Attention Gated Recurrent U-Net for CT Image Segmentation of COVID-19 Lung Infection Region

Haoyu Chen
Dept. Of Artificial Intelligence
Chonnam National University
South Korea
leochy554@gmail.com

Kyungbaek Kim
Dept. Of Artificial Intelligence
Chonnam National University
South Korea
kyungbaekkim@jnu.ac.kr

ABSTRACT

The corona virus disease 2019 (COVID-19) outbreak peaked in early 2020, posing a significant threat to human health. Computed tomography (CT) images are one of the most important screening tools for the diagnosis of COVID-19 today. CT images not only have the advantage of convenience and speed, but also can show the characteristics of lung lesions in COVID-19 patients. However, due to the strong noise of lung CT images and the small area of COVID-19 infection in the early stage, the traditional image segmentation method is prone to misdiagnosis. In recent years, with the development of deep learning technology, image segmentation can provide a reliable diagnostic basis for practical clinical medical applications, and improve the efficiency of doctors and the accuracy of disease diagnosis. Therefore, this paper proposes an Attention Gates Recurrent U-Net (AGRU-Net) scheme for CT image segmentation of COVID-19 lung infection region. The scheme increases the depth of the network by using the recurrent block in the convolutional layer of U-Net, and secondly introduces the attention gates mechanism into the U-shaped network to improve the multi-scale generalization ability of the U-Net model, so as to enhance the response and sensitivity of the model to the features of COVID-19 lung region, and the loss function of the model adopts weighted binary cross-entropy dice loss function to solve the segmentation task in highly unbalanced scenes and to improve the extraction accuracy of COVID-19 lung infection regions.In the experiments on the public dataset of COVID-19 lung CT scans, the dice coefficient of the AGRU-Net model proposed in this paper is 0.8979 in the infected region of COVID-19 lung, which is better than the segmentation effect of similar models and achieves effective segmentation of the CT images of the infected region of COVID-19 lung.

KEYWORDS

COVID-19, Image segmentation, Computed tomography, Attention Gates, U-Net, AGRU-Net, Dice coefficient

1 INTRODUCTION

Since December 2019, a number of cases of unexplained pneumonia have been detected in some hospitals in Wuhan, Hubei Province, which were confirmed to be acute respiratory infections with a novel 2019 coronavirus infection and subsequently began to spread rapidly in the surrounding countries [1]. Rapid and accurate diagnosis of COVID-19 is very important for the prevention and control of the epidemic, and a large amount of clinical experience has shown that pulmonary CT images play a very critical role in the diagnosis and disease assessment of COVID-19 and are an important part of the diagnostic criteria and discharge criteria of patients.In the CT imaging of the lungs of COVID-19, multiple small patchy shadows and shadows were seen in the early stage, which were obvious in the extrapulmonary area, and then developed into multiple ground-glass shadows in both lungs. In severe cases, solid lung changes may appear with a few pleural effusions [2-3]. With the recent advances in computer vision in the past decade, deep learning is increasingly being used for medical image analysis. Although the application of deep learning in computer vision has seen rapid growth in many different areas, it is still extremely challenging in the area of AI-based segmentation of medical CT images. For example, Zhou et al [4] proposed a Unet++ model to enhance image segmentation by nesting dense skip connections between the encoder and decoder to obtain more semantic information. Liu et al [5] transferred Mask R-CNN, which performs well on natural images, to CT images for lung nodule segmentation and trained Mask R-CNN to predict nodule location and nodule size.Saood et al [6] used two different deep learning techniques, SetNet and U-Net, for semantic segmentation of infected tissue regions in CT lung images.

Despite the great progress in COVID-19 CT image infected region segmentation studies, there are still many problems. First, most of the datasets used in the studies are now non-public, which leads to small training samples, prone to overfitting, poor generalization of the study results, and the resulting system cannot assist in clinical diagnosis. Second, COVID-19 lung CT images are unusually complex and easily confused with other lung-related diseases, and it is very difficult for the encoder to extract effective segmentation features. In addition, the infected region of new coronary pneumonia is characterized by indeterminate location, unclear boundary, and variable shape, which requires segmentation models with extremely strong detail feature extraction ability.

