POINT OF CARE IMAGE ANALYSIS FOR COVID-19

Daniel Yaron[1*]   *Daphna Keidar*[2*]   *Elisha Goldstein*[1*]   *Yair Shachar*[3]   *Ayelet Blass*   *Oz Frank*[1]
Nir Schipper[4]   *Nogah Shabshin*[5]   *Ahuva Grubstein*[7,8]   *Dror Suhami*[7,8]   *Naama R. Bogot*[6]
Chedva S. Weiss[6]   *Eyal Sela*[9]   *Amiel A. Dror*[9]   *Mordehay Vaturi*[7,8]   *Federico Mento*[10]
Elena Torri[11]   *Riccardo Inchingolo*[12]   *Andrea Smargiassi*[12]   *Gino Soldati*[13]   *Tiziano Perrone*[14]
Libertario Demi[10]   *Meirav Galun*[1]   *Shai Bagon*[1]   *Yishai M. Elyada*[15]   *Yonina C. Eldar*[1]

[1]Weizmann Institute of Science, Israel   [2]ETH Zürich, Switzerland   [3]Eyeway Vision Ltd. Israel
[4]The Hebrew University of Jerusalem, Israel   [5]HaEmek Medical Center, Israel   [6]Shaare Zedek Medical Cener, Israel
[7]Sackler school of medicine, Tel Aviv University, Israel   [8]Rabin Medical Center, Israel
[9]Galilee Medical Center, Azrieli Faculty of Medicine, Bar-Ilan University, Israel   [10]University of Trento, Italy   [11]Bresciamed, Italy
[12]Fondazione Policlinico Universitario A. Gemelli IRCCS, Italy   [13]Valle del Serchio General Hospital, Italy
[14]Fondazione IRCCS Policlinico San Matteo di Pavia, Italy   [15]Mobileye Vision Technologies Ltd., Israel
*These authors contributed equally to this work.

ABSTRACT

Early detection of COVID-19 is key in containing the pandemic. Disease detection and evaluation based on imaging is fast and cheap and therefore plays an important role in COVID-19 handling. COVID-19 is easier to detect in chest CT, however, it is expensive, non-portable, and difficult to disinfect, making it unfit as a point-of-care (POC) modality. On the other hand, chest X-ray (CXR) and lung ultrasound (LUS) are widely used, yet, COVID-19 findings in these modalities are not always very clear. Here we train deep neural networks to significantly enhance the capability to detect, grade and monitor COVID-19 patients using CXRs and LUS. Collaborating with several hospitals in Israel we collect a large dataset of CXRs and use this dataset to train a neural network obtaining above 90% detection rate for COVID-19. In addition, in collaboration with ULTRa (Ultrasound Laboratory Trento, Italy) and hospitals in Italy we obtained POC ultrasound data with annotations of the severity of disease and trained a deep network for automatic severity grading.

1. INTRODUCTION

Coronavirus Disease 2019 (COVID-19) was declared a global pandemic [1], and has had severe economic, social and healthcare consequences. In order to contain the disease, an immediate concern is to rapidly identify and isolate SARS-CoV-2 carriers. This requires means for mass testing of the general population, with low cost, high sensitivity and fast processing times. The prevalent test today is Reverse Transcription Polymerase Chain Reaction (RT-PCR) [2, 3], which suffers from a number of problems: Testing reagents and kits are expensive and suitable for single-use only, processing the samples requires dedicated personnel and equipment, and it can take hours or days to obtain results. Most significantly, the test has a limited sensitivity rate of as low as 71% [4, 5]. Due to these shortcomings, finding alternative testing and identification methods is crucial. A strong candidate is diagnosis of

patients based on medical imaging of the chest, since COVID-19 presents primarily in the lower respiratory tract. Medical imaging, specifically computerized tomography (CT) scans, chest X-ray (CXR), and lung ultrasound (LUS), can provide an alternative approach, affording advantages that can readily complement the testing capabilities of RT-PCR. In the case of COVID-19, disease characteristics such as consolidations and ground-glass opacities can be identified in images of the lung [6, 7] which raises the possibility of using chest and lung imaging for detection and severity grading of COVID-19 patients.

