Deep convolutional neural networks for predicting the presence and severity of COVID-19 from chest X-rays

Reece Huff

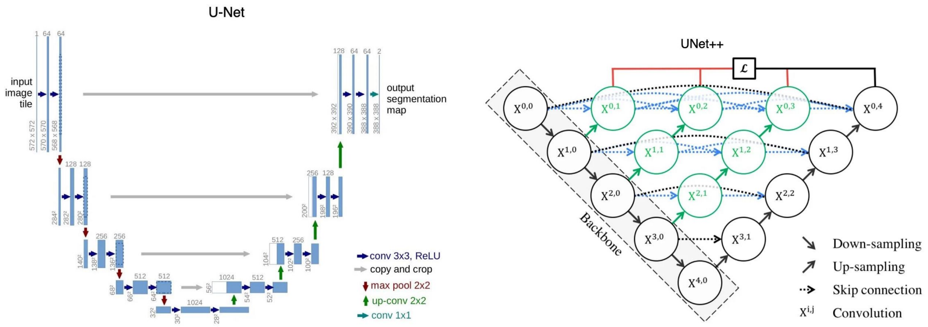
University of California, Berkeley

Abstract. COVID-19 has had heavy economic and medical burdens on society over the past two years, making its detection of utmost importance. Chest x-ray images offer an alternative for COVID-19 detection to the current gold-standard, PCR-testing. Segmentation-based classification and direct classification algorithms were compared as machine-learning approaches for diagnosis of COVID-19 with chest x-rays. Our custom direct classification artificial neural network, COVID-Net, outperformed both segmentation-based approaches, UNet and UNet++, by more than 20%, with a final test accuracy of 95.20%. While UNet and UNet++ have lower classification performance, they have the added benefit of measuring COVID-19 localization and severity as a percentage of of the area of infected lung tissue of total lung tissue, and provide information beyond what a PCR test could provide. Our findings suggest that machine-learning algorithms can enable powerful detection of COVID-19 through chest X-ray images. The best approach depends on the desired application; if diagnostic accuracy is the priority, COVID-Net should be utilized, whereas segmentation-based algorithms should be used if COVID severity and location is desired.

1 Introduction

The COVID-19 pandemic has had wide-reaching effects over the past couple years, making the development of accurate testing methods necessary. While PCR tests remain the gold-standard in testing, they are not able to assess severity or location of the infection. To address this, researchers have developed extensive datasets of chest x-ray images along with ground truth lung segmentation and COVID infection segmentation data in COVID-19 patients, patients with non-COVID infections, and patients with no infection [1][2]. Combining lung segmentation and COVID-19 infection segmentation allows for both COVID-19 detection, as well as location identification and severity determination.

The U-Net architecture [3], which won the ISBI challenge for cell tracking in 2015, is effective for image segmentation tasks. Image segmentation is particularly useful in biomedical imaging, where often information about the classification of each pixel is desired, rather than overall classification of an image. The structure (Fig. 1a) consists of symmetric contraction and expansion operations. The contractions use convolution layers with ReLU activation functions



(a) UNet [3] (b) UNet++ [4]

Fig. 1: UNet and UNet++ architectures.

and max pooling. As the image width and height are reduced, the number of feature channels are increased. Then, during the expansion, up-sampling is used to restore the image dimensions. At each up-sampling step, the data is concatenated with the corresponding image from the contraction path. The image from the contraction path provides high-resolution data, while the image from the up-sampling path provides context. This is an improvement on the previously used image segmentation neural net structures, which used small local patches, and therefore had limited context outside of a small area.

UNet++ further improved the generalizability of UNet by nesting a series of traditional UNets within itself [4]. As a result, earlier feature maps are concatenated onto later iterations of UNet, essentially reducing the semantic gap between the encoder and decoder phases of the network. By reducing these semantic gaps, UNet++ has outperformed on images of cells and livers by an average of ∼3%.

While image segmentation networks such as UNet and UNet++ have enabled both detection and location identification of COVID-19, they may not provide the highest detection accuracy compared to standard classification networks. Therefore, **our goal** was to compare the accuracy and efficiency of a segmentation-based COVID diagnostic (UNet and UNet++) with a direct classification diagnostic. We constructed, tuned, and validated our own architecture, COVID-Net (Fig. 2), for direct diagnosis of COVID, which uses a cascade of both convolutional neural networks and fully-connected neural networks to improve robustness. The **key outcome** of this study was a comprehensive assessment of segmentation-based classification and direct classification algorithms for COVID-19 diagnosis from chest x-ray images, with careful consideration of each method's strengths and weaknesses.