To solve the above problems, a segmentation scheme combining attention Gate and recurrent U-Net is proposed in this paper. Firstly, the network structure of U-Net is improved by

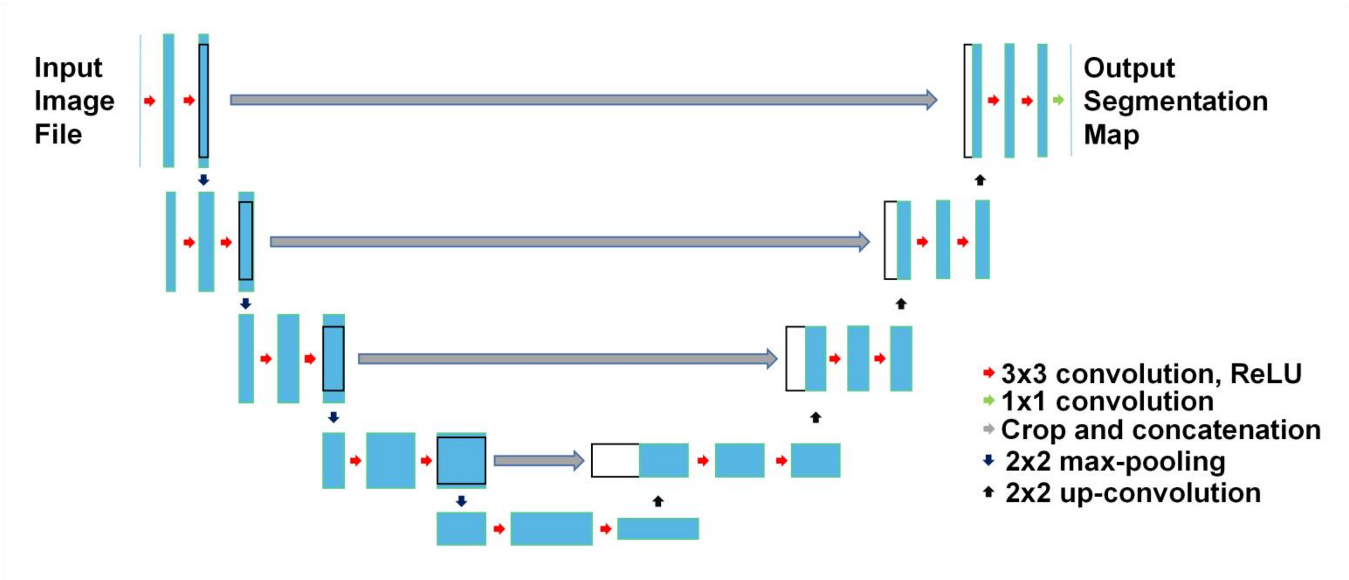

Figure 1: Traditional U-Net model framework.

using recurrent blocks in the convolutional layers of encoder and decoder in U-Net to increase the level of U-Net network, and the superior learning effect is obtained by stacking the convolutional blocks. Next, attention gates are introduced into the decoder of the U-Net, the output features of the encoder are readjusted to improve the multi-scale generalization ability of the network model, and the model loss function adopts a weighted binary cross-entropy dice loss function to solve the segmentation task in highly unbalanced scenes in order to strengthen the segmentation ability of the model. Finally, the automatic segmentation model AGRU-Net is proposed for the infected region of COVID-19 CT images.

2 RELATED WORK

2.1 U-Net Model

Traditional semantic segmentation algorithms are mainly composed of feature extraction as well as classifiers. Most of the commonly used methods are based on machine learning methods, such as support vector machines, decision trees, artificial neural networks, etc. These methods are shallow structured models, which can only use shallow features for judgment and cannot segment complex images well. Deep learning techniques use high-level semantic information for segmentation.In 2014, Long et al [7] first used a fully convolutional neural network (FCN) for end-to-end segmentation of natural images, achieving a breakthrough from traditional methods to deep learning methods.In 2015 Ronneberger et al [8] proposed a U-Net structure based on FCN.The U-Net network mainly consists of encoder, decoder and skip connections. The encoder performs the downsampling operation to extract the image spatial features, and the decoder performs the upsampling operation to generate images based on the spatial features extracted by the encoder. The encoder of U-Net uses two layers of 3×3 convolution, ReLu activation function and 2×2 maximum pooling to extract the image features, which not only reduces the feature map size but also increases the channels in the process. The decoder part uses 2×2 deconvolution for upsampling to reduce the number of channels and gradually recover the feature map size.The U-Net model is shown in Fig. 1.

