Predicting Severity of COVID-19 From Patients' Chest CT Images Using Self-Supervised Image Segmentation Approaches

Daryl Fung, PingZhao. Hu, Carson Leung, Qian Liu, Judah Zammit
University of Manitoba, MB, Canada

ABSTRACT COVID-19 is the new outbreak of a contagious disease that infects the lungs. Currently, no vaccines or antiviral medicines exist for COVID-19 as COVID-19 is a newly infectious disease that was first discovered around December 2019. As COVID-19 is a very contagious disease, cases appear faster than the amount of test kit available. Currently, the most common testing used is PCR(Polymerase Chain Reaction) test. These test samples are sent to a centralized lab for analysis which would take several days for the test results to be available. Due to the exponential rate of infections, the limited amount of test kits, and the long wait time for the test results to be available, many infected patients are unable to get tested and receive treatments. An alternative approach to test for COVID-19 patients is through computerized tomography (CT) scan of the lungs. CT scan can drastically reduce the time taken for test results to be available and this could speed up the testing time as well as the limiting number of testing kits available. We will propose a deep learning architecture that can evaluate different segmentation of the lungs from CT images to detect if a patient is infected with COVID-19 so that we can reduce the amount of time taken to carry out testing to determine if patients are infected with COVID-19. In addition, we will calculate the severity of the lungs affected by the disease from the CT images. The lungs will be subdivided into different regions, a calculation of the severity of each region of the lungs will be carried out through evaluation of CT severity score (CT-SS) or Dice Similarity Coefficient (DSC).

INDEX TERMS Deep Learning, COVID-19, Machine Learning, CT Lungs Images, Self-Supervised

IMPACT STATEMENT The authors should include here a significance statement of no more than 30 words. The statement should summarize the main findings of the research work reported in the manuscript.

I. INTRODUCTION

COVID-19 is a newly identified disease that is very contagious and has been rapidly spreading across different countries around the world. The virus that was first identified in Wuhan has now infected more than 3.5 million people around the whole world and causes more than 245,000 deaths. Common symptoms from COVID-19 are fever, dry cough, but in more serious cases, patients can experience difficulty in breathing. As more people are infected, communities that have been in close contact with infected patients are getting tested for COVID-19. The test used to carry out the test for COVID-19 uses PCR(Polymerase Chain Reaction) test which could take several days for the test results to be available as the test samples are sent to a centralized lab for analysis and can be time consuming. There is a limited number of supplies of PCR tests which is a bottleneck for testing to be efficient. Several alternative methods have been considered to test patients that are COVID-19 positive including CT scan of the lungs. CT scans of the lungs are faster and easier to detect COVID-19 presence in patients. As the number of infected patients increases exponentially, it can be hard to provide testing scans for patients because of the limited number of doctors. It is recommended that Artificial Intelligence systems are used to analyse the CT scans of lung patients to determine the severity of COVID-19 and monitor the disease progression as well as to compensate for the high number of patients. Specifically, we propose using deep learning to analyze and create a pixel-level segmentation of CT scan images of patients' lungs to determine the severity of COVID-19 in their lungs.

II. RELATED WORKS

There are several works that have been proposed to create image segmentation for CT scan lung images of COVID-19 positive patients. They have demonstrated effective solutions using deep neural networks to accurately predict if a patient has COVID-19 positive or negative.

A study has been conducted that uses multiple models for different tasks where the study uses both classification and image segmentation tasks for COVID-19 detection through multi-tasks learning. The study uses Inception Residual Recurrent Neural Network (IRRCNN) for the classification of COVID-19 detection and uses Nabla-Net (NABLA-N) network for infected region segmentation for X-ray and CT images scan. [1] Transfer learning is used to retrain the IRRCNN model with samples to differentiate between COVID-19 positive samples and negative samples in the classification phase. Mathematical Morphological approaches are implemented for selecting appropriate contours for chest region selection in the segmentation phase with NABLA-N network. Some classical imaging and adaptive threshold approaches are applied to

extract the features to identify infected regions of COVID-19. They used a total number of 5,216 samples of which 3,875 samples are pneumonia and 1,341 samples are normal.

