COVID-19-Lung-CT-Lesion-Segmentation-Challenge

Ilan Valencius William Gibbons Cole Gvozdas

1. Introduction

Over 10 million people worldwide have had documented cases of COVID and many of them have died, most often as a result of fluid buildup in the lungs. As a result, doctors and researchers began looking into ways to detect and diagnose the symptoms of COVID-19 at earlier stages in order to prevent people from reaching a critical condition. To accomplish this, researches at NVIDIA, Childrens National Hospital in DC, and the National Institute of Health hosted a challenge with the goal of automatic detection and quantification of COVID-19 lesions in chest computed tomography (CT) [2].

255 teams competed in the validation and testing phase, with the top submissions achieving average dice coefficients of greater than 0.6 on the validation data [2]. As there are already accurate models available, we decided to examine how different model architectures fared at the segmentation of lung lesions without any additional modification. Altering models tailors them for a specific modality but we are interested in what models perform well on CT segmentation regardless of additional support. To do this, we will investigate four model networks, BasicUNet, DynamicUNet, VNet, and HighResNet.

2. Related Work

The baseline model provided by the MONAI team was given in the challenge description. The team used the BasicUNet from the MONAI framework as the model for comparison. They trained their model for 500 epochs. Their final Dice score on the challenge validation set was 0.6904 $\pm$ 0.1801. This was the base score to which all participants could compare their models [2]. The highest Dice score upon submission was 0.7709 $\pm$ 0.1450 [2]. Therefore, we can see that participants significantly improved upon the baseline model and score.

The VNet model, which is described more in-depth below, has been used in a wide variety of biomedical image segmentation tasks. It was used for segmentation of MRIs of prostates in the original paper that the model was published in and has been put to further use throughout the medical field [5]. One task that the model had particular success in was segmentation of MRIs of inner ear fluid space. This paper was published this year in Frontiers in Neurology. The researchers displayed the accuracy and robustness of the Vnet architecture, with their final scores being the following: "Dice overlap coefficient: 0.9 $\pm$ 0.02, Hausdorff maximum surface distance: 0.93 $\pm$ 0.71 mm, mean surface distance: 0.022 $\pm$ 0.005 mm" [1]. These results indicate that the VNet architecture may be a great candidate for lung lesion segmentation

According to Microsoft's data on high resnet, the model should be able to outperform most other models in terms of training and testing scores. It should also perform better on memory cost, but some of the early testing we've done seems to refute this idea. Microsoft also stated that the runtime for training and inference on high resnet should take a little bit longer than other models and this seemed to hold up in our initial testing [9]. As it has become a standard for pose estimation, HighResNet may be outpeformed by models tailored for cell segmentation [9].

3. Methods

Our team utilized the MONAI framework to test out our models. MONAI is a PyTorch-based framework for medical artificial intelligence [6]. They provide a variety of frameworks for deep learning models with adjustable parameters. Due to the size of the image files the specific implementation graphs of each network cannot be included in this paper and can be found here

3.1. BasicUNet

The BasicUNet architecture is provided by MONAI as a baseline metric for segmentation [8]. It involves two parts, an encoder and a decoder. The encoder involves a series of convolutions and pooling layers. More specifically, the Basic UNet model uses pairs of 3x3 convolutions, each followed by a ReLu activation layer and lastly, a max pooling layer. The decoder is essentially the exact opposite of the encoder layer, with multiple series of 3x3 convolutions followed by up-convolutional layers. The main intuition within the decoder section of the model that makes it robust and accurate is the concatenation of the feature maps at the same level coming from the encoder side of the model. This makes the UNet model a fantastic launching point for image segmentation. For this project we instantiate a BasicUNet with a (32, 32, 64, 128, 256, 32) feature map.

3.2. DynamicUNet

DynamicUNet (DynUNet) is an implementation of the no-new UNet (nnUNet) architecture proposed by Isensee et al. [3]. It builds upon the UNet architecture with the intuition that DynUNet automatically adapts itself to the specifics of a given dataset. This is accomplished in a variety of ways but for this project this was accomplished by the addition of residual connections within the convolutional blocks. This allows stochastic descent gradients to more directly alter the convolutional blocks.