The U-Net model has many advantages in the field of image segmentation, as it allows the use of both global location and contextual information, and requires few training samples to perform well in medical image segmentation. However, the COVID-19 CT lung infection region image is very complex, and U-Net uses simple convolution and pooling operations in the encoder to extract features. Such a feature extraction method tends to cause the model to fail to extract all the useful feature information, and a part of the features will be lost. In addition, U-Net uses simple convolution and deconvolution in the decoder for image recovery, which in turn leads to a certain loss of feature information, making the network unable to fully recover the complex feature information of the image.

2.2 Attention Mechanism

Attention is an indispensable human cognitive function, and an important property of perception is that humans do not tend to process all information at once, but tend to selectively focus on a part of information at the desired time and place while ignoring other perceptible information. Influenced by the attention mechanism, researchers gradually refer attention mechanism to the field of computer vision. John et al [9] first introduced the attention mechanism to the field of computer vision, arguing that the attention role is to optimize traditional visual search methods by reducing the number of samples processed and increasing the feature matching between samples. Wang et al [10] proposed a

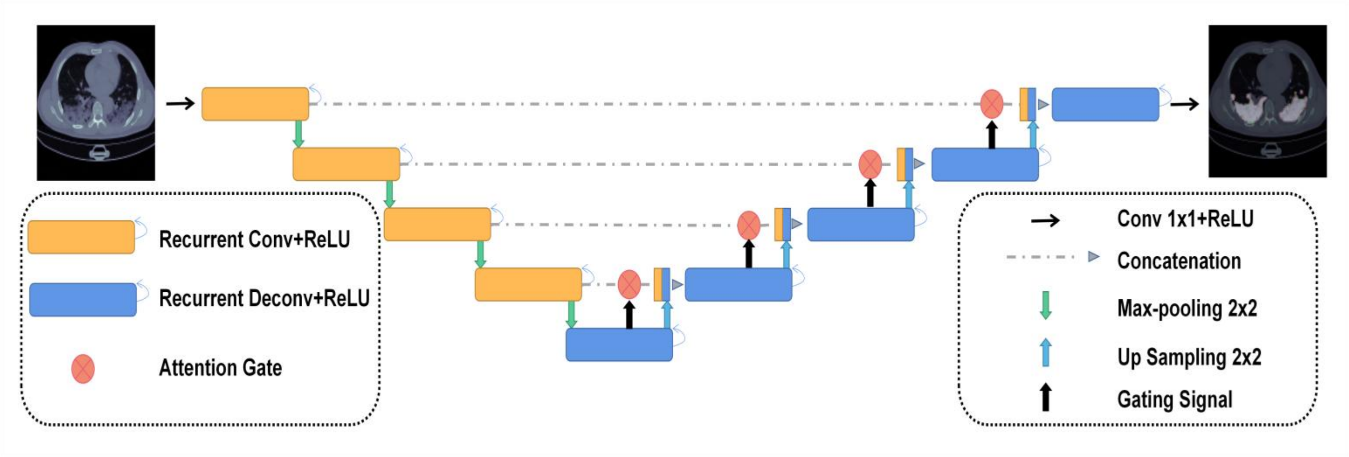

Figure 3: AGRU-Net model framework.

residual attention network to use the attention mechanism for more accurate dense prediction in semantic segmentation tasks. Zhang et al [11] proposed an edge-attention-guided network (ET-Net), which embeds an edge-attention representation to guide the segmentation network. By combining the attention mechanism with convolution,downsampling and upsampling, it can enhance the information region in the feature map and suppress irrelevant information, thus adding the feature extraction capability of the network.

2.3 Recurrent Neural Network

Recurrent neural network (RNN) are a class of artificial neural networks. RNN use back propagation to update the weights of neurons for better network learning. During forward propagation, the inputs move forward and through each layer to calculate the output state. In the case of backpropagation, go back and change the weights of the neurons to get a better way to learn. As deep learning continues to develop, RNN are gradually introduced into convolutional neural network (CNN) to form recurrent convolutional neural network (RCNN) to solve the image segmentation problem.Chen et al [12] proposed a deep learning framework based on a combination of FCN and RNN to solve the segmentation problem of 3D images in medical images.Vuola et al [13] developed a U-Net and Mask-RCNN ensemble framework with good performance in the segmentation task of cell nuclei. By adding a cyclic structure to the network, superior segmentation results can be obtained.