Another study [2] introduces a feature variation block and progressive atrious spatial pyramid pooling block using COVID-segNet, a high accuracy network that is able to create segmentation of COVID-19 infection from chest CT images. The network consists of an Encoder and a Decoder with residual skip connection connecting the encoder and the decoder at their respective layer, following the architecture of UNET [3]. Their main findings include the introduction of an FV block and a PASPP block. FV block consists of three branches - contrast enhancement branch, position sensitive branch, and identity branch. These branches can enable automatic change of parameters to display positions and boundaries of COVID-19. The PASPP block takes features extracted from the FV block to acquire semantic information with a variety of receptive fields. The dataset that they used consists of 21,658 labeled chest CT images, of which 861 CT images are confirmed COVID-19.

Both the papers above however conducted the study with a good amount of data samples to train the network to achieve a high performance. The papers are only able to recognize the presence of COVID-19 in a patient, but the papers could not quantify the severity of the disease.

While there is a limited number of public data samples available for CT COVID-19 lung images, it will not be feasible to train a network to achieve high performance. There are a different number of research that resolve this issue. One method is to use self-supervised learning to mitigate the problem of having a low number of data samples to improve the performance of deep neural networks. Instead of having to manually annotate the data, self-supervised learning instead finds and exploits the relationship between data samples.

Novosel et al. [8] used elf-supervised learning to boost the performance of a supervised pixel-level segmentation task. They used image colorization and depth prediction to improve the feature representation from the encoder network to improve the performance of the network by learning multi-tasks of the image domain. They show that self-supervised learning by training the network on semantic segmentation, image colorization and depth prediction instead of solely training on semantic segmentation improves the performance of the network.

Another study [9] uses fine-grained segmentation networks (FGSN) to produce a dense segmentation map. Instead of using manually annotated labels, they created labels in a self-supervised manner. Features were extracted in certain layers in the network and clustered using k-means clustering. The cluster assignments are then used as supervision for the training. They showed that the learned representations can contain useful information for visual localization. The performance of the model improved by as much as 15

III. Problem Statements

The related works on segmentation for COVID-19 was trained on an abundant amount of annotated data samples of CT lung images. Getting a high performance in deep neural networks requires an abundant amount of annotated samples. Performance can be drastically reduced if there are not enough data samples to compensate for the model's complexity. Likewise, complex data distributions to learn require a higher model complexity to be able to fit the distribution with better performance. As pixel-level segmentation on CT images is a complex task, pixel-level segmentation requires a high model complexity to fit the distribution. Unfortunately, there is a limited number of publicly available COVID-19 dataset especially in the form of pixel-level segmentation. The limited number of samples available greatly reduce the performance of modeling complex distribution for pixel-level segmentation of CT scans lung images.

The related work does not also consider the severity of the lungs of patients as a result from COVID-19. We will propose a model and technique that utilizes self-supervised learning to mitigate the limited number of publicly available COVID-19 CT lung images samples as well as a method to calculate severity score of the segmented regions of CT lung images.

IV. Methodology

In this section, we will propose the details of the baseline model that we will be comparing with including the network architecture, the data preprocessing steps, and the loss function. We will then show our proposed methods using self-supervised learning and how it helps to improve generalisation and performance of the model while having a limited number of data samples. We will also show the extension our data preprocessing steps which further improves the performance of our model.

A. Baseline

The baseline that we will be comparing with is Inf-Net (Lung Infection Segmentation Network) [14] without using any semi-supervised learning algorithm. This is to show that the self-supervised learning method improves the performance of the baseline supervised learning Inf-Net.

The network architecture of Inf-Net is shown in 1. The CT images are first passed through two convolution layers to extract the low-level features. The low-level features extracted will then be passed into three convolutional layers to obtain the high-level features. The low-level features are also passed into an edge attention module that helps to improve the representation of different boundaries for the segmentation. The high-level features from the three convolution layers are passed into a *parallel partial decoder* (PPD) to aggregate the features and generate a map for the lung infections. The features generated by the PPD will then be calculated together with the high-level features generated by the previous three convolutional layers with a *reverse attention* module. The equation of R is obtained by:

$$R_i = Concat(f_i, Downs(e_{att})) \odot A_i \tag{1}$$

where f_i refers one of the high-level features generated by one of the three convolutional layers, e_{att} refers to the

