COVID19 infection region segmentation

YingJhi Lin 411551007 and MuEn Huang 311551173

1 Introduction

The spread of the Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) which causes CoronaVirus Disease 2019 (COVID-19) has caused grave concerns for governments around the world. Apart from the number of people who have succumb to death due to the disease, the pandemic has results in great economic impact on societies around the world. In this sense, identifying the COVID-19 cases is crucial for contact tracing and curbing the outbreaks. There are many testing methods used to identify the cases, and among the widely used ones is Reverse Transcription-Polymerase chain reaction (RT-PCR). This method uses reverse transcription to obtain DNA and then applies Polymerase Chain Reaction (PCT) to amplify the DNA for analysis. This method is accurate, but costly, requires time and medical resources. Therefore, developing low-cost and rapid methods for identifying the infected cases is a matter of importance.

Computed Tomography scan(CT scan) is a medical imaging technique used to obtain detailed internal images of the body, which can provide clear and specific information while rapidly acquiring the image, in addition, it can image a small portion or all the body during the same examination. Therefore, CT image is now frequently used to identify disease or injury within various regions of the body.

Early in the epidemic, scientists discovered that analyzing CT images can be used for identifying COVID-19 cases. It was shown that bilateral pulmonary parenchymal ground glass and consolidative pulmonary opacities is observed in the chest CT images of covid-19 patients. The main objective of this project is to implement a Unet based model for automatically segmenting COVID-19 infection areas as well as lung from chest CT images, which plays a supportive role in the diagnosis of COVID-19. With the model's help, we can find evidence of COVID-19 and/or characterize its findings can play a crucial role in optimizing diagnosis and treatment, especially in areas with a shortage of expert radiologists, and on top of that, it can also help find out the symptoms of COVID-19, which is labor and time consuming.

2 Methods

In this project, we implemented the paper[1]. The authors proposed 2 novel convolution blocks including Feature Variation block (FV block) and Progressive Atrous Spatial Pyramid Pooling (PASPP). Based on these methods, we furthermore designed a deep supervision to calculate the loss at different output features of decoder. The proposed architecture is shown as Fig. 1.

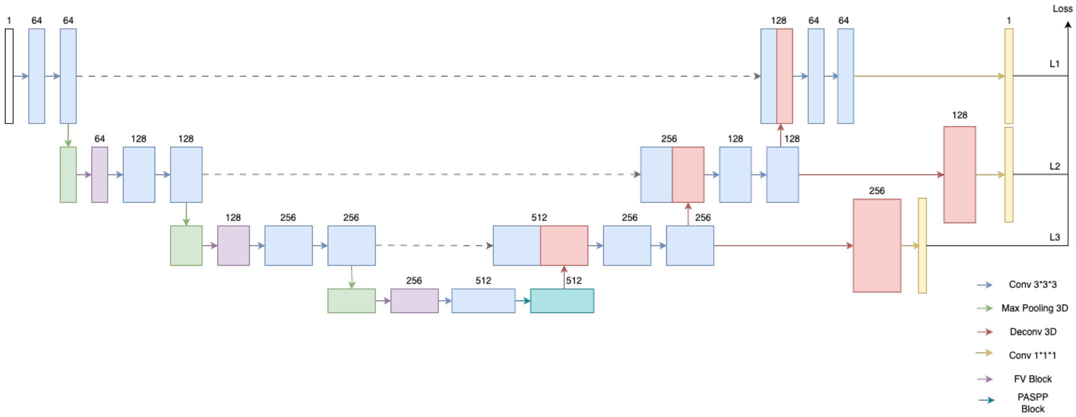

Fig. 1: Architecture of proposed model

2.1 FV block

The FV block we implemented is an attention block that included channel-wised and spatial-wised attention. The author claimed that this block can enhance the capability of feature representation effectively and adaptively for diverse cases.