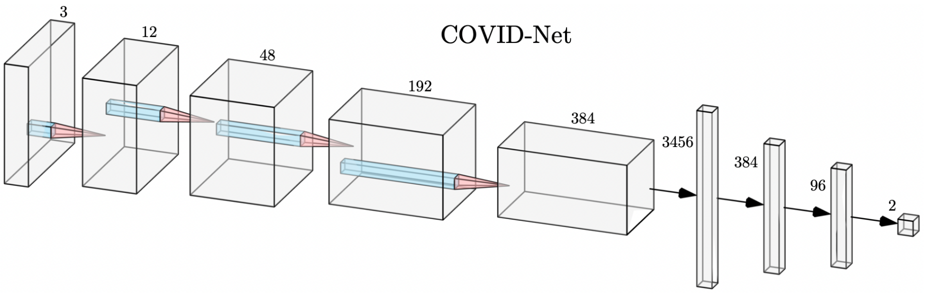


Fig. 2: COVID-Net classifier architecture

2 Methods

2.1 Dataset

The dataset used is publicly available through Kaggle. The dataset creation is detailed in [1][2]. Briefly, a human-machine collaborative method was used to create the ground truth segmentation masks, in which medical doctors created, evaluated, and modified segmentation masks generated by a U-Net segmentation neural network. The model was iteratively trained with the added input from doctors. All ground truth masks in the final dataset were verified by humans.

In summary, the dataset consisted of 5,836 chest x-rays in total, of which, 2,916, 1,460, and 1,460 chest x-rays were from patients with COVID-19, non-COVID infections (e.g., viral or bacterial pneumonia), and no infection (i.e., healthy). Of the 5,836 scans, approximately 60% was set aside for the training set, and the remaining 40% was equally divided into the validation and test set. While the images were randomly split into training, validation, and test sets, we made sure that each set had approximately the same proportion of COVID-19, non-COVID infections, and healthy x-rays.

2.2 UNet & UNet++

Custom implementations of UNet and UNet++ were employed in PyTorch and inspired by Zhou et al [4]. Modifications were made to the dataset creation and data loader to match our dataset. No additional preprocessing was applied to the data. The model hyperparameters were not tuned due to computational expense and time constraints.

The probabilistic image segmentation output was used to first determine if COVID-19 was present (positive if at least one pixel was identified as a COVID-19 infection area with greater than 50% probability). If it was, lung area data was used in combination with COVID-19 infection area to determine the severity of infection.

The UNet and UNet++ models for lung segmentation and infection segmentation were both trained over 100 epochs, with the final saved model being the one with the highest validation accuracy throughout the training. The infection segmentation models were trained using the Adam optimizer. The loss function used was a combination of cross-entropy loss and dice loss.

2.3 COVID-Net

Inspired by AlexNet, COVID-Net used a cascade of convolutional neural network layers and fully-connected neural network layers to perform COVID-19 detection and diagnosis (Fig. 2). Specifically, the input x-ray was first passed through a CNN block that contains a 2D convolution (i.e., cross-correlation function), batch normalization, max-pooling, and a rectified linear unit (ReLU) four times. Each pass through the CNN block increased the number of output channels, resulting 12, 48, 192, and 384 feature maps after each block, respectively. After passing through the four-layered CNN, the final feature maps were flattened and fed into a three-layer fully-connected neural network. Dropout was incorporated in between layers of the MLP as a form of regularization, thus enhancing robustness and generalizability. The final outputs provided a probability of the input chest x-ray containing COVID. Probabilities greater than 50% were considered COVID positive. Additionally, the training set was synthetically augmented with a series of randomized rotations, flips, and crops to improve accuracy. COVID-19 was trained for 50 epochs using the Adam optimizer and the cross-entropy loss function, ultimately reaching approximately 94% validation accuracy.