3 MATERIALS AND METHODS

3.1 Data and Preprocessing

Since most of the existing COVID-19 lung CT imaging studies are based on large and private datasets, and it is very difficult to obtain these data from a single institution, the COVID-19 lung CT scan dataset provided by Ma et al [14], which contains 20 CT scans of COVID-19 patients and expert segmentation of lungs and infections, was used in this experiment. The dataset includes a total of 3,520 images and corresponding ground truth with lung, right lung and infection tagged by two radiologists and validated by experienced radiologists. The dataset was divided into 3,205 images for training, and 315 images for validation. Considering that the original image size was too large compared to the COVID-19 lung infection region image, which might not be conducive to training, the original image containing the COVID-19 lung infection region was rescaled to 128 × 128 pixels size as the network input. All data were enhanced by random cropping, scaling and rotation, and pixel values in the range of 0 to 1 were normalized to improve the training accuracy before training. Fig. 2 shows some CT images of the dataset and their corresponding ground truths.

3.2 Neural Network Model

3.2.1 Improved U-Net model. In order to solve the problem of the general segmentation ability of U-Net model with fewer network parameters, inspired by the attention mechanism and recurrent convolutional neural network, we propose an improved U-Net network combining recurrent block and attention gate, which is named AGRU-Net, and the network framework is shown in Fig. 3.The AGRU-Net model retains the encoding and The

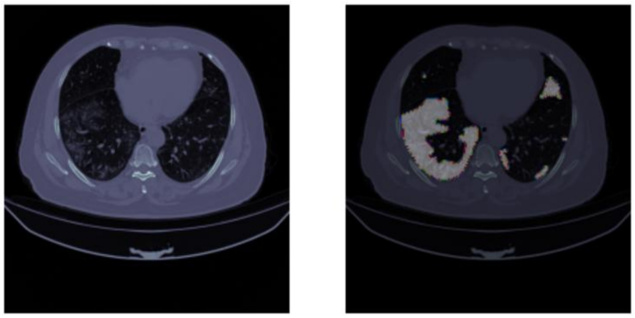

Figure 2: Lung CT images and infection masking in patients with COVID-19 infection. Images are from the COVID-19 lung CT scan public dataset [14] used in this article.

AGRU-Net model retains the encoding and decoding processes in U-Net, but differs from the U-Net network in that firstly, the AGRU-Net model uses a recurrent convolutional layer (RCL) [15] instead of a convolutional layer in the encoding and decoding processes, which helps increase the depth of the model and effectively preserves the features in the image. Secondly, attention gates are added in the decoding process to suppress feature responses that are not related to the background region, reducing the number of parameters and computational burden associated with the increase in network depth.

3.2.2 Recurrent module. The RCL unit used in this paper is gradually evolved over discrete time steps, assuming that the x_n input sample in the RCL unit is located at the n' layer, and for the pixel unit located on the kth feature map in the RCL unit (i, j), Its output $H_{ijk}(t)$ in step size t is as follows:

$$H^n_{ijk}(t) = (w^f_k)^T u^{f(i,j)}_n + (w^r_k)^T x^{r(i,j)}_n (t-1) + b_k \tag{1}$$

Here, $u^{f(i,j)}_n$ and $x^{r(i,j)}_n$ represent the input in the regular convolutional layer and the n'th RCL unit, respectively, w^f_k and w^r_k are the regular convolutional layer weights and the RCL weights of the n'th feature map, respectively, and b_k is the bias. The output of the RCL unit is activated by the RuLU function with the following formula:

$$f(H^n_{ijk}(t)) = Max(0, H^n_{ijk}(t)) \tag{2}$$

The expanded RCL structure is shown in Fig. 4. When t = 2 time steps, a feedforward sub-network with a maximum depth of 3 and a minimum depth of 1 is formed, consisting of a subsequence of one convolutional layer and two recurrent convolutional layers.