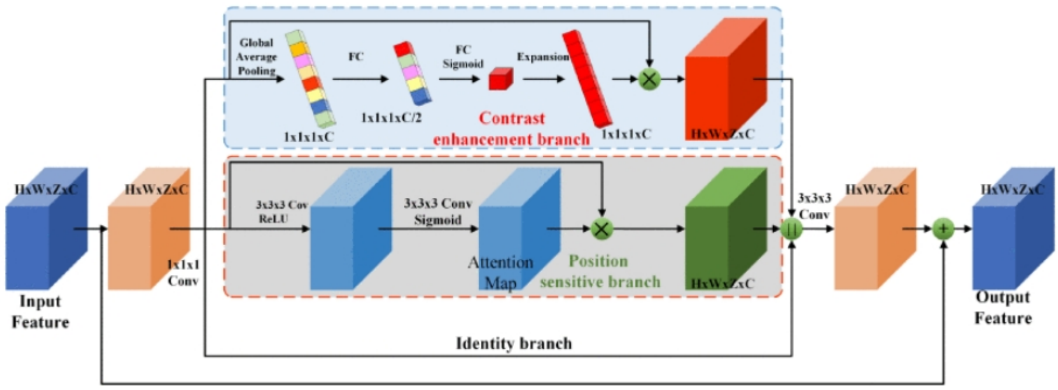

Fig. 2: Feature Variation block

2.2 PASPP

PASPP block is a refined ASPP block[2]. Although ASPP has been proposed to capture global information for semantic segmentation, the author claimed that aggregating information progressively is a more reasonable approach to get effective features. The PASPP block adopts atrous convolutions with different dilation rates to obtain features with various scales. The final output is generated straightforwardly to assemble residual branches in parallel.

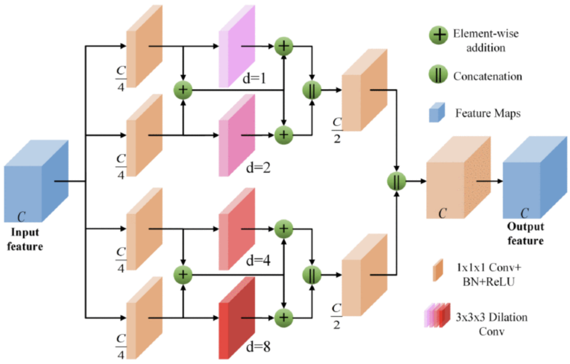

Fig. 3: PASPP

2.3 Deep supervision

Deep supervision (DS) is a method commonly used to avoid the problem of exploding or vanishing gradients in deep networks by forcing intermediate layers to produce more discriminative features[3]. Feature maps from intermediate decoding levels are first upsampled to the output dimension and then fed into a $1 \times 1 \times 1$ convolutional layer with sigmoid activation to produce prediction masks. From each output of the model, loss is computed with respect to the ground truth labeling.

3 Experiment and result

3.1 Training details

The setting of the hyperparameters is shown in table. We trained the model 200 epochs and choose the model with best validation dice score to inference in test set. The COVID19 dataset is downloaded from kaggle. A CT scans dataset with expert segmentations of infection regions in patients with COVID-19. Total 20 subjects of CT scans are divided into 14, 2, 4 as train, validation and test dataset respectively. The loss function is defined as

$$\mathbf{Loss} = \mathbf{L}_{BCE} * 0.5 + \mathbf{L}_{DSC}$$

which is 0.5 multiply binary cross entropy loss adding dice score loss.

Table 1: Setting of the hyperparameters

Epochs	Batch size	Optimizer	Learning rate	Patch size
200	5	Adam	0.0001	128*128*8

3.2 The details of the model

This is a UNet based model and we futher designed the model by inserting novel blocks from the paper. The FV block and PASPP block is used for attention mechanism and fusing features at different scales respectively, enabling model to handle the sophisticated infection areas with diverse appearance and shapes. Besides these novel blocks published in the paper, we further refine the model using deep supervision. In deep supervision, from each output of the model, loss is computed with respect to the ground truth labeling. The loss coefficients are weighted and summed to create the training loss of the framework. To reflect more emphasis on the final prediction thus assign 0.5 weight to the final layer and distribute the remaining 0.5 across intermediate outputs. The performance of adding these blocks of mechanism is shown in ablation study.