Physiological pulmonary properties of COVID-19 on CT scans are typically detectable and clear to see [6]. However, the availability of CT equipment is limited both by its price and operational requirements such as rooms and staff. Moreover, there is a need to decontaminate the machine between suspected COVID-19 patients, a lengthy process that results in a very slow rate of scanning. In contrast, with portable X-ray and ultrasound machines, imaging can be done rapidly and without needing to bring patients into radiography rooms. These machines are also less costly, use less radiation, and can be readily distributed and deployed to point-of-care (POC) locations outside hospitals and primary care centers. The drawback of these modalities is that their analysis requires qualified personnel and the unique characteristics of the prognostic properties of the images can make them much harder to analyze than CT scans.

In this paper we consider a combination of signal processing and deep learning tools to develop deep network architectures that can lead to high detection rates of COVID-19 and to severity grading of disease from POC imaging using X-ray and ultrasound. Early approaches to develop detection methods based on X-ray used data from publicly available image sources relied on limited data containing compressed images with lack of detail, and coming from many different makes and models of x-ray machines [8, 9, 10]. This illustrates one of the main challenges in this field, namely the collection of large amounts of COVID-19 positive and negative images of full resolution and similar sources. Notably, one recent effort has shown more reliable results based on a larger dataset

that was more uniformly sourced and comes closer to the goal of developing tools that can be used in clinical settings [11]. They achieved a sensitivity of 88% with a specificity of 79%.

Here, we collected a large dataset of images from portable X-rays and used them to train a network that can detect COVID-19 with high reliability. Our algorithm adds external information in the form of lung segmentation based on a deep learning model, which together with several other pre-processing methods, boosts performance to over 90% detection rate. We further develop a tool for retrieving CXR images that are similar to a given query image based on that image's embedding in a low-dimensional space generated by the model.

For LUS, our goal is to develop a network with high reliability to grade disease severity. To this end, we rely on ICLUS dataset presented in Roy *et al.* [12][1]. There, the authors propose a sophisticated neural network to automatically predict the severity grade from annotated LUS frames, which results in an F1 score (the harmonic mean of precision and recall) of 65.1%. Here we enhance their method by developing a signal processing approach to "rectify" images taken by convex probes, and then similar to the X-ray network, we input both the original and rectified images, creating a channel of "side information". This results in an F1 score of 68.7%.

In the following sections we detail the approach taken in developing the networks for both the X-ray and LUS data. In both cases signal processing tools are used to pre-process the data and to form an additional input channel that boosts the network results.

2. COVID-19 DETECTION BASED ON X-RAY

Deep learning approaches have shown impressive abilities in image related tasks, including in many radiological contexts [13] [14]. However, despite their potential in assisting COVID-19 management efforts, these methods require large amounts of training data. In order to address this challenge, a large dataset of images from portable X-rays was sourced and used to train a network that can detect COVID-19 in the images with high reliability, and to develop a tool for retrieving CXR images that are similar to each other. We rely on a combination of deep learning tools including standard pre-processing and augmentations of the images and additional information in the form of an extra lung segmentation channel.

We collected CXR images from 1384 patients, 360 with a positive COVID-19 diagnosis and 1024 negative, totaling 2426 CXRs. All COVID-19 negative images come from before the start of the pandemic. Patients' COVID-19 positive labels were determined according to positive RT-PCR testing. The COVID-19 positive images include all CXRs performed with portable X-ray machines on patients admitted to

---
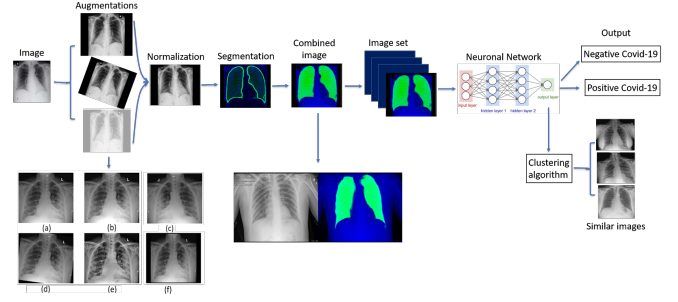
Fig. 1. Full pipeline workflow overview for COVID-19 classification model of X-ray images.