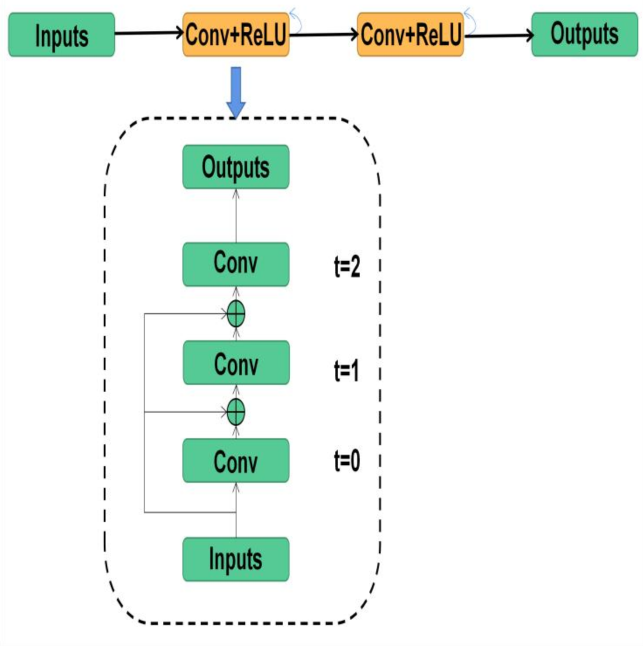

Figure 4: Expanded recurrent convolution structure for t = 2.

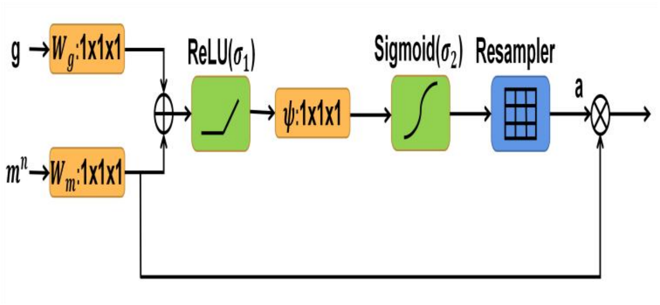

Figure 5: Attention Gate module framework.

3.2.3 Attention Gate module. The structure of the attention gate module used in this paper is shown in Fig. 5, Where n denotes the number of layers, i denotes the spatial pixel, m is the feature map of the encoding process extracted by the attention gate, g denotes the gating vector, W_m and W_g denotes the 1x1x1 convolution, ψ^T denotes the linear transform calculated for the channel 1x1x1 convolution, σ_1 denotes the ReLU activation function, σ_2 denotes the Sigmoid activation function, bg and bφ are the corresponding convolution bias terms, Θ_{att} are the network parameters, and the resampling operation resamples the feature map to the original size of m. a_i are the attention coefficients in the range of [0, 1]. a_i generally obtains larger values in the target region and smaller values in the background region and is the output of the attention gate model. The attention gate module equation is as follows:

$$q^n_{att} = \psi^T(\sigma_1(W^T_m m^n_i + W^T_g g_i + b_g)) + b_\psi \tag{3}$$

$$a^n_i = \sigma_2(q^n_{att}(m^n_i, g_i; \Theta_{att}) \tag{4}$$

The application of attention gates enables the model to automatically respond to feature areas and suppress irrelevant area responses without localizing the features, while simulating the global feature location relationship, so that similar features can enhance each other and thus improve the model accuracy.

3.3 Loss Function

In order to solve the category imbalance problem in medical images, this paper uses a weighted binary cross-entropy dice loss function, which is shown in the following equation:

$$L_M = L_{W-B} + L_D \tag{5}$$

Here, L_M is the weighted binary cross-entropy dice loss function, L_{W-B} is the weighted binary cross-entropy loss function, and L_D is the dice loss function.

The weighted binary cross-entropy loss function is a variant of the binary cross-entropy loss function that enhances the weights at the boundary of the segmented object, and its formula is shown below:

