COVID-Net UV: An End-to-End Spatio-Temporal Deep Neural Network Architecture for Automated Diagnosis of COVID-19 Infection from Ultrasound Videos

Hilda Azimi — National Research Council Canada
Ashkan Ebadi — National Research Council Canada
Jessy Song — University of Waterloo
Pengcheng Xi — National Research Council Canada
Stéphane Tremblay — National Research Council Canada
Alexander Wong — University of Waterloo
Email: {hilda.azimi,ashkan.ebadi,pengcheng.xi,stephane.tremblay}@nrc-cnrc.gc.ca
{jessy.jia.song,alexander.wong}@uwaterloo.ca

Abstract

Besides vaccination, as an effective way to mitigate the further spread of COVID-19, fast and accurate screening of individuals to test for the disease is yet necessary to ensure public health safety. We propose COVID-Net UV, an end-to-end hybrid spatio-temporal deep neural network architecture, to detect COVID-19 infection from lung point-of-care ultrasound videos captured by convex transducers. The COVID-Net UV comprises a convolutional neural network that extracts spatial features and a recurrent neural network that learns temporal dependence. After careful hyperparameter tuning, the network achieves an average accuracy of 94.44% with no false-negatives for COVID-19 cases. The goal of COVID-Net UV is to assist front-line clinicians in the fight against COVID-19 as a decision support tool via accelerating the screening of lung point-of-care ultrasound videos and automatic detection of COVID-19 positive cases.

1 Introduction

The Coronavirus Disease 2019 (COVID-19) has resulted in a dramatic loss of life worldwide and posed an unprecedented public health challenge. There is no doubt that vaccination has been helping in mitigating the further spread of COVID-19. However, fast screening of individuals to test for the disease is still necessary to ensure public health safety [1]. Chest x-ray (CXR) and computed tomography (CT) are two modalities that are often used for screening patients suspicious of COVID-19 infection. Another imaging modality for diagnosing lung-related diseases is the lung point-of-care ultrasound (POCUS). This modality has been suggested as the most helpful in contexts/environments that are resource-limited, such as emergency settings or low-resource regions/countries [1]. Compared to CXR and CT, POCUS is much cheaper to acquire and has higher portability and accessibility, thus enhancing the ability for possible COVID-19 screening [2].

Deep learning (DL) networks have been applied to POCUS images for different tasks and analyses such as segmentation, disease classification, and detection [3]. However, the protocol for physicians to perform an ultrasound (US) examination requires them to capture and analyze the US video, often from various angles, views, and positions [1, 2]. This means that the sequences of US video frames from one position or view to another can provide physicians with more information to make an accurate diagnosis; and perhaps not all frames of US videos contain signs and symptoms of a suspected disease. Therefore, applying DL to frames of US videos only and without considering their temporal information is not the ideal solution to adopt POCUS data for screening and diagnostics purposes.

Motivated by this challenge and the promise of artificial intelligence (AI) tools to aid clinicians, we propose COVID-Net UV, an end-to-end spatio-temporal deep neural network architecture to detect COVID-19 positive cases from POCUS videos. Our contributions can be summarized as follows: 1) the COVID-Net UV is an effective tool for automatic detection of COVID-19 positive cases from POCUS videos without requiring the need for technician intervention and any further processing, 2) it also bridges the gap in the current diagnostic procedure of POCUS data by eliminating the need for the time-consuming and costly training of human experts, as interpreting US data requires domain knowledge [4]. We hope the COVID-Net UV helps front-line physicians and radiologists in screening patients, especially in resource-limited settings/environments/regions where current commonly used modalities, e.g., CXR and CT, are rare or unavailable.

2 Related Work

Several techniques have been proposed so far for classifying various features in POCUS images and videos related explicitly to COVID-19 disease. Some are DL models, trained and built mostly using frame-based data as input. Roy et al. applied a DL model derived from spatial transformer networks to predict the COVID-19 severity score associated with POCUS video frames. Their model provided localization of pathological artifacts in a weakly-supervised way and adopted a uninorms-based method for frame score aggregation at the video-level [5]. First, they classified every single frame of a POCUS image sequence into one of the four levels of disease severity. Next, they predicted a score for the entire frame sequence based on the same scoring scale by applying video-level grading. In [6], authors took one step beyond [5] and presented a technique for directly classifying POCUS videos based on a Two-Stream Inflated 3D ConvNet (I3D). They categorized the main imaging features seen in POCUS videos, such as A-lines, B-lines, consolidation, and pleural effusion, to unveil the degree to which the infection had affected the lungs. Both these works had different approaches to analyzing PUCUS videos in the presence of COVID-19 disease. However, what should be given special attention is that due to the rapid progression of COVID-19 into a very critical condition, quick diagnosis of the disease is very crucial. While reverse transcription polymerase chain reaction (RT-PCR) as the standard test to initially diagnose COVID-19 disease may take up to 24 hours and requires multiple tests for definitive results, diagnosis using POCUS videos can be much quicker. We aim to bridge the gap in the current diagnostic procedure of POCUS videos to initially detect and diagnose COVID-19 as fast as possible by learning spatio-temporal features, combining Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) architectures.