four hospitals in Israel. For the non COVID-19 images we obtained CXRs taken by the same X-ray machines prior to December 2019. These are patients without COVID-19, typically with another respiratory disease. The test set was taken from the full CXR dataset and contains 350 CXR (15%) of which 179 (51%) are positive for COVID-19 and 171 (49%) are negative. Many patients have multiple images. To deal with this, each patient's images were used either for the test set or for the train set, but never for both. This is done to prevent the model from identifying patient-specific image features (e.g., medical implants) and associating them with the label. In the analysis, 4% (101/2426) of the images were excluded due to lateral positioning, or due to rectangular artifacts in the image, of these 98 were COVID-19 positive.

2.1. Image Processing and Network architecture

The model pipeline (Fig. 1), begins with a series of preprocessing steps, including augmentation, normalization, and segmentation of the images. Augmentations are transformations that change features such as image orientation and brightness. These properties are irrelevant for correct classification, but may vary during image acquisition, and can affect the training performance of the network because of its rigid registration with respect to orientation and pixel values. Importantly, augmentations should correspond to normal variation in CXR acquisition; to ensure this we consulted with radiologists when defining the augmentation parameters. Not all augmentations are applied each time, but rather each augmentation has a certain probability of being applied, represented by p: brighten (p=0.4); gamma contrast (p=0.3); CLAHE (p=0.4); rotate $d \in [-7°, 7°]$ (p=0.4); shear $d \in [-7°, 7°]$ (p=0.4); scale up to 0.2 on each axis (p=0.4); flip from left to right (p=0.5); either sharpen or apply Gaussian blur; horizontal flip (p =0.5).

The normalization process consisted of cropping black edges, standardizing the brightness and scaling the size of each image to 1024X1024 pixels using bilinear interpolation.

To enhance performance we created an additional image channel using lung segmentation via a U-net [15] pre-trained on a different dataset. This network produces a pixel-mask

Model	Accuracy	Sensitivity	Specifity
	% (/350)	% (/179)	% (/171)
ResNet34	89.4 (313)	87.1 (156)	91.8 (157)
ResNet50	**89.7 (314)**	**87.1 (156)**	**92.4 (158)**
ResNet50 - No preprocessing	85.1 (298)	82.1 (147)	88.3 (151)
ResNet152	86.0 (304)	83.2 (149)	90.6 (155)
Chexpert	84.0 (294)	86.5 (155)	81.2 (139)
VGG16	87.7 (306)	87.1 (156)	88.3 (151)
Ensemble	**90.5 (317)**	**91.1 (164)**	**90.0 (154)**

Table 1. **Performance of different models:** Comparison of accuracy, sensitivity and specificity of various deep networks trained and tested on the same test set.

of the CXR indicating the probability that each pixel belongs in the lungs, allowing the network to access this information while training. The final input images to the network contain 3 channels: the original CXR, the segmentation map, and one filled with zeroes. This is done to accommodate the pretrained models we used that use 3-channel RGB images.

We compared five networks: ResNet34, ResNet50, ResNet152 [16], VGG16 [17] and Chexpert [14], all trained using transfer learning. We additionally classify the images with an ensemble model by averaging over the results of the first four networks, with exclusion of Chexpert due to low results. Training was performed with the Adam optimizer with an initial learning rate of 1e-6 which was exponentially decreased as epochs progressed. We used cross-entropy as a loss function with an L_2 regulariser with regularisation coefficient 0.01. The best test scores were achieved after 32 epochs.

In addition to classification, we propose a method for retrieving a number of CXR images that are the most similar to a given image. The activation of layers of the neural network serve as embeddings of the images into a vector space, and should capture information about clinical indications observed in the images. We use the embeddings obtained from the networks last layer to search for similarity between the resulting vectors, and retrieve the nearest neighbors of each image.

