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, REMOVE THIS: TODO: update abstract, add multi-seg figure, add severity score performance

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. In order to obtain the severity score, the model is first trained to segment the infected region of the lungs. Then, t he severity score will be calculated by calculating the overlapping ratio between the segmented region for infected regions and the parenchyma of the 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.

The paper above however conducted the study with a good amount of data samples to train the network to achieve a high performance. They obtained their dataset from hospitals through obtaining permission. We would like to create a network that does not require much labeled dataset to be able to achieve good performance. By doing this method, we could bring this network forward to detect new lung diseases when there are not much dataset available. Besides that, the paper is 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 segmentation, 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 semi-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, semi-supervised learningutilizes the unlabeled data samples to aid in the training for the network.

Deng Pin Fan et al. [14] used semi-supervised learning to enlarge the limited number of training samples for CT lung image segmentation. They developed a model called InfNet and semi-InfNet. The InfNet version of the model uses fully supervised method to predict the segmentation of the CT images for ground-glass opacities and consolidations. The model outputs 4 images of the segmentation for the CT lung images that contains either ground-glass opacities or consolidations with different image sizes. The segmentation of the different image sizes are resized to the same size as the ground truth of the segmentation to compute the loss function. They also uses a edge loss to guide the model to predict the bounding area of the segmentation. To improve InfNet, they use semi-supervised by progressively enlarging the training dataset with unlabeled data using random sampling strategy. Specifically, they generate pseudo labels for unlabeled CT lung images. The advantage of using semi-supervised learning is that we can generate pseudo labels to increase the number of data samples. However, semi-supervised learning still require the unlabeled CT images before being able to undergo its learning procedure.

Another study [27] uses Task-Based Feature Extraction Network (TFEN) and Covid-19 Identification Network (CIN). They propose to use task-specific feature extraction network that is tailored to CT lung images with three different classes: Healthy, pneumonia, and COVID-19 cases. They also mentioned that dataset for COVID-19 is still limited and there is not enough high quality dataset. They treat the task-specific feature extraction network as autoencoders and train the overall TFEN module to extract the relevant features from the CT images. Then, they use CIN to perform classification on the extracted features from the TFEN module. Due to the fact that by providing a person with limited CT images, they can easily detect the abnormal regions and differentiate between them very accurately by making use of prior information. This helped them develop a semi-supervised feature extraction network that allows obtaining the relevant prior information to perform the classification in order to mimic human behaviours. However, this study does not undergo segmentation of the CT lung images for better diagnosis of the CT lung images. It also does not provide the severity score of the CT lung image.

There is a study that predicts the severity score of COVID-19 on chest x-ray with deep learning [28]. They use a DenseNet model from the TorchXRayVision library as DenseNet models have been shown to predict Pneumonia well. They use a pre-training step to train the feature extraction layers and a task prediction layer. The pre-training step was used to generate a general representations of lungs and other CXRs that they would have unable to achieve from the small set of COVID-19 images available. They use a network that outputs 18 outputs of a representation of the image, 4 outputs that are a hand picked subset which contain the radiological findings (pneumonia, consolidation, lung opacity, and infiltration), and a lung opacity output. This study however did not use infected region segmentation to predict the severity score. They do not use self-supervised learning but pre-training steps to counter the limited data samples available for COVID-19.

III. Problem Statements

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. The related works utilizes semi-supervised learning to increase the amount of data samples to achieve higher 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

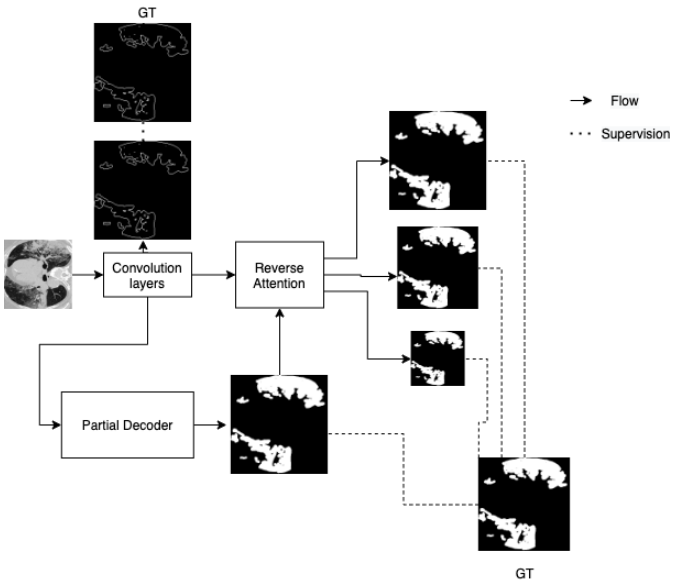

Fig. 1. The architecture of the supervised InfNet. The CT lung image is first passed into the convolutional layers to extract the features of the CT lung image. Then, the features generated from the convolutional layer are fed into the partial decoder module, reverse attention module and the edge detection module. The edge detection module is to help the network with detection of the boundaries of the segmentation. The reverse attention and the partial decoder generates the segmentation of the infection regions of the CT lung images.

available greatly reduce the performance of modeling complex distribution for pixel-level segmentation of CT scans lung images.