$$L_{W-B} = -w_i(\beta * Y_i \log(P_i) + (1-Y_i)\log(1-P_i)) \tag{6}$$

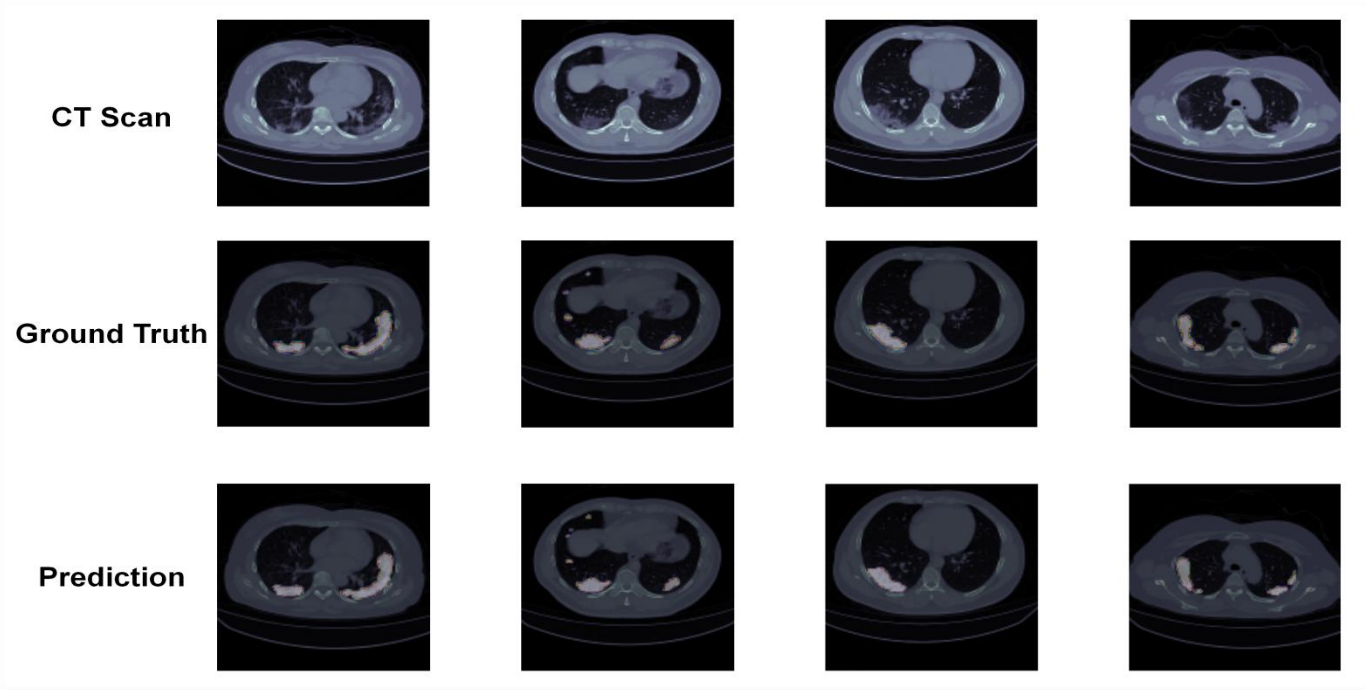

Figure 6: COVID-19 Lung Infection Region Segmentation Results.

Here, w_i denotes the weight value of each pixel point, P_i denotes the predicted value of each pixel, and Y_i denotes the label value of each pixel, and β value can be used to adjust the false positives and false negatives.

The Dice loss function is widely used in segmentation and classification of medical images. The Dice loss function solves the problem of positive and negative sample imbalance by comparing the similarity between the predicted and labeled values, and its formula is shown as follows:

$$L_D = 1 - \frac{2 * w_i * P_i * Y_i}{w_i * P_i + w_i * Y_i} \tag{7}$$

In this paper, the segmentation details of the model are optimized by combining the advantages of the weighted binary cross-entropy loss function and the dice loss function to improve the segmentation accuracy of images.

4 Results and Evaluation

4.1 Implementation

The network framework is based on the public framework tensorflow 2.9.1, trained and tested on a computer with an Intel Core i7-10700 2.9 GHz CPU (with 64 GB of memory) and an NVIDIA GeForce RTX 2060 SUPER GPU (with 8 GB of video memory), on a CUDA 11.7 architecture platform The parallel computation is performed on a CUDA 11.7 architecture platform and CuDNN 8.4.1 is invoked for computational acceleration. The proposed network uses the Adam[16] optimization algorithm with default hyperparameters, where the initial learning rate is set to 10^{-4},and the detailed parameters of the experiment are shown in Table 1.

Table 1: Experimental operation parameters.