3.3 Experiment of FV block

The architecture of FV block is shown as Fig. 2. The contrast enhancement (CE) branch the author proposed is a refined block from SE block[4]. The difference between CE and SE is at the part of fully connected layer. We can see that in SE block, the feature is put into 2 fully connect layers with the shape 1*1*C to 1*1*C/r and 1*1*C/r to 1*1*C following a sigmoid function; in CE block, the feature is put into 2 fully connect layers with the shape 1*1*C to 1*1*C/r (r=2) and 1*1*C/r to 1*1*1 following a sigmoid function and the feature is expanded to the same dimension of the input. To further test which block performs better, we do a experiment and the result is shown in Table 2. The result shows different trends compared with original paper, which SE block performs a slightly better than CE block the paper proposed with higher dice score(72.29% vs 72.03%).

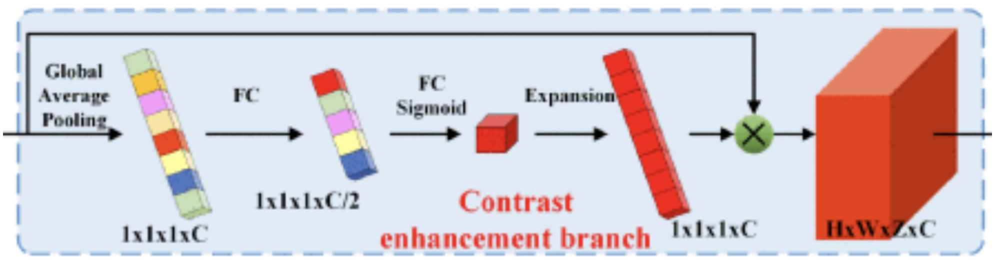

Fig. 4: Contrast enhancement branch

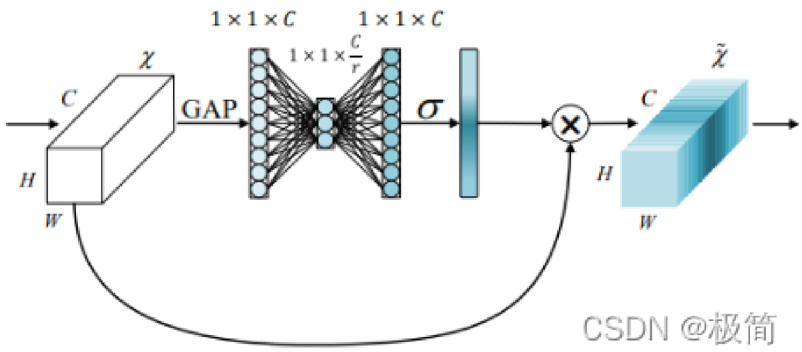

Fig. 5: Architecture of SE block

Table 2: Performance of the FV block With different channel-wise attention

Model	Dice Score	IoU	Sensitivity	Precision
UNet+FV (SE Block)	72.29%	57.31%	70.09%	77.27%
UNet+FV (CE Block)	72.03%	57.22%	70.02%	75.78%

3.4 Ablation study

To know how the novel blocks improve our model, we further do ablation study to observe each performance of model. The result showed that adding the FV block(with SE block) improved 0.3% of DSC ; adding the PASPP block further improved 2.4% of DSC. Finally, we do deep supervision and improved 1.5% of DSC. This shows that the architecture we proposed, the performance of DSC is the best and improved the original paper.