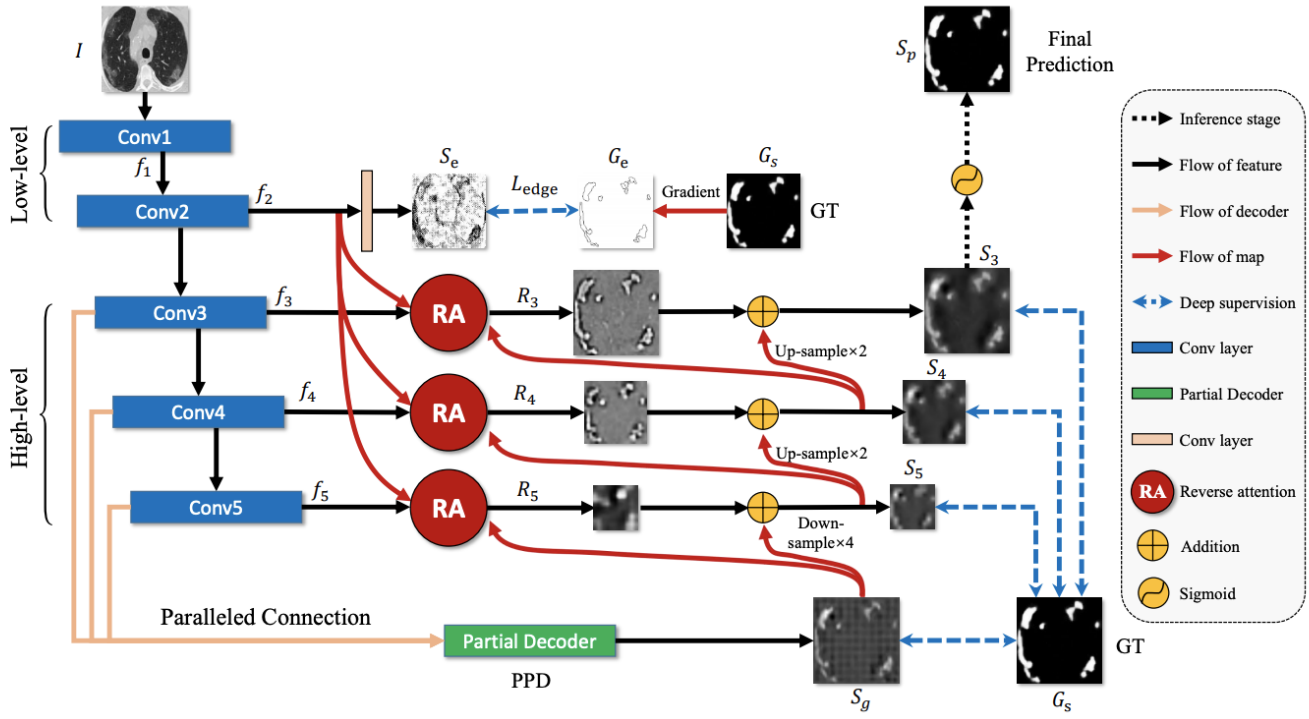

Fig. 1. The architecture of the inf-net model [14].

features before feeding into the edge attention module. The A_i is obtained by:

$$A_i = \eta(\ominus(\sigma(Upsample(S_{i+1})))) \tag{2}$$

where S is the smaller predicted output from one level higher. σ is the sigmoid operation to convert the output to range between [0, 1]. $\ominus$ is the reverse operation by substracting the input from a matrix with all ones. η is the expansion of the single tensor to 64 repeated tensors, reversing each candidate of the tensor.

There are several loss functions proposed for this baseline model. The first loss function is the loss edge, L_{edge} which guides the model in representing better segmentation boundaries. The other loss function is the segmentation loss, L_{seg}. The segmentation loss combines both the loss of Intersection over Union (IoU) and the binary cross entropy loss. The segmentation loss equation is as follows:

$$L_{seg} = L_{IoU} + \lambda L_{BCE} \tag{3}$$

The λ is set to 1 for this experiment. The segmentation loss is adapted to all of the S_i predicted output where S_i are created from f_i such that $i = 3,4,5$. The pseudo-code for the training of baseline model is relatively straightforward as can be seen in 1.

The total loss function for the baseline Inf-Net model is then:

$$L_{total} = L_{seg}(G_t, S_g) + L_{edge} + \sum_{i=3}^{5} L_{seg}(G_t, S_i) \tag{4}$$

The summation of the loss functions are calculated from the output of the three convolutional layers. G_t refers to the ground truth labels. S_g is the output from the parallel partial decoder to match with the ground truth label.