3.3. VNet

The Vnet was originally created in 2016 with the intention of being one of the first CNNs built for 3D images. Due to the inherent 3D nature of any type of medical scan, whether it be MRI, CT, PET, etc., it would make sense to train models using the full 3D inputs rather than a bunch of 2D slices of those 3D images. Previously, all image analysis or segmentation was done through processing hundreds of 2D slices of the scans, which may pose issues when it comes to model accuracy, biases, robustness, etc. The Vnet is based off of the Basic UNet described above, yet adapted for 3D inputs. Some of the adaptations in the VNet include 3D convolutions instead of 2D ones, replacement of pooling layers with convolutional ones, and using PreLu as the nonlinearity activation function. The 3D convolutions involve kernels of 5x5x5 voxels and the replacement of pooling layers removes the need for back-propagation, leading to decreased amount of memory required during training. Overall, the VNet is an intuitive derivative of the Basic UNet that may prove best fit for our image segmentation.

3.4. HighResNet

HighResNet was created by Li et al. to address the effieciency bottlenecks of other 3D segmentation algorithms. This is accomplished through the use of dilated convolutions which allow for greater spatial resolution in convolution blocks when using very dense datasets, such as CT scans [4]. Unfortunately, this network was not able to be evaluated as loading the full model into GPU memory was infeasible. To load the model, convolutional blocks had to be stripped and the patch size decreased to such an extent that it defeated the purpose of inter-model comparison.

4. Experiments

4.1. Dataset

We trained all the networks on the challenge dataset provided. This consists of chest CT scans of 219 patients who have all tested positive for COVID-19 using a Reverse Transcription Polymerase Chain Reaction (RT-PCR). CT scans are drawn from the "CT Images in COVID-19" dataset and "COVID-19-AR" dataset [2]. Each training sample consists of a full patient CT scan at a resolution of 0.79mm x 0.79mm x 4.8 mm for a voxel volume of $3mm^3$. The number of CT scans varies by patient and lesions have been annotated by a team of radiologists in ITKSnap [2].

Name	Lr	Notes
BasicUNet-1	0.0001	bare
DynUNet-1	0.0001	bare
DynUNet-2	0.0001	res-block and trans-bias = True
DynUNet-3	0.01	new strides and kernels, res-block = True, patch size 192x192x8
VNet-1	0.0001	dropout-dim = 1
VNet-2	0.01	dropout-dim = 1, dropout-prob=0.5 patch size 192x192x16

Table 1. Training Parameters

4.2. Hyperparameters

We base our segmentation pipeline off of the provided MONAI baseline [8]. Dice Cross-Entropy loss is used for training and evaluation and the training patches were augmented by random cropping of pos neg ratio, random flipping of channels, and the addition of gaussian noise with μ=0.15 and σ=0.01 [7]. Models were optimized using Adam and run for 60 epochs. Initially, models were trained with a momentum of 0.95 and the baseline learning rate of 0.0001 but this was based on the assumption that models could be trained for 500 epochs as the baseline pipeline had done [8]. To remedy this discrepancy in epochs, learning rate was increased to 0.01. All models used a dropout probability of 0.1. For VNet-2 this was increased to 0.5 as the architecture itself cannot be altered. A 0.8-0.2 train validation split was used for training providing 160 training samples and 39 validation samples. 50 unseen samples were used as a test set. Models were trained on a NVIDIA GeForce GTX 1660 Ti using 6 Gb of memory. If this memory limit was exceed, the default patch size of 192x192x32 pixels was decreased in the z direction, preserving the 192x192 2D resolution. For each model, the weights that performed the best on the validation dataset were loaded and evaluated on the test set by determining the average and standard Dice Cross-Entropy loss across the test dataset. Results are shown in Figure 1.

5. Conclusion

Of the three models tested, the VNet architecture performed the best, maximizing DiceCE during run VNet-2. For this run, there are a variety of confounding factors compared to run VNet-1. Some are due to memory constraints such as a smaller patch size while some are heuristic such as a lower learning rate and higher dropout probability. Re-

gardless, when run on similar parameters, a learning rate of 0.0001 and dropout probability of 0.1, all three models have similar loss scores. This indicates that any of these three models may be fit to be expanded upon, which has happened. As stated in the results of the challenge:

> All top-10 teams used a 2D/3D U-Net variant with at most minor modifications. [2]

6. Contribution

All code is available from https://github.com/ivalencius/COVID-19-Lung-CT-Lesion-Segmentation-Challenge

- Introduction - Cole Gvozdas
- Related Work - William Gibbons
- Methods - Everyone
- Experiments - Ilan Valencius
- Conclusion - Ilan Valencius

References

[1] Seyed-Ahmad Ahmadi, Johann Frei, Gerome Vivar, Marianne Dieterich, and Valerie Kirsch. Ie-vnet: Deep learning-based segmentation of the inner ear's total fluid space. *Frontiers in Neurology*, 13, 2022. 1

