOHIB ULLAH received a bachelor's degree in Electronic and Computer engineering from Politecnico Di Torino, Italy, in 2012, and a master's degree in Telecommunication Engineering from the University of Trento, Italy, in 2015. He received his Ph.D. degree in computer science from the Norwegian University of Science and Technology (NTNU), Norway, in 2019. Currently, he is working as a post-doctoral research fellow at NTNU and is involved in several industrial projects related to video surveillance. He is the reviewer of well-reputed conferences and journals (Elsevier Neurocomputing, Elsevier Neural Computing and Applications, Spring Multimedia Tools and Applications, IEEE Access, Journal of imaging, IEEE CVPRw, IEEE ICIP, IEEE AVSS, etc). He served as a chair of the technical program at the European workshop on visual information processing. He also served as a program committee member of the International Workshop on Computer Vision in Sports (CVsports). He teaches courses on machine learning and computer vision. His research interests include medical imaging, crowd analysis, object segmentation, behavior classification, and tracking. In these research areas, he published over 30 peer-reviewed journals, conferences, and workshop papers.

HAO WANG [Member, IEEE] is an Associate Professor in the Department of Computer Science in Norwegian University of Science and Technology, Norway. He has a Ph.D. degree (2006) and a B.Eng. degree (2000), both in computer science and engineering, from South China University of Technology, China. His research interests include big data analytics, industrial internet of things, high performance computing, and safety-critical systems. He served as a TPC co-chair for IEEE DataCom 2015, IEEE CIT 2017, ES 2017, and IEEE CPSCom 2020, and a senior TPC member for CIKM 2019. He is the Chair for Sub-TC on Healthcare of IEEE Industrial Electronics Society Technical Committee on Industrial Informatics.

MARIUS PEDERSEN is professor at the Norwegian University of Science and Technology in Gjøvik, Norway, and also the head of the Norwegian Colour and Visual Computing Laboratory. Pedersen has a BsC in computer engineering and a MsC in Media Technology from Gjøvik University College. His PhD in Color Imaging is from the University of Oslo in 2011. His research focuses on print and image quality.

FAOUZI ALAYA CHEIKH received his Ph.D. in Information Technology from Tampere Univ. of Technology, in Tampere, Finland in April 2004; where he worked as a researcher in the Signal Processing Algorithm Group since 1994. From 2006, he has been affiliated with the Department of Computer Science and Media Technology at Gjøvik University College in Norway, at the rank of Associate Professor. From January 2016 he is with the Norwegian University of Science and Technology (NTNU). He teaches courses on image and video processing and analysis and media security. His research interests include e-Learning, 3D imaging, image and video processing and analysis, video-based navigation, biometrics, pattern recognition, embedded systems and content-based image retrieval. In these areas, he has published over 100 peer-reviewed journal and conference papers, and supervised four post-doc researchers, five PhD and a number of MSc thesis projects. Dr. Alaya Cheikh is currently the co-supervisor of five PhD students. He has been involved in several European and national projects among them: ESPRIT, NOBLESS, COST 211Quat, HyPerCept, IQ-Med and H2020 ITN HiPerNav. He is on the editorial board of the IET Image Processing Journal and the editorial board of the Journal of Advanced Robotics & Automation and the technical committees of several international conferences. Dr. Alaya Cheikh is an expert reviewer to a number of scientific journals and conferences related to the field of his research. He is a senior member of IEEE, member of NOBIM and Forskerforbundet (The Norwegian Association of Researchers - NAR)