A Plug-and-play Attention Module for CT-Based COVID-19 Segmentation

Zhengqian Zhang\textsuperscript{1,a}, Haoqian Xue\textsuperscript{1,b}, Guanglu Zhou\textsuperscript{1,c,*}

\textsuperscript{1}Harbin Institute of Technology Computer College, Weihai, Shandong, China
\textsuperscript{a}20S130445@stu.hit.edu.cn, \textsuperscript{b}hxqashh@163.com, \textsuperscript{c}vip633@163.com
\textsuperscript{*}Corresponding author’s e-mail: vip633@163.com

Abstract. At the end of 2019, a new type of coronavirus (COVID-19) rapidly spread globally, even if the penetration of vaccination is getting higher and higher, the emergence of viral variants has increased the number of new coronal pneumonia infections. The deep learning model can help doctors quickly and accurately divide the lesion zone. However, there are many problems in the segmentation of the slice from the CT slice, including the problem of uncertainty of the disease area, low accuracy. At the same time, the semantic segmentation model of the traditional CNN architecture has natural defects, and the sensing field restrictions result in constructing the relationship between pixels and pixels, and the context information is insufficient. In order to solve the above problems, we introduced a Transformer module. Visual Transformer has been proved to effectively improve the accuracy of the model. We have designed a plug-and-play spatial attention module, on the basis of attention, increased positional offset, effective aggregate advanced features, and improve the accuracy of existing models.

1. Introduction
At the end of 2019, a new type of coronavirus (COVID-19) began to spread throughout the world. According to the World Health Organization\cite{5}, as of August 16, 2021, a total of 207,173,086 confirmed diagnoses worldwide. The deaths exceeded 4,361,996. It has spread to more than 100 countries and regions around the world. It has caused severe disasters to people in all countries around the world. The new crown vaccine, which is based on high hopes, has been raging under the raging of various variant viruses, and the effectiveness of the vaccine has continued to decline. More and more evidence shows that human society will coexist with COVID-19 for a long time. Facing the continuous surge of confirmed cases, hospitals and medical staff are facing great pressure on diagnosis and treatment. In the diagnosis process, virus accounting testing is considered to be a more scientific criterion. X-rays and computed tomography (CT) also widely used in the diagnosis process. After analysing 1014 patients in Wuhan\cite{6}, China using chest CT, the accuracy can reach 0.68. In general, chest CT can play an important role in the early diagnosis of new coronary pneumonia and the direct treatment of the later stage.

Over time, the application of deep learning in the diagnosis of new coronary pneumonia has increased. For example, many scientific research institutions from Wuhan have developed a new coronary pneumonia detection neural network(COVNet)\cite{7}. After testing with a large amount of local Wuhan data, the experimental results show that the deep learning model can accurately predict Covid-19 and distinguish between community-acquired pneumonia (CAP) and other non-pneumonia abnormal CT scans. In the early stage of the epidemic, Fan DP and others proposed a new coronary pneumonia lung
infection segmentation network (Inf-Net)[8] based on implicit recursive reverse attention (RA) module and explicit edge attention guidance, which mainly uses a parallel partial decoder (PPD) that aggregates advanced features to combine contextual information. At the same time, he also proposed a semi-supervised segmentation system for data scarcity. Although more and more artificial intelligence systems are applied to front-line clinical work, most of them are classification tasks, that is, to distinguish whether they are infected with new coronary pneumonia. CT image segmentation for new coronary pneumonia is also very rare. Segmentation data can effectively provide doctors with important reference information, but due to the scarcity of data sets and the difficulty of making segmentation data sets, as well as the weak differences in the characteristics of CT image classes. As a result, CT affects the low accuracy of segmentation, and there are few related researches.

In order to solve the above problems, we propose a plug-and-play spatial attention module that adds an appropriate positional offset in the traditional non-local network, which can effectively aggregate advanced features, and improve the accuracy of the split model, help the medical staff effectively assess the infection area of pneumonia in the CT image, facilitating medical staff for diagnosis.

2. Related Work
In this part, we will give a brief introduction to the recent new crown segmentation field and the plug-and-play transformer field.

Plug-and-play transformer:
The plug-and-play visual transformer can significantly improve the performance of existing visual models. More and more papers prove this, and there is a lot of related work in a hurry. Nanjing University Zhou Q et al. proposed an efficient multi-head self-attention module [10] to solve the problem of huge training and inference overhead and too low channel dimension in the subset, so that the dot product of query and key can no longer constitute an information matching function. These two questions. Stand-Alone Self-Attention [11] is proposed to replace spatial convolution to improve performance and reduce the amount of calculation. Gu R et al. proposed a joint spatial attention module [12], a new channel attention module, and a scale attention module, which are used to focus on the foreground area and highlight the most relevant feature channels, Emphasize the most significant feature maps in multiple scales.