Algorithm 1 Pseudo code for Inf-Net

Input: Labeled data $D_{labeled}$
for each batch of $D_{labeled}$: **do**
  $P_{labeled}$ = Preprocess $D_{labeled}$
  Perform training of baseline Inf-Net, M, on $P_{labeled}$
  calculate test performance of M on testing labeled data.
  save model weights, w.
end for

B. Self-supervised method

In order to mitigate the problem of a limited number of data samples available and to aid in the generalisation of deep learning models, we determine to use self-supervised learning to learn good representations of the CT scan of lung images. Self-supervised learning generates auxiliary tasks from the labeled data samples. For instance, when undergoing data augmentation with rotation, we could train the network to predict if the images have been rotated 0 degree, 90 degree, 180 degree to learn representations of the images.

We will propose using a self-supervised method to improve the performance of deep neural networks to create pixel-level segmentation for CT scan for lung images of COVID-19 patients. We will integrate self-supervised inpainting to pre-train our network. As shown in [25], image inpainting is similarly related to image segmentation. By learning features from Image inpainting, the model can learn more features that

are related to image segmentation. As creating mask can be a complex task for the network to learn to inpaint, the mask can either be too complex for the network to start learning or too simple to be able to learn good representations. We will be using a coach network [25] that increases the complexity of the masking of the CT images throughout the training of the network. The mask created will initially be relatively simple, once the network is able to predict the inpainting of the CT images with good performance, the coach will increase the complexity of the masking to reduce the performance of the network, similar to how Generative Adversarial Network (GAN) [20] works. The loss fuction for the coach network is:

$$L_{coach}(x) = 1 - L_{rec}(x \odot M) \tag{5}$$

where $M = C(x)$ which is created by the coach network. A constraint is apply to this loss function because the coach network would just create a mask that masks all region because no context information would be present for the network to learn and a maximum loss will be achieved. The constraint is:

$$\hat{B}(x) = B(x) - SORT(B(x))^{k|B(x)} \tag{6}$$

$$M = C(x) = \sigma(\alpha \hat{B}(x)) \tag{7}$$

The backbone, B, of the coach network has a similar network architecture with the model that inpaints the CT images. SORT(B(x)) sorts the features in descending order over the activation map. k represents the k^{th} elements in the sorted list and k helps to control the fraction of the image to be erased. The region that has scores lesser than the k^{th} element will be erased from the images. If k is 0.75 then 0.75 fraction of the images will not be erased. The score is scaled into a range of [0, 1] using a sigmoid activation function. We keep $\alpha = 1$ while training the coach network.

Algorithm 2 Pseudo code for self-supervised with Inf-Net

Input: Labeled data $D_{labeled}$
for each epoch **do**
  for each coach step **do**
    Generate Mask using coach network
    Use the generated mask to masks the CT images
    calculate the loss for coach network
    backpropagate and adjust the mask accordingly
  end for
  for each network step **do**
    Preprocessed $D_{labeled}$ with data augmentation to create $P_{labeled}$
    Generate inpainted output using inf-net network with last layer replaced by an inpaint layer
    Calculate inpainting loss and back-propagate
  end for
end for
for each batch of $D_{labeled}$: **do**
  $P_{labeled}$ = Preprocess $D_{labeled}$
  Perform training of baseline Inf-Net, M, on $P_{labeled}$
  calculate test performance of M on testing labeled data.
  save model weights, w.
end for

We will also implement different data augmentation that includes random cropping, rotation, and random cutout [7], [15], [16] to increase the number of available annotated data samples and labels as well as improve the model generalization as the data samples for CT images from COVID-19 can be limited.

Once the network is able to predict the pixel-level segmentation of the CT scan images, we will determine the severity score of the lung regions through several methods. The severity score of the lungs can be determined by using CT severity score (CT-SS) [11]. The score uses lung opacification for extension of the infections in the lungs. CT-SS is an adaptation from the method previously used in patients after severe-acute respiratory syndrome (SARS) [10] to describe ground-glass opacity, interstitial opacity, and air trapping. The lungs will be divided into 20 regions, the posterior apical segment of upper left lobe was divided into apical and posterior segmental regions, the anteromedial basal segment of lower left lobe will be divided into anterior and basal segmental regions. For each region, there contain a system attributing scores of 0, 1, 2, either parenchymal opacification involves 0

Another method to determine the severity score is to use Dice similarity coefficient (DSC) [12] to evaluate the overlap ratio between the automatically segmented regions by the network and the reference region provided by radiologists. The equation is as follows:

$$DSC(R,S) = \frac{2|R \cap S|}{|R| + |S|} \tag{8}$$

Where R refers to the reference region provided by the radiologists and S refers to the automatically segmented regions by the network. $|.|$ is the operator that is used to calculate the number of voxels in a given region.