Table 3: Performance of the network with different blocks

Model	Dice Score	IoU	Sensitivity	Precision
UNet(Baseline)	71.96%	57.04%	64.80%	84.35%
UNet+FV	72.29%	57.31%	70.09%	77.27%
UNet+FV+PASPP	74.71%	60.55%	71.06%	79.83%
UNet+FV+PASPP+DS	76.27%	62.37%	69.38%	85.63%

3.5 Performance in test set

The CT data in the Kaggle dataset we utilized, existed resolution difference. Owing to the total 20 CT data are from 2 different resource. The resolution can be different from 512*512*300 to 630*630*40. To proof our model can inference in such a situation, we listed the total 4 CT data result in test set.

Table 4: Performance of the network with different blocks

Subject	Resolution	Dice Score	Sensitivity	IoU
A	(512, 512, 301)	83.04%	76.93%	71.00%
B	(512, 512, 256)	65.66%	57.22%	48.88%
C	(630, 401, 110)	70.74%	58.98%	54.72%
D	(630, 630, 93)	85.64%	84.36%	74.88%

3.6 Experiment of data preprocessing

The CT data in the Kaggle dataset offered lung mask and the ground truth of infection region mask. We compared the two data preprocessing scheme on one of which first mask the original CT image with lung mask, letting the model to learn the feature on lung region only. On the other hand, we do not do anything, and put the original data into model for training. We used the final proposed model to do the experiment, the result showed as below.

Table 5: Performance of the network with different data preprocessing

Data	Dice Score	IoU	Sensitivity	Precision
Lung only	76.27%	62.37%	69.38%	85.63%
Original image	50.74	39.55%	73.28%	52.47%

4 Discussion

In this project, we implemented a paper and experimented a series of novel techniques including FV block, PASPP block and deep supervision. To pursue a better performance, we also compared the FV block with CE block which is the paper proposed and the FV block with original SE block. We utilized both of the strength of binary cross entropy and dice similarity by combining them together and giving the dice similarity score a higher weight when designing the loss function. The result showed that the final model based on Unet with FV and PASPP block and training in deep supervision scheme performed the best. This showed that these attention mechanism and training techniques did help the model to learn the infection features. Within this model architecture we implemented and improved, the dice score of segmenting the infection region in our project is higher than the paper we referenced(76.27% vs 72.6%). However, our project still exists some limitations. First, we have got only 20 CT scans in kaggle dataset, which is not a large dataset, the training procedure and the result may affected by the data randomization without doing cross validation. Second, the 20 CT scans are from 2 different resources as the reslolution exists huge difference (i.e 512*512*300 and 630*630*40), although we train the CT data in patches, and show the final result of the 4 different subjects with satisfied result, the difference of resolution may affect in training with a larger dataset. Last, within a small training dataset, the generalization of the model should be further experimented.

5 Conclusion

In this project, we implemented a model to segment the COVID-19 infection region and lung from CT image. The model included a feature variation block and progressive ASPP block which are beneficial to highlight the boundary and position of COVID-19. In FV block, the SE block performs higher dice score and sensitivity than CE block. With the final model we propose, we are able to segment the COVID-19 infection area accurately with 76.27% of dice score from CT image.

References

[1] Yan, Q., Wang, B., Gong, D., Luo, C., Zhao, W., Shen, J., Ai, J., Shi, Q., Zhang, Y., Jin, S., Zhang, L., You, Z.: Covid-19 chest ct image segmentation network by multi-scale fusion and enhancement operations. IEEE Transactions on Big Data **7**(1), 13–24 (2021) https://doi.org/10.1109/TBDATA.2021.3056564

[2] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)

[3] Hilbert, A., Madai, V.I., Akay, E.M., Aydin, O.U., Behland, J., Sobesky, J., Galinovic, I., Khalil, A.A., Taha, A.A., Wuerfel, J., *et al.*: Brave-net: fully automated arterial brain vessel segmentation in patients with cerebrovascular disease. Frontiers in artificial intelligence **3**, 552258 (2020)

[4] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)