3 Data and Methods

To train and evaluate the COVID-Net UV, we used the COVIDx-US dataset v1.4. [7, 8] that contained 242 curated US videos, collected and integrated from nine different data sources. The videos comprised four different classes: COVID-19 infection, non-COVID-19 infection, other lung diseases/conditions, and normal control cases. The *other lung diseases/conditions* class included US videos of patients with various lung diseases and conditions, e.g., chronic obstructive pulmonary disease (COPD), pneumothorax, and hemothorax. We filtered out the other lung diseases/conditions class due to the heterogeneity of the cases. And, we only included lung US videos captured with a convex probe, as we observed that including data captured with the linear probe in training increased noise and influenced the performance of the network negatively.

We formulated the problem as a binary classification problem, i.e., the COVID-19 cases were labeled as positive and the normal and non-COVID-19 cases as negative. This resulted in 119 videos in total that were split into a training set with 76 videos (38 positives and 38 negatives), a validation set with 25 videos (12 positives and 13 negatives), and an unseen test set with 18 videos (10 positives and 8 negatives).

We employed a hybrid architecture that included convolutional and recurrent layers to process spatial and temporal aspects of lung ultrasound (LUS) videos, respectively. We adopted the InceptionV3 model [9], pre-trained on the ImageNet dataset [10], as the spatial feature extraction backbone and added two layers of Gated Recurrent Units (GRU) units (16 + 8) to capture temporal features. Since a video is an ordered sequence of frames, the frames can be extracted and stored in a 3D tensor. However, the number of frames may vary from video to video, making it impossible to stack them in

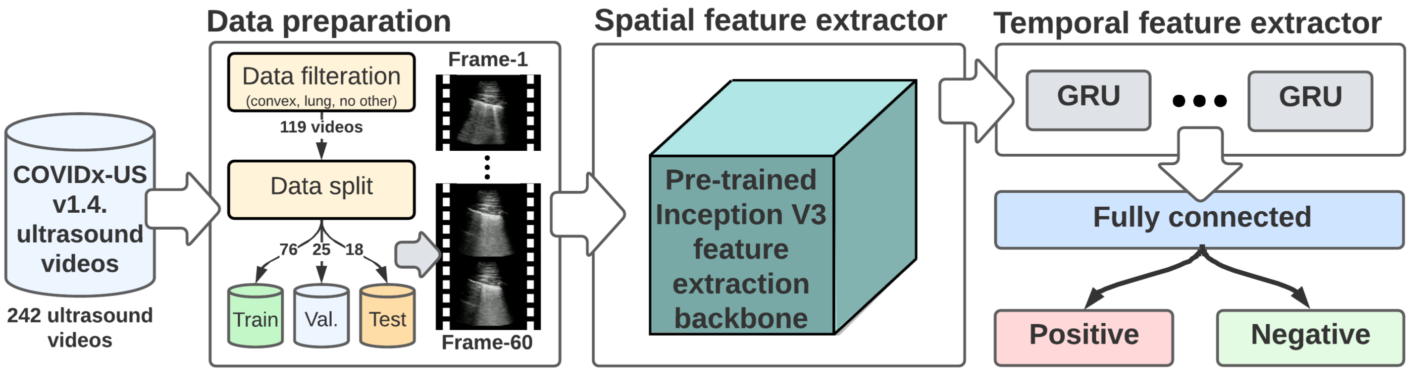

Fig. 1: COVID-Net UV: a CNN-RNN architecture to classify POCUS videos into two classes of positive, i.e., COVID-19 infection, and negative, i.e., non-COVID-19 infection.

batches. To overcome this problem, we first captured the frames of a video. Next, we extracted frames from the videos until a maximum frame count was reached. In this case, if the frame count was lower than the maximum frame count, the video was padded with zeros. We chose 60 as the maximum frame count, considering the characteristics of the videos in the dataset. To optimize the network and avoid over-fitting, we employed two callbacks in our training strategy: 1) *learning rate scheduler*, where we decayed the learning rate by a factor of 0.5 after three epochs with no performance improvement on the validation set, and 2) *early stopping*, where training was stopped after seven epochs with no performance improvement on the validation set. The initial learning rate and the maximum number of epochs were set at 0.001 and 30, respectively. The final network was trained for 18 epochs (following the *early stopping* strategy). Fig. 1 shows the high-level architecture of the COVID-Net UV as well as the conceptual flow of the analysis.