[2] Carlos Tor Diez et al. Holger Roth, Ziyue Xu. Rapid artificial intelligence solutions in a pandemic - the covid-19-20 lung ct lesion segmentation challenge, 2021. 1, 2, 3

[3] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, dec 2020. 2

[4] Wenqi Li, Guotai Wang, Lucas Fidon, Sebastien Ourselin, M. Jorge Cardoso, and Tom Vercauteren. On the compactness, efficiency, and representation of 3d convolutional networks: Brain parcellation as a pretext task. pages 348–360, 2017. 2

[5] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016. 1

[6] MONAI. Project monai, 2020. 1

[7] MONAI. Transforms, 2020. 2

[8] MONAI. A u-net model for lung lesion segmentation from ct images, 2021. 1, 2

[9] Microsoft Research. High-resolution network: A universal neural architecture for visual recognition, 2020. 1

7. figures

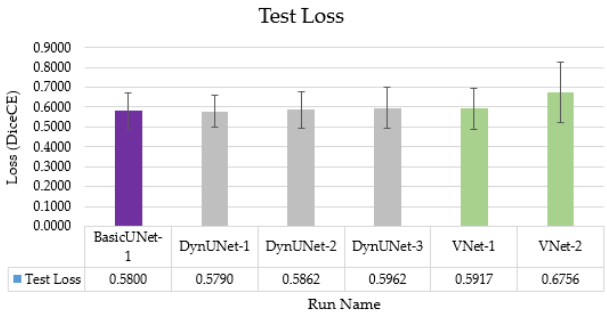

Figure 1. Test loss (DiceCE) determined on the unseen test dataset. Higher loss scores indicate better performance.

Figure 2. Example segmentation of lung lesions. The lesion area is outlined in green.

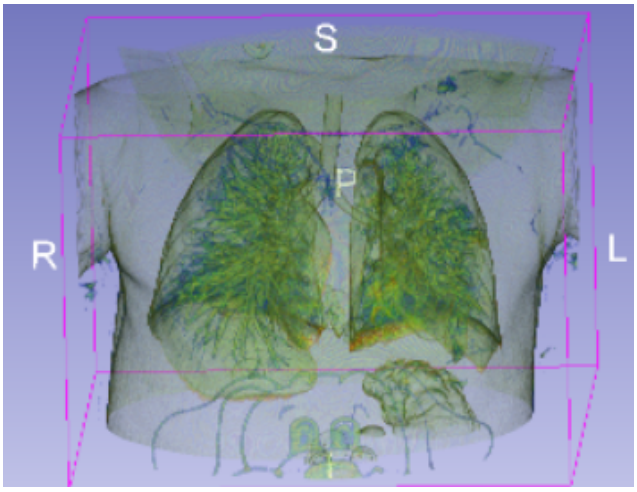

Figure 3. Reconstruction of lung tissue in Slicer.

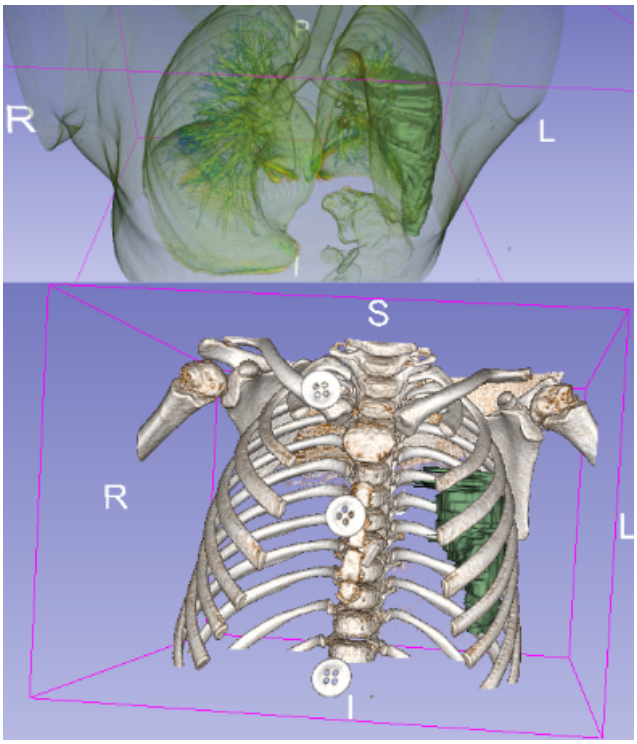

Figure 4. Reconstruction of lesion area after segmentation. The afflicted are is reconstructed in green.