We will compare our method against supervised and semi-supervised [13], [14] models trained on COVID-19 dataset. For comparing supervised learning, we will compare against the paper [13]. We will train and follow using the same network structure but change from supervised learning to self-supervised learning and compare the performance between supervised and self-supervised.

When comparing with the semi-supervised model, we determine that our model is successful if our model is able to reach close to or better than the performance of the semi-supervised model as semi-supervised model is able to obtain a higher amount of data samples by looking at both unannotated and annotated data samples while self-supervised model only have access to the annotated labels. A self-supervised learning method will create its own training annotated labels without any manual human labelling and trained without any unlabeled data samples. We will compare our method's performance against Inf-Net [14] which uses semi-supervised learning by generating pseudo labels from randomly selected unlabeled CT images.

Our method will be novel compare to the other methods mentioned as our method will be integrating both the segmentation of the CT lung images as well as the calculation of the severity score through caluclation of the segmented infected

lung areas.

V. Experiments

A. Datasets

The dataset that we will be using is an integrative resource of chest computed tomography images and clinical features of patients with COVID-19 pneumonia (ICTCF) [23] which contains the severity score for each CT lung image. ICTCF contains 127 types of clinical features and laboratory confirmed cases of COVID-19 from 1170 patients. The dataset can be found here: http://ictcf.biocuckoo.cn/ .

The datasets that we will be using to compare with for the segmentation results will be the datasets from Inf-Net [14] and the datasets from medseg [26]. The dataset from Inf-Net [14] contains 68 labeled CT images while the dataset from medseg contains 911 labeled CT images. In total, the number of labeled segmented CT images is 979.

The other datasets that we will be focusing on are COVID-CT-Dataset [21] and covid-chestxray-dataset [22]. COVID-CT-Dataset can be found at https://github.com/UCSD-AI4H/COVID-CT. COVID-CT-Dataset consist of 349 CT images obtianed from 216 patients. Covid-chestxray-dataset can be found at https://github.com/ieee8023/covid-chestxray-dataset. COVID-CT-Dataset was created through assembling medical images from publications and websites and contains 123 frontal view X-rays of the lungs. An additional dataset that we can use is from Cell [24], http://ncov-ai.big.ac.cn/download?lang=en if we have enough capacity to load the dataset as the dataset can be huge. This dataset is constructed from the China Consortium of Chest CT Image investigation (CC-CCII). The CT images are classified into novel coronavirus pneumonia (NCP) caused by SARS-CoV-2 virus, common pneumonia and normal controls. It consists of 617,775 CT images obtained from 4,154 patients.

B. Baseline

We have trained the supervised inf-net as baseline with comparison to the supervised inf-net with additional data augmentation. The data augmentation that we used includes *vertical flipping, horizontal flipping, random crop, and random cutout*. The comparison for the methods can be seen in I. We can see that the additional data augmentation improves the performance of the baseline model. The performance of the multiple segmentation inf-net improves by more than 50 percent with data augmentation of random cutout with a value of 0.5.

VI. Conclusion

A conclusion section is required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the major findings and significance of the work or suggest applications and extensions. Do not exceed 300 words for the conclusion section.

	single Inf-Net	multi Inf-Net
Baseline	7.95	1.17
Baseline+data aug (0.2)	5.09	0.78
Baseline+data aug (0.4)	5.09	0.71
Baseline+data aug (0.5)	**5.09**	**0.66**

TABLE I. comparison of baseline model with added data augmentation. The floating value after the data augmentation refers to the fraction of the image randomly cutout. 0.2 shows that 0.2 of the image is cutout to be empty.

References

[1] Alom, MZ., Rahman, MMS., Nasrin, MS., Taha, TM., Asari, VK. *COVID-MTNet: COVID-19 Detection with Multi-Task Deep Learning Approaches.* arXiv:2004.03747, 2020.

[2] Yan, Q., Wang, B., Gong, D., et al. *COVID-19 Chest CT Image Segmentation – A Deep Convolutional Neural Network Solution.* arXiv:2004.10987, 2020.

[3] Ronneberger, O., Fischer, P., and Brox, T. *U-net: Convolutional networks for biomedical image segmentation.* In MICCAI, pages 234–241. Springer, 2015. 2

[4] Kalluri, T., Varma, G., Chandraker, M., and Jawahar, CW. *Universal semi-supervised semantic segmentation.* CoRR, abs/1811.10323, 2018.