The related works 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 first show the details of the supervised InfNet for imaging segmentation (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 of our data preprocessing steps which further improves the performance of our model.

A. Supervised InfNet for imaging segmentation

Supervised InfNet (Lung Infection Segmentation Network) will be used as our baseline [14] to compare 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 InfNet for imaging segmentation. We will extend our work on supervised InfNet by adding self-supervised method to it.

We will not change the structure of the InfNet model and use the default parameters as included in their GitHub code. There will be two different types of the InfNet model. The first one will create a single-labeled segmentation of the image for the infected region. This model predicts as true of the region is either ground-glass opacities or consolidations. This will act as a prior to be fed into the multple-labeled segmentation model concatenate with the original CT image as proposed by the author of InfNet. The multiple-labeled segmentation model will be used to predict the background, ground-glass opacities, and consolidations for the infected region. The multiple-labeled segmentation model will give each of the label a different value instead of grouping them as one as what the single-segmentation model does.

B. Self-supervised InfNet for imaging segmentation

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:

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

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:

$$\hat{B}(x) = B(x) - SORT(B(x))^{k|B(x)} \tag{2}$$

$$M = C(x) = \sigma(\alpha\hat{B}(x)) \tag{3}$$

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.

After the self-supervision training is finished, the single segmentation InfNet and multi segmentation InfNet net-

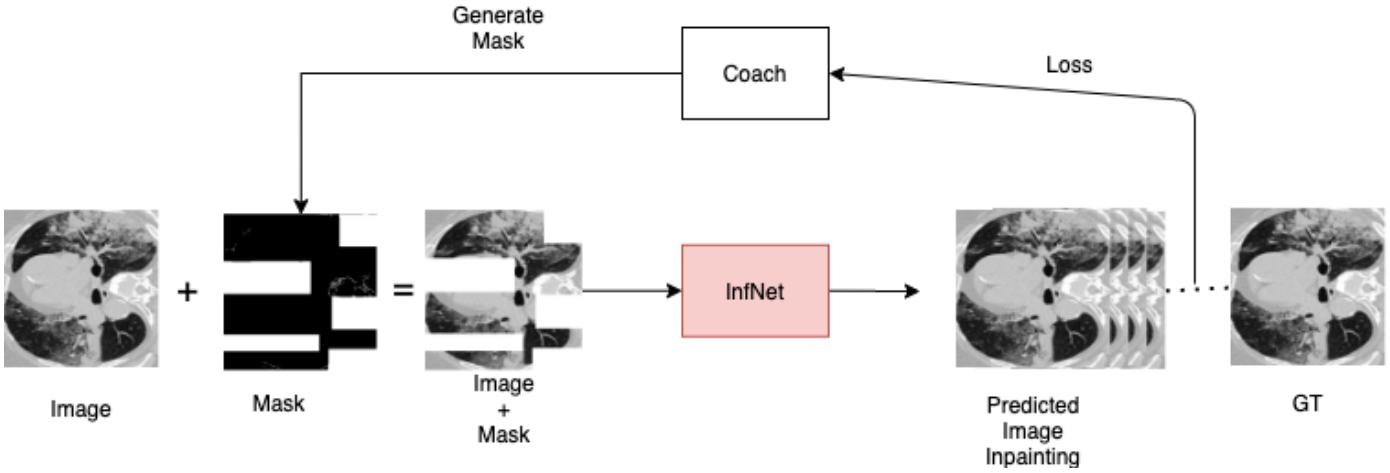

Fig. 2. The architecture of the coach network for self-supervised inpainting. The loss for the coach network is constructed from the loss of the image inpainting from the InfNet. The coach network and the InfNet both work together as a MinMax algorithm. The InfNet will try and minimize the loss to generate better image inpainting while the coach network will try to increase the loss of the image inpainting through generating more complex masks. In the beginning, the masks generated by the coach network will be less complex. Through training of the coach network, as the InfNet gets better at predicting image inpainting, the coach network generates a more complex masks.

work would reuse the weights that were trained during self-supervised training to train normally on the segmentation for the infected regions.