2.2. Evaluation and Results

We trained five deep networks whose accuracy, sensitivity (detection rate) and specificity rates can be seen in Table 1. We selected ResNet50 and the ensemble model for the rest of the analysis, as they achieved the best performance in our task. The ensemble model achieved accuracy 90.6%, (95% CI: 84%, 92.9%) sensitivity 91.1% (95% CI: 83%, 92.4%) and specificity of 90% (95% CI: 80.7%, 95.9%) on the test images.

The area under the curve (AUC) of the ROC curve is 0.96 (95% CI: 0.92, 0.97). The ROC curve is provided in Fig. 2a, and demonstrates that for a broad range of thresholds, both high true positive rate (TPR) and low false positive rate (FPR)


(a)   (b)
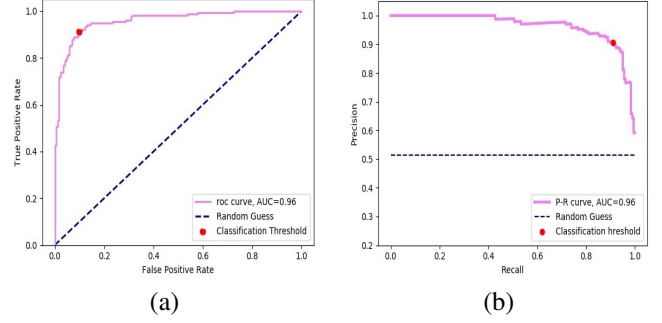
Fig. 2. **Model performance:** (a) Receiver Operating Characteristic (ROC) curve. The area under the ROC curve is 0.96 and the classification threshold we used is marked, where true positive rate (TPR) is 91.1% and false positive rate (FPR) is 10%; (b) Precision-Recall curve. The area under the curve is 0.96 and the classification threshold we used is marked where the model achieved precision of 90% and recall of 91%.
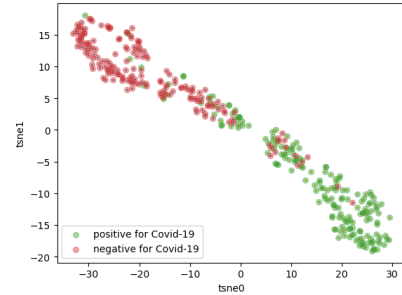
Fig. 3. **t-distributed Stochastic Neighbor Embedding (t-SNE):** The arrangement of the points, as a result of transforming the features of each image, shows two main distinct clusters, indicating strong association of the features with the ground truth.

are achievable. In Fig. 2b we present the precision-recall (P-R) curve, which shows the trade-off between precision and recall (sensitivity) as the value of the threshold is varied. This P-R curve shows a broad range of thresholds for which both high precision and high recall are attainable. The AUC of the P-R is 0.96 (95% CI: 0.93, 0.98). These ROC and P-R curves attest to the stability of the model across different thresholds. We additionally train ResNet50 on the dataset with and without all the preprocessing stages. As seen in Table 1, preprocessing incurs an improvement of 4% in accuracy and 5% in sensitivity.

We additionally visualize the distinction made by the model using t-distributed Stochastic Neighbor Embedding (t-SNE) [18] which uses a nonlinear method to reduce high dimensional feature vectors into two dimensions, as seen in Fig. 3. This makes it possible to visualize the data points and reveal similarities and dissimilarities between them. We used one of the last layers of the networks, which essentially provides an embedding of the images into a vector space. These vector embeddings of the images are given as input to the t-
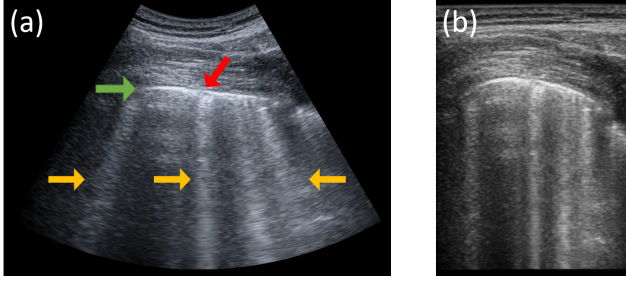
Fig. 4. **Lung Ultrasound:** (a) A LUS frame captured using a convex probe. Pleural line is indicated by green, and the red arrow points at consolidations. Sonographic artifacts appearing as bright vertical lines ("B-lines") indicated by the yellow arrows. The B-lines are *not* axis aligned and appear to emit for a single focal point. (b) The same frame rectified using its polar coordinates. The B-lines now appear axis-aligned, while the pleural line is distorted.