[5] Misra, I., and van der Maaten, L. *Self-supervised learning of pretext-invariant representations.* arXiv preprint arXiv:1912.01991, 2019.

[6] Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. *A simple framework for contrastive learning of visual representations.* arXiv:2002.05709, 2020.

[7] Newell, A., Deng, J. *How Useful is Self-Supervised Pretraining for Visual Tasks?* arXiv:2003.14323, 2020.

[8] Novosel, J., Viswanath, P., and Arsenali, B. *Boosting Semantic Segmentation With Multi-Task Self-Supervised Learning for Autonomous Driving Applications.* In Proc. of NeurIPS - Workshops, pages 1–11, Vancouver, BC, Canada, Dec. 2019.

[9] Kahl, F. *"Fine-grained segmentation networks: Self-supervised segmentation for improved long-term visual localization,"* in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 31–41.

[10] Chang, YC., Yu, CJ., Chang, SC., et al. *Pulmonary sequelae in convalescent patients after severe acute respiratory syndrome: evaluation with thin-section CT.* Radiology 2005; 236(3):1067-1075.

[11] Yang, R., Li, X., Liu, H., Zhen, Y., Zhang, X., Xiong, Q., et al. *Chest CT Severity Score: An Imaging Tool for Assessing Severe COVID-19. Radiol Cardiothorac Imaging.* 2020;2(2):e200047.

[12] Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., Xue, Z., and Shi, Y. *Lung Infection Quantification of COVID-19 in CT Images with Deep Learning.* arXiv preprint arXiv:2003.04655, 1-19, 2020.

[13] Yan, Q., Wang, B., Gong D., et al. *COVID-19 Chest CT Image Segmentation – A Deep Convolutional Neural Network Solution.* arXiv preprint arXiv:2004.10987, 2020.

[14] Fan, DP., Zhou, T., Ji, GP., et al. *Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Scans.* arXiv preprint arXiv:2004.14133v2, 2020.

[15] Alexander Kolesnikov, Xiaohua Zhai, and Lucas Beyer. *Revisiting self-supervised visual representation learning.* In Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[16] Trinh, TH., Luong, MT., and Le, QV. *Selfie: Self-supervised pretraining for image embedding.* arXiv preprint arXiv:1906.02940, 2019.

[17] Frinken, V., Zamora-Martinez, F., Espana-Boquera, S., Castro-Bleda, M. J., Fischer, A., and Bunke, H. (2012). *Long-short term memory neural networks language modeling for handwriting recognition.* In Pattern Recognition (ICPR), 2012 21st International Conference on, pages 701–704. IEEE.

[18] LeCun, Y., Haffner, P., Bottou, L., and Bengio, Y. *Object recognition with gradient-based learning.* In Shape, contour and grouping in computer vision, pages 319–345. 1999.

[19] Kingma, Diederik, P. and Welling, M. *Auto-Encoding Variational Bayes.* In The 2nd International Conference on Learning Representations (ICLR), 2013.

[20] Goodfellow IJ., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, AC., and Bengio, Y. *Generative adversarial nets.* In Proceedings of NIPS, pages 2672– 2680, 2014.

[21] Zhao, JY., Zhang, YC., He, XH., Xie, PT. *COVID-CT-Dataset: a CT scan dataset about COVID-19.* arXiv preprint arXiv: 2003.13865, 2020.

[22] Cohen, JP., Morrison, P., and Dao, L. *COVID-19 Image Data Collection.* arXiv preprint arXiv: 2003.11597, 2020. https://github.com/ieee8023/covid-chestxray-dataset.

[23] Ning, Lei, WS., Yang SJ., et al. (2020). *iCTCF: an integrative resource of chest computed tomography images and clinical features of patients with COVID-19 pneumonia.* 10.21203/rs.3.rs-21834/v1.

[24] Zhang, K., Liu, XH., Shen, J., et al. *Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements and Prognosis of COVID-19 Pneumonia Using Computed Tomography.* DOI: 10.1016/j.cell.2020.04.045.

[25] Singh, S., Batra, A., Pang, G., Torresani, L., Basu, S., Paluri, M., and Jawahar, C. V. *Self-supervised feature learning for semantic segmentation of overhead imagery.* In BMVC, 2018.

[26] *COVID-19 CT segmentation dataset*. Retrieved from http://medicalsegmentation.com/covid19/.