The proposed self-supervised single-labeled segmentation InfNet network architecture can be seen in 3. The last layer for each output prediction is replace to a different linear activation layer. The linear activation layer will re-create the original image that is covered by the masks.

The proposed self-supervised multi-labeled segmentation InfNet network architecture is shown in 4. The changes in the architecture for the multi-labeled segmentation InfNet is similar to the single-labeled segmentation InfNet where the last layer of the layer is replace with a different linear actiivation layer to output the inpainting of the original image.

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. We will show that proper data augmentation affect the performance of the model by a marginal amount.

The output of the single segmentation InfNet will include the edge of the segmentation and four single-labeled segmentation of the infected region of the CT lung images with different sizes as shown in 1. A loss will be calculated for each of the output from the InfNet 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:

$$L_{seg} = L_{IoU} + \lambda L_{BCE} \tag{4}$$

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 $i = 3,4,5$.

Algorithm 1 Pseudo code for self-supervised with InfNet

Input: $D_{labeled}$ = [($inputImage_1$, G_{t1}), ...]
for each epoch **do**
  for each coach step **do**
    mask = M(x)
    maskedInput = $mask \odot inputImage$
    $predictedImage = network(maskedInput), inputImage$
    $L_{rec} = CrossEntropy(predictedImage, inputImage)$
    $L_{coach}(x) = 1 - L_{rec}$
    update coach weights
  end for
  for each network step **do**
    $P_{labeled} = Preprocess(D_{labeled})$ // data aug
    $inpaintingOutput = network(P_{labeled})$
    $L_{rec} = CrossEntropy(InpaintingOutput, inputImage)$
    backpropogate and save network weights
  end for
end for
for each batch of $D_{labeled}$: **do**
  $P_{labeled}$ = Preprocess ($D_{labeled}$)
  $trainLoss = train(P_{labeled})$
  Backpropagate train loss
  $testLoss = test(P_{labeled})$
  save model weights, w.
end for

The total loss function for the baseline InfNet model is then:

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

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.

As for the multiple segmentation infected region InfNet.