SNE which transformed each vector to a data point on the 2 dimensional space. Each data point was then colored according to its ground truth. We see that the arrangement of dots, representing features of the images colored by their true labels, depicts two distinct clusters, revealing a similarity between most of the images with the same label.

Finally, we applied K-Nearest Neighbors (KNN) on the image embeddings in order to retrieve images similar to each other. For each image we retrieve 4 images with the closest image embeddings; averaging over these images' predictions achieves 87% accuracy and 83.2% sensitivity, meaning that the nearest images typically have the same labels.

3. SEVERITY GRADING USING ULTRASOUND

It has been recently shown that Lung Ultrasound (LUS) can be used to assess the severity of COVID-19 patients [19, 20, 21]. Soldati *et al.* [19] devised a 4-scale grading system for POC LUS scans. The grades are based on observing both anatomical features, (e.g., integrity of the pleural line and presence of consolidations) as well as sonographic artifacts (aka "A-lines" and "B-lines"). Fig. 4a shows a typical LUS frame with prominent anatomical and sonographic artifacts annotated. These features can be observed regardless of the probe used (linear or convex). However, the *orientation* of the sonographic artifacts changes between probe type: While they appear "axis-aligned" when using a linear probe, when using a convex probe, the artifacts appear "tilted" as if emitted from the focus point of the probe. This difference has little effect on a human observer, but can be confusing for an automatic grading system.

Roy *et al.* [12] proposed a sophisticated neural network to automatically predict the proposed 4-scale clinical score from LUS frames, for frames captured by both linear and convex probes. They treat all frames in the same manner and rely on a Spatial Transformer Network [22] to focus on relevant parts

| Model | Settings 1 | Settings 2 | | | |
| | | Drop Transision Frames (K) | | | |
	All Frames	K=1	K=3	K=5	K=7
Resnet-18 [12]	62.2	63.9	65.5	66.9	67.8
Full [12]	65.1	66.7	68.3	69.5	70.3
Ours	**68.7**	**70.0**	**72.1**	**73.9**	**75.3**

Table 2. **Grading severity of LUS frames:** Comparing F1 scores (%) to [12] using the same dataset and evaluation settings (for definition see [12]). Our Resnet-18 based ensemble surpasses even the complicated STN-based model of [12].

of the frame and overcome the difference between "convex" and "linear" frames. They collected a dataset of 58,924 LUS frames of which 78% were captured using a convex probe, and used this data to train and evaluate their network.

In this work we took a more explicit approach to deal with the discrepancy between frames captured with linear and convex probes and suggest a simpler network with improved performance. We noted that frames captured with convex probes are formed to depict the anatomy as accurate as possible. However, forming the frames this way results in tilting the sonographic artifacts making them appear diagonal rather than axis-aligned. Therefore, one can rectify the frame using its underlying polar coordinates, making the sonographic artifacts appear axis-aligned, as in frames captured using a linear probe. This process is exemplified in Fig 4b. Once we make the artifacts axis aligned it is easier for a standard convolutional architecture to handle this type of information.

We then trained an *ensemble* of two Resnet-18 [16] networks: One receives the original frames as input and the second processes the rectified frames (for frames captured with a linear probe we treated the rectification as an identity). We used the same dataset and train/test split as in [12].

Table 2 compares F1 scores of our proposed system to that of [12]. The results shows that explicitly treating the problematic alignment of the sonographic artifacts allows us to achieve better F1 scores with a *simpler* architecture compared to [12]. This demonstrates the potential of LUS as a method for grading the severity of patients.