3 Results

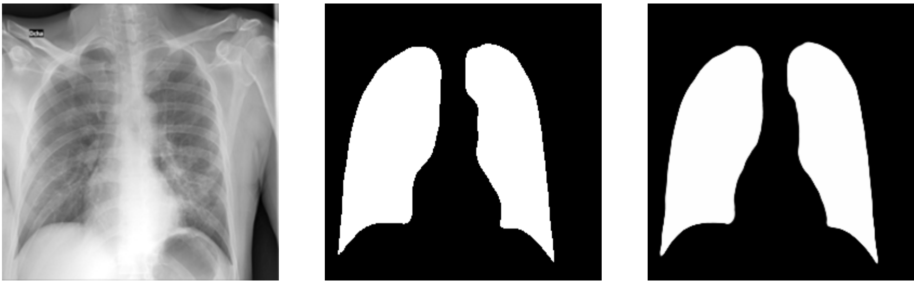
3.1 UNet vs. UNet++ Segmentation

Learning algorithm	Final test IoU	
	lung segmentation	infection segmentation
UNet	93.08%	70.80%
UNet++	94.00%	64.75%

Table 1: UNet, UNet++, and COVID-Net IoU on the test set.

After 100 epochs of training, both UNet and UNet++ achieve high performance. Specifically, they achieve an IoU (the area of the intersection divided by the area of the union) above 90% (Table 1) on lung segmentation. However, their performance decreases for infection segmentation, perhaps due to challenging delination of the infected region of the lung (Table 1).

An example of the performance of UNet++ on both the lung segmentation and infection segmentation is in Fig. 3 and Fig. 4, respectively. The predicted lung segmentation (Fig. 3c) is nearly identical to the target mask (Fig. 3b) given



(a) Chest x-ray. (b) Target mask. (c) UNet++ prediction.

Fig. 3: Representative performance of UNet++ on **lung segmentation** of chest x-ray.

(a) Chest x-ray. (b) Target mask. (c) UNet++ prediction.

Fig. 4: Representative performance of UNet++ on **infection segmentation** of chest x-ray.

(a) Chest x-ray. (b) Target mask. (c) UNet prediction.

Fig. 5: Representative performance of UNet on **lung segmentation** of chest x-ray.

the input chest x-ray (Fig. 3a). In contrast, UNet++ struggles with prediction of the infection mask (Fig. 4b), resulting in blurred edges and irregularity along the boundary of the mask (Fig. 4c).

(a) Chest x-ray. (b) Target mask. (c) UNet prediction.

Fig. 6: Representative performance of UNet on **infection segmentation** of chest x-ray.

Similarly, examples of the performance of UNet on both the lung segmentation and infection segmentation are in Fig. 5 and Fig. 6, respectively. The predicted lung segmentation (Fig. 5c) is nearly identical to the target mask (Fig. 5b) given the input chest x-ray (Fig. 5a). Much like UNet++, UNet struggles with prediction with the target infection mask (Fig. 6b), resulting in blurred and incomplete predicted masks (Fig. 6c).

3.2 Classification Accuracy

COVID-Net reached the highest classification accuracy of the three algorithms, with a final test classification accuracy of 95.20% (Table 2). In comparison, UNet and UNet++ have relatively poor performance, with detection accuracies of 72.88% and 66.78% respectively (Table 2). We were not able to replicate the high detection accuracy (~99%) seen in [1] and [2].

Learning algorithm	Final test classification accuracy
UNet	72.88%
UNet++	66.78%
COVID-Net	95.20 %

Table 2: UNet, UNet++, and COVID-Net classification accuracy on the test set.

4 Conclusion

Using our custom COVID-Net architecture, we were able to achieve a detection accuracy of 95.20%, beating the detection accuracy of our UNet and UNet++ implemenations (72.88% and 66.78% respectively). However, COVID-Net is not able to provide any information about location or severity of infection, which limits its usefulness in comparison to easier and more accessible detection methods, such as PCR tests. The UNet and UNet++ implementations showed high accuracy in identifying lung area (IoU of 93.08% and 94.00% respectively), and lower accuracies in identifying infection area (IoU of 70.80% and 64.75% respectively). Though our implementation did not reach high accuracy in location and severity identification, this model architecture shows promising results in providing valuable infection information not available through PCR tests.

References

1. Tahir, A.M., Chowdhury, M.E., Khandakar, A., Rahman, T., Qiblawey, Y., Khurshid, U., Kiranyaz, S., Ibtehaz, N., Rahman, M.S., Al-Maadeed, S., Mahmud, S., Ezeddin, M., Hameed, K., Hamid, T.: Covid-19 infection localization and severity grading from chest x-ray images. Computers in Biology and Medicine **139** (2021) 105002
2. Degerli, A., Ahishali, M., Yamac, M., Kiranyaz, S., Chowdhury, M.E.H., Hameed, K., Hamid, T., Mazhar, R., Gabbouj, M.: Covid-19 infection map generation and detection from chest x-ray images. (2020)
3. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015)
4. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation (2018)