4 Results

The learning curves through the process of training and optimizing the network are illustrated in Fig. 2. The training of the network was stopped right before the loss on the validation set started increasing and avoided overfitting, as seen in Fig. 2-b. Following the *learning rate scheduler* strategy, during the process of training, the learning rate was decayed two times through epochs 15 and 18 (Fig. 2-c). We checked the performance of COVID-Net UV on the unseen test set. The network learned to classify COVID-19 positive and negative classes with an overall accuracy of 94.44%. The network achieved a sensitivity of 100% and 87.50% for positive and negative classes, respectively, meaning no false-negatives for the COVID-19 positive cases. The network reached a precision of 90.91% for the positive class and 100% for the negative class.

5 Discussion and Future Work

In this work, we proposed the COVID-Net UV, a hybrid end-to-end network architecture to classify lung POCUS videos for the diagnosis of COVID-19 disease. Our network comprises two modules: 1) a pre-trained InceptionV3 to extract spatial features from LUS video frames, and 2) RNN structure containing GRU units to learn the temporal dependence between the video frames. The network has a sensitivity of 100% in detecting COVID-19 positive cases. This suggests that the COVID-Net UV can be used a powerful AI-based decision support tool to assist clinicians. One may note that human experts have a sensitivity of 86.4% [11] in performing a similar task and similar models solely based on spatial architecture are reported to have the highest accuracy of 83.2% [12]. We used a public LUS dataset (i.e., the COVIDx-US) that includes data of various sources and quality. The network could be potentially improved and further validated by using larger datasets. The proposed methodology and pipeline could be used in future studies on larger datasets and the COVID-Net UV can be considered as a baseline for future works in examining more complex models on the larger POCUS video dataset. The network architecture can be tuned for new disease categories. Future research could evaluate the performance of the proposed architecture in other diseases. Another potential research direction could be adding explainability component to the pipeline.

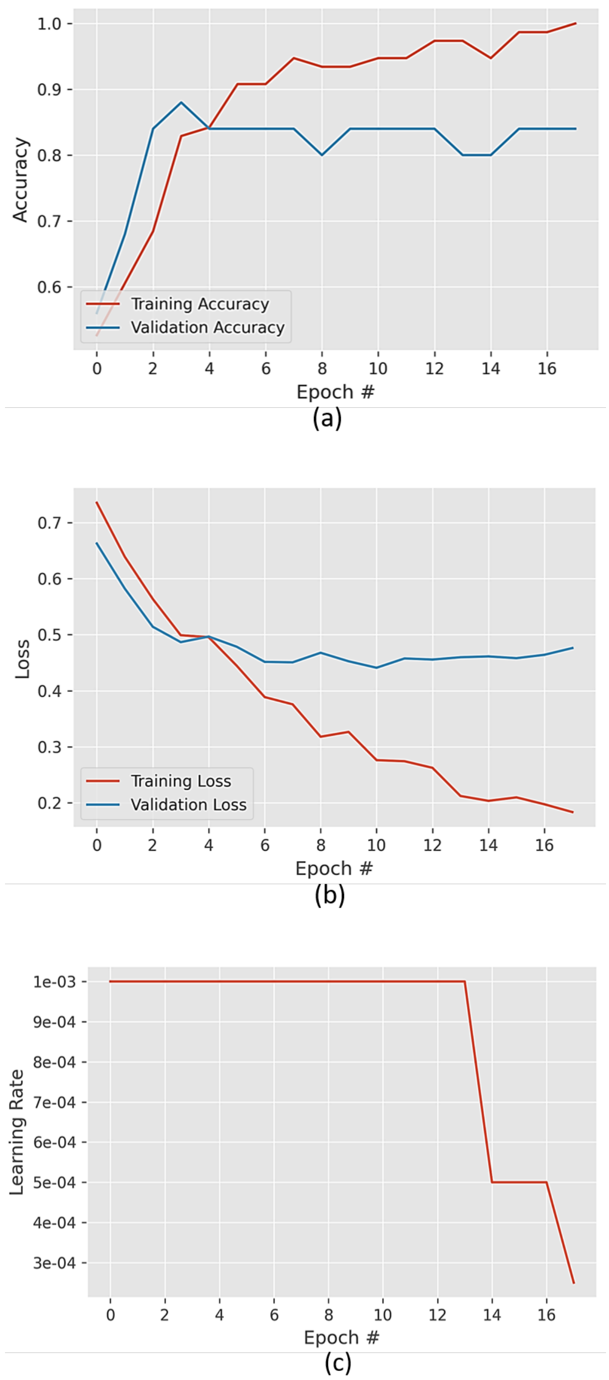

Fig. 2: Learning curves through the process of training and optimizing the network. (a) Accuracy, (b) Loss and (c) Learning rate