4. CONCLUSION

In this paper we showed how a combination of signal processing and machine learning tools can enhance imaging in COVID-19 patients. We first demonstrated how proper preprocessing and lung segmentation combined with transfer learning can lead to a simple deep network for COVID-19 detection based on portable X-ray images, reaching a detection rate of over 90%. We then showed how proper rectification of ultrasound images can result in an efficient deep network for grading severity from LUS data. We believe this work can pave the way to a wider use of available POC modalities in treating COVID-19 patients.

5. REFERENCES

[1] World Health Organization, "Who director-general's opening remarks at the media briefing on COVID-19 - 11 march 2020," 2020.

[2] C.B.F. Vogels, A.F. Brito, A.L. Wyllie, and others., "Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets," *Nature Microbiology*, vol. 5, no. 10, pp. 1299–1305, Oct 2020.

[3] Buddhisha Udugama, Pranav Kadhiresan, Hannah N. Kozlowski, Ayden Malekjahani, et al., "Diagnosing COVID-19: The disease and tools for detection," *ACS Nano*, vol. 14, no. 4, pp. 3822–3835, Apr 2020.

[4] L Luo, L Dan, L Xin-long, W Xian-bo, J Qin-long, Z Jia-zhen, et al., "Modes of contact and risk of transmission in COVID-19 among close contacts," *medRxiv*, 2020.

[5] I. Arevalo-Rodriguez, D. Buitrago-Garcia, D. Simancas-Racines, et al., "False-negative results of initial rt-PCR assays for COVID-19: A systematic review," *medRxiv*, 2020.

[6] Weifang Kong and Prachi P. Agarwal, "Chest imaging appearance of COVID-19 infection," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, pp. e200028, 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*, vol. 296, no. 2, pp. E115–E117, 2020, PMID: 32073353.

[8] Faisal M shah, Sajib K S Joy, Farzad Ahmed, Mayeesha Humaira, Amit S Ami, Shimul Paul, and Abidur R K Jim, "A comprehensive survey of COVID-19 detection using medical images," *engrXiv*, Jul 2020.

[9] Linda Wang and Alexander Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-Ray images," *arXiv*, 2020.

[10] Mohammad Mudasir Bhat Asif Iqbal Khan, Junaid Latief Shah, "Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, vol. 196, pp. 105–581, 2020.

[11] Ran Zhang, Xin Tie, Zhihua Qi, et al., "Diagnosis of COVID-19 pneumonia using chest radiography: Value of artificial intelligence," *Radiology*, vol. 0, no. 0, pp. 202944, 0, PMID: 32969761.

[12] 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.

[13] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, 2020.

[14] Irvin J., Rajpurkar P., and Ko M. et al., "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," *arXiv*, 2019.

[15] 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.

[16] 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.

[17] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv*, 2015.

[18] Laurens van der Maaten and Geoffrey Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.

[19] Gino Soldati, Andrea Smargiassi, Riccardo Inchingolo, Danilo Buonsenso, Tiziano Perrone, Domenica Federica Briganti, Stefano Perlini, Elena Torri, Alberto Mariani, Elisa Eleonora Mossolani, et al., "Proposal for international standardization of the use of lung ultrasound for patients with COVID-19: A simple, quantitative, reproducible method," *Journal of Ultrasound in Medicine*, 2020.

[20] Yael Lichter, Yan Topilsky, Philippe Taieb, Ariel Banai, Aviram Hochstadt, Ilan Merdler, Amir Gal Oz, Jacob Vine, Or Goren, Barak Cohen, et al., "Lung ultrasound predicts clinical course and outcomes in COVID-19 patients," *Intensive care medicine*, pp. 1–11, 2020.

[21] Andrea Smargiassi, Gino Soldati, Elena Torri, Federico Mento, Domenico Milardi, Paola Del Giacomo, Giuseppe De Matteis, Maria Livia Burzo, Anna Rita Larici, Maurizio Pompili, et al., "Lung ultrasound for COVID-19 patchy pneumonia: extended or limited evaluations?," *Journal of Ultrasound in Medicine*, 2020.

[22] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al., "Spatial transformer networks," in *Advances in neural information processing systems*, 2015, pp. 2017–2025.