Segmentation in Covid-19 Chest CT:
The segmentation of new coronary CT on chest has great medical diagnostic aid value, but there are few related papers. Xie W et al. proposed an RTSU network [13] to utilize structured relationships by introducing a new non-local neural network module. The proposed module learns the visual and geometric relationships between all convolutional features to generate self-attention weights. Zhou L et al. proposed a new model that can segment and quantify infected areas on CT scans from different sources[14]. Mainly through two innovative methods of the new coronary pneumonia CT scan simulator and the decomposition of the three-dimensional segmentation problem into three 2D problems. Tilborghs S et al. compared 12 deep learning algorithms [15]. Combining different methods can improve the overall test set performance of lung segmentation, binary lesion segmentation and multiple lesion segmentation.
3. Materials and Methods

Figure 1. An overview of the network

Figure 2. On the left is a non-local block[3]. On the right is our proposed spatial attention module, "⊕" the regular addition operation. "⊗" Stands for denotes element-wise sum. Red represents the position shift. Earthy yellow represents 1×1 convolution.

The dataset uses 100 new coronary pneumonia segmentation data sets provided by [8], 70 images are used for training, and 30 pictures are used to verify. All models use an SGD optimizer with an input resolution of 128 × 128, the learning rate is 0.01, and the training is 40000 iterations. Other parameters are unified based on MMSegmentation[9].

Figure 1 shows our model architecture. The feature map output by backbone is used as input. After the spatial attention module, the output result of the same dimension is obtained. Finally, it is fused with the initial feature map to obtain the final result.

Figure 2 and Equation (1) shows the left side shows the traditional attention mechanism in non-local, but the high-dimensional matrix multiplication leads to the complexity of the algorithm $O(hw \times hw)$, which is $O(n^2)$. And in the segmentation task, especially the new coronary pneumonia segmentation task, capturing the spatial structure is very important.

$$y = softmax(qk)v$$

In response to the above two problems, first effectively reduce the computational complexity and change the algorithm complexity from $O(n^2)$ to $O(n)$. Then we propose to add a position offset. The
position offset strengthens the ability of the attention mechanism to build the positional relationship between pixels and capture pixels Affinity. The Equation (2) shows process.

$$y = \text{softmax} \left( q R^v \left( k v + R^v k v^v + \right) + q \right)$$  \hspace{1cm} (2)

4. Experiments and Results

Table 1. Segmentation results on Covid-19 Chest CT testing set.”-s” represents the use of the spatial attention module.

| Method      | Mean IoU % | Mean Dice % |
|-------------|------------|-------------|
| Unet-s [2]  | 74.00 +2.36| 83.85 +2.13|
| FCN-s [1]   | 71.86 +3.40| 82.25 +3.28|
| PSPNet-s [4]| 70.42 +1.25| 80.78 +1.43|

Figure 3. Segmentation results on Covid-19 Chest CT testing set.

After adding the spatial attention module to the mainstream segmentation model, using the new coronary pneumonia segmentation data set as a benchmark, the results show that both mIoU and mDice indicators have improved by 1%-3%, and the segmentation effect is significant. It is proved that in the
CT-based Covid-19 segmentation task, using the spatial attention mechanism of this paper to build long-distance dependence between pixels has a good effect.

5. Conclusion
In order to solve the above problems, we propose a plug-and-play spatial attention module, adding an appropriate position offset to the traditional non-local network, which can effectively aggregate advanced features and improve the accuracy of the segmentation model, help medical staff to effectively evaluate the area of infection of new coronary pneumonia in CT images, which is conducive to medical staff in diagnosis and treatment.

References
[1] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.
[2] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.
[3] Wang X, Girshick R, Gupta A, et al. Non-local Neural Networks[J]. IEEE Computer Society, 2016.
[4] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[C]// IEEE Computer Society. IEEE Computer Society, 2016.
[5] World Healthy Organization Coronavirus Dashboard (2021). https://covid19.who.int/table.
[6] T. Ai, Z. Yang et al., “Correlation of chest CT and rt-pcr testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases,” Radiology, vol. 2019, p. 200642, Feb 2020.
[7] Li L, Qin L, Xu Z, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT[J]. Radiology, 2020:200905.
[8] Fan D P, Zhou T, Ji G P, et al. Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images[J]. IEEE Transactions on Medical Imaging, 2020, PP(99):1-1.
[9] MMSegmentation Contributors.(2020) OpenMMLab Semantic Segmentation Toolbox and Benchmark. https://github.com/open-mmlab/mmsegmentation
[10] Zhang Q, Yang Y. ResT: An Efficient Transformer for Visual Recognition[J]. arXiv preprint arXiv:2105.13677, 2021.
[11] Ramachandran P, N Parmar, Vaswani A, et al. Stand-Alone Self-Attention in Vision Models[J]. 2019.
[12] Gu R, Wang G, Song T, et al. CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation[J]. 2020.
[13] Xie W, Jacobs C, Charbonnier J P, et al. Relational Modeling for Robust and Efficient Pulmonary Lobe Segmentation in CT Scans[J]. IEEE Transactions on Medical Imaging, 2020.
[14] Zhou L, Li Z, Zhou J, et al. A Rapid, Accurate and Machine-Agnostic Segmentation and Quantification Method for CT-Based COVID-19 Diagnosis[J]. IEEE Transactions on Medical Imaging, 2020, PP(99):1-1.
[15] Tilborghs S, Dirks I, Fidon L, et al. Comparative study of deep learning methods for the automatic segmentation of lung, lesion and lesion type in CT scans of COVID-19 patients[J]. 2020.