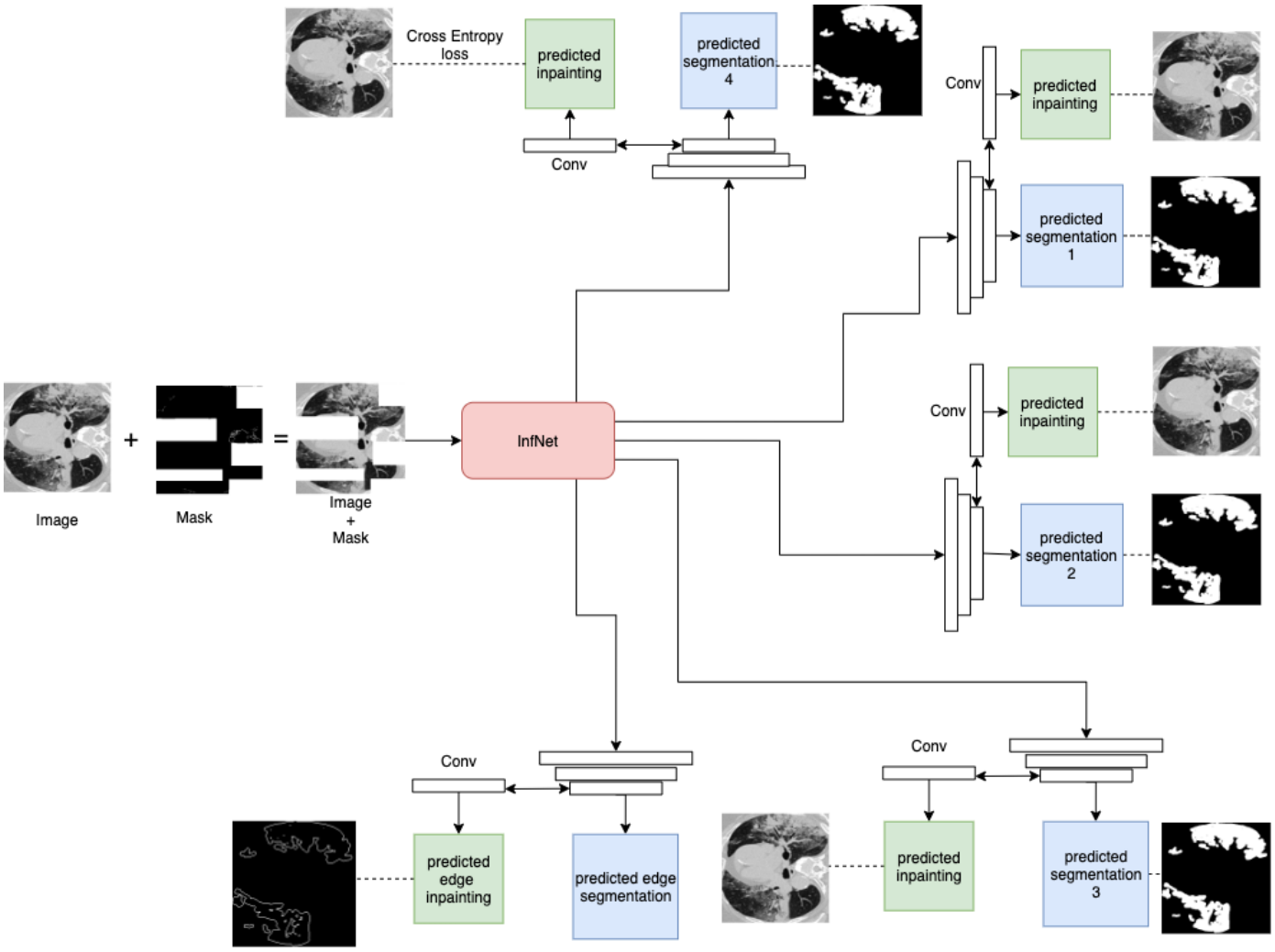

Fig. 3. The architecture of our self-supervised InfNet model. The original InfNet model would generate 5 different predictions: the edge segmentation prediction, and the other 4 are segmentation of the infected regions but of different sizes. In order to utilise the ability of self-supervised method for InfNet segmentation, we generate masks to be fed into the InfNet model. The last layer for each output prediction is not used for the self-supervised case. However, the last layer is replace with a different layer to reconstruct the image and the edge appropriately. Everything else is kept the same as the InfNet architecture. This way the network will learn meaningful representations of the CT images and we can use these meaningful representations to learn the segmentation of the infected regions of the CT lung images. After learning the self-supervised features for InfNet, the training continues as normal similar to the InfNet algorithm. The training will start with the weights trained using the self-supervised inpainting method. The last layer will be changed to its original layer instead of the replaced convolutional layer.

We also use the default model and hypermaters from the InfNet code. We will however train the network without using any unlabeled images to be used as a supervised version. The CT lung images and prior for the CT lung images are concatenate together before being fed into the multiple segmentation InfNet. The prior is generated from the single segmentation InfNet. The prior would contain the area of the infected region. However, the prior does not contain the labels for ground-glass opacities and consolidations. It just shows the infected regions. The multiple segmentation InfNet will label the CT lung images with background, ground-glass opacities, and consolidations. The architecture for multiple segmentation InfNet can be seen in 4.

In order to improve the performance of the model and to aid in the generalisation, 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.

C. Estimation of Severity of COVID-19 from CT images

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

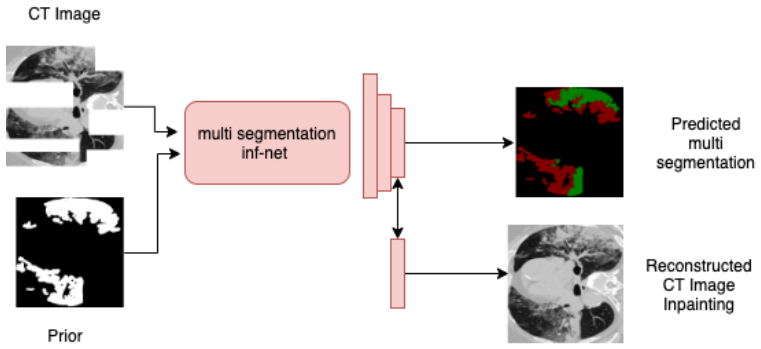

Fig. 4. The architecture of our self-supervised multi segmentation InfNet model. Similar to the supervised version of multiple InfNet, we use the same network weights and architectures. However, the last layer of the multiple segmentation InfNet is replace with a linear activation layer to predict the inpainting for the CT images. After the self-supervised InfNet has been trained, we will train the multiple segmentation for multiple segmentation InfNet using the weights of the self-supervised InfNet. The last layer however will be changed to its original form instead of using the replaced linear activation layer. This way, all the network weights are loaded from the self-supervised version except for the last layer.

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:

$$DSC(R,S) = \frac{2|R \cap S|}{|R| + |S|} \tag{6}$$

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 InfNet [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.