References

[1] A. MacLean, S. Abbasi, A. Ebadi, A. Zhao, M. Pavlova, H. Gunraj, P. Xi, S. Kohli, and A. Wong, “Covid-net us: A tailored, highly efficient, self-attention deep convolutional neural network design for detection of covid-19 patient cases from point-of-care ultrasound imaging,” *arXiv:2108.03131 [cs, eess]*, 8 2021, arXiv: 2108.03131. [Online]. Available: http://arxiv.org/abs/2108.03131

[2] Y. Amatya, J. Rupp, F. M. Russell, J. Saunders, B. Bales, and D. R. House, “Diagnostic use of lung ultrasound compared to chest radiograph for suspected pneumonia in a resource-limited setting,” *International Journal of Emergency Medicine*, vol. 11, no. 1, p. 8, 3 2018, pMID: 29527652 PMCID: PMC5845910.

[3] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, “Deep learning in medical ultrasound analysis: A review,” *Engineering*, vol. 5, no. 2, pp. 261–275, 4 2019.

[4] M.-Y. Ng, E. Y. P. Lee, J. Yang, F. Yang, X. Li, H. Wang, M. M.-s. Lui, C. S.-Y. Lo, B. Leung, P.-L. Khong, C. K.-M. Hui, K.-y. Yuen, and M. D. Kuo, “Imaging profile of the covid-19 infection: Radiologic findings and literature review,” *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, 2 2020, publisher: Radiological Society of North America.

[5] S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, I. Huijben, N. Chennakeshava, F. Mento, A. Sentelli, E. Peschiera, R. Trevisan, G. Maschietto, E. Torri, R. Inchingolo, A. Smargiassi, G. Soldati, P. Rota, A. Passerini, R. J. G. van Sloun, E. Ricci, and L. Demi, “Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2676–2687, 8 2020, event: IEEE Transactions on Medical Imaging.

[6] S. Erfanian Ebadi, D. Krishnaswamy, S. E. S. Bolouri, D. Zonoobi, R. Greiner, N. Meuser-Herr, J. L. Jaremko, J. Kapur, M. Noga, and K. Punithakumar, “Automated detection of pneumonia in lung ultrasound using deep video classification for covid-19,” *Informatics in Medicine Unlocked*, vol. 25, p. 100687, 2021, pMID: 34368420 PMCID: PMC8332742.

[7] A. Ebadi, P. Xi, A. MacLean, S. Tremblay, S. Kohli, and A. Wong, “Covidx-us – an open-access benchmark dataset of ultrasound imaging data for ai-driven covid-19 analytics,” *arXiv:2103.10003 [cs, eess]*, 4 2021, arXiv: 2103.10003. [Online]. Available: http://arxiv.org/abs/2103.10003

[8] A. Ebadi, P. Xi, A. MacLean, A. Florea, S. Tremblay, S. Kohli, and A. Wong, “Covidx-us: An open-access benchmark dataset of ultrasound imaging data for ai-driven covid-19 analytics,” *Frontiers in bioscience (Landmark edition)*, vol. 27, no. 7, 2022.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” 12 2015, arXiv:1512.00567 [cs]. [Online]. Available: http://arxiv.org/abs/1512.00567

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database.” 2009 IEEE Conference on Computer Vision and Pattern Recognition, 6 2009, pp. 248–255, iSSN: 1063-6919.

[11] N. Islam, S. Ebrahimzadeh, J.-P. Salameh, S. Kazi, N. Fabiano, L. Treanor, M. Absi, Z. Hallgrimson, M. M. Leeflang, L. Hooft, v. d. C. B. Pol, R. Prager, S. S. Hare, C. Dennie, R. Spijker, J. J. Deeks, J. Dinnes, K. Jenniskens, D. A. Korevaar, J. F. Cohen, d. A. V. Bruel, Y. Takwoingi, v. d. J. Wijgert, J. A. Damen, J. Wang, M. D. McInnes, and C. C.-. D. T. A. Group, “Thoracic imaging tests for the diagnosis of covid-19,” *Cochrane Database of Systematic Reviews*, vol. 2021, no. 3, 2021, [Online; accessed 2022-03-04]. [Online]. Available: https://www.readcube.com/articles/10.1002%2F14651858.cd013639.pub4

[12] N. Awasthi, A. Dayal, L. R. Cenkeramaddi, and P. K. Yalavarthy, “Mini-covidnet: Efficient lightweight deep neural network for ultrasound based point-of-care detection of covid-19,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 6, pp. 2023–2037, 6 2021, event: IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control.