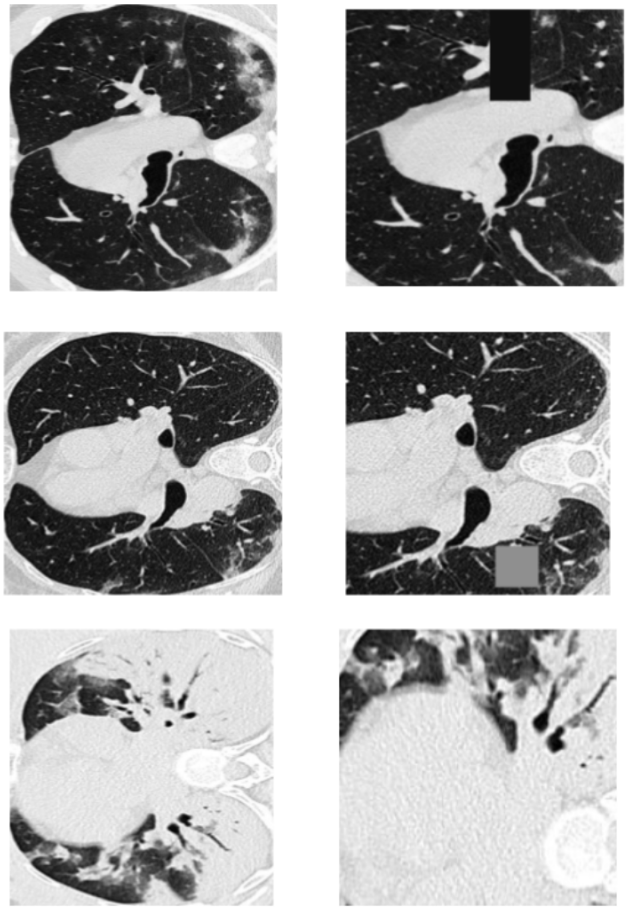

Fig. 5. Example CT lung images of our data augmentation. The left column is the original CT lung images while the right column is the augmented CT lung images. The first row of the CT lung images involves random cropping and random cutout. The second row of the CT lung images involves random cropping and random cutout. The third row of the CT lung images involves random cropping and vertical flipping. We can see the random cutout involves patching the image with colors of the same value of rgb. For instance, if the value of r is 10, then the value of g and b are also 10. If the value of r is 50, then the value of g and b are also 50.

B. Data Augmentation

We used data augmentation to increase our data samples size. The data agumentation that we used includes *vertical flipping, horizontal flipping, random crop, and random cutout*. For the random cutout percentage, we experimented that 0.5 cutout of the CT lung images yield higher performance than the rest of the value. This is because entropy at 0.5 is the highest which could increase more variability of the images. Examples of the data augmentation can be seen in figure 5.

C. Performance evaluation of image segmentation

We have trained the supervised inf-net as baseline with comparison to the supervised inf-net with additional data augmentation. 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. For the single segmentation Inf-Net comparison, we can see the performance of the single segmentation Inf-Net improves the most when the random cutout is set to a value of 0.5 without any self-supervised training.

	single Inf-Net	multi Inf-Net
SInfNet	**0.60**	0.93
SInfNet+data aug (0.4)	0.67	
SInfNet+data aug (0.5)	0.65	
Self-super	0.67	
Self-super+data aug	0.71	

TABLE I. Comparison of supervised inf-net (SInf-Net) model with added data augmentation. The values listed in the table represent the test loss. 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.

	single Inf-Net	multi Inf-Net
SInfNet	0.26	0.12
SInfNet+data aug (0.4)	0.31	
SInfNet+data aug (0.5)	0.30	
Self-super	0.31	
Self-super+data aug	**0.34**	

TABLE II. Comparison of supervised inf-net (SInf-Net) model with added data augmentation. The values listed in the table represent the dice similarity coefficient. 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.

However, the loss calculation does not truly evaluate the performance of the model. We use an additional performance metrics to calculate the model performance. The performance metrics we will be using is called dice similarity coefficient. As shown in II, we see that the best performing model is the self-supervised single Inf-Net using the dice similarity coefficient.

D. severity estimation results

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.

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/.

[27] Khobahi, S., Agarwal, C., Soltanalian, M. *CoroNet: A Deep Network Architecture for SemiSupervised Task-Based Identification of COVID-19 from Chest X-ray Images.* In medRxiv, 2020.

[28]] Cohen, J. P., Dao, L., Morrison, P., Roth, K., Bengio, Y., Shen, B., Abbasi, A., Hoshmand-Kochi, M., Ghassemi, M., Li, H., Duong, T. Q. *Predicting covid19 pneumonia severity on chest x-ray with deep learning*, arXiv preprint arXiv:2005.11856.