Name	Parameter configuration
Network framework	Tensorflow2.9.1
Epoch	80
Batch size	12
Learning rate	0.0001
Optimizer	Adam

4.2 Performance Indicators

The experimental results are quantitatively analyzed by considering several performance metrics, including accuracy (AC), Dice coefficient (DC). For this purpose, four variables are used in this paper: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

The accuracy is used to evaluate the precision of pixel classification and is obtained from the following equation:

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

The Dice coefficient is an ensemble similarity measure function that is used to calculate the similarity of two samples in an experiment. The calculation formula is as follows:

$$DC = 2 * \frac{|GT \cup SR|}{|GT| + |SR|} \tag{8}$$

Here, GT is the label of the segmented image and SR is the network segmentation result.

4.3 Experimental results and analysis

In order to demonstrate that the AGRU-Net network model has superior results for COVID-19 lung infection region segmentation, three deep learning-based models, U-Net, Attention U-Net [17], and Residual U-Net [18], were used in the COVID-19 CT scan dataset for COVID-19 lung infection region segmentation process. The results are shown in Table 2, the Dice coefficient and accuracy of the AGRU-Net model were as high as 0.8979 and 0.9956, respectively, and the obtained results were higher than the other schemes. Fig. 6 shows the segmentation results of ARGU-Net on the COVID-19 lung CT scan dataset. The first row is the CT scan image of COVID-19 lung, the second row is the ground truth mask, and the third column is the segmented image of the infected region of COVID-19 lung from ARGU-Net. It can be seen that the segmentation results obtained by the AGRU-Net model are already very close to the expert labeling results, reflecting the strong segmentation capability of the AGRU-Net model for the COVID-19 lung infection region.

Table 2: Performance comparison of ARU-Net with other networks.

Model	DC	AC
U-Net	0.8878	0.9955
Attention U-Net	0.8895	0.9955
Residual U-Net	0.8882	0.9954
AGRU-Net	**0.8979**	**0.9956**

4 CONCLUSIONS

In this paper, we develop and analyze a method to segment the COVID-19 lung infection region from CT images. Based on the U-Net network structure, we incorporate the recurrent block and attention gate mechanism to enhance the localization ability and feature extraction ability of the network for the target region, improve the overall robustness of the network with the increase of the number of network layers, and enhance the segmentation ability of the network. In addition, the loss function training model combining the weighted binary cross-entropy loss function and the dice loss function can effectively alleviate the model instability problem caused by sample imbalance in COVID-19 CT images, and improve the model segmentation prediction ability. Ultimately, we outperform the currently popular semantic segmentation methods in terms of COVID-19 lung infection region. With the increasing status of deep learning in medical image segmentation, we will make more changes to the model to improve its performance in the future.

ACKNOWLEDGMENTS

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF)& funded by the Korean government (MSIT) (NRF-2019M3E5D1A02067961). This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program(IITP-2022-RS-2022-00156287) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)

REFERENCES

[1] Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. The lancet, 395(10223), 470-473.

[2] Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O., Gu, J., ... & Zheng, C. (2020). Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. The Lancet infectious diseases, 20(4), 425-434.

[3] Ye, Z., Zhang, Y., Wang, Y., Huang, Z., & Song, B. (2020). Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. European radiology, 30(8), 4381-4389.

[4] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 3-11). Springer, Cham.

[5] Liu, M., Dong, J., Dong, X., Yu, H., & Qi, L. (2018, September). Segmentation of lung nodule in CT images based on mask R-CNN. In 2018 9th International Conference on Awareness Science and Technology (iCAST) (pp. 1-6). IEEE.

[6] Saood, A., & Hatem, I. (2021). COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. BMC Medical Imaging, 21(1), 1-10.

[7] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[8] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[9] Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. Artificial intelligence, 78(1-2), 507-545.

[10] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., ... & Tang, X. (2017). Residual attention network for image classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3156-3164).

[11] Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., & Shao, L. (2019, October). Et-net: A generic edge-attention guidance network for medical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 442-450). Springer, Cham.

[12] Chen, J., Yang, L., Zhang, Y., Alber, M., & Chen, D. Z. (2016). Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. Advances in neural information processing systems, 29.

[13] Vuola, A. O., Akram, S. U., & Kannala, J. (2019, April). Mask-RCNN and U-net ensembled for nuclei segmentation. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (pp. 208-212). IEEE.

[14] Jun, M., Cheng, G., Yixin, W., Xingle, A., Jiantao, G., Ziqi, Y., et al., 2021. COVID-19 CT lung and infection segmentation dataset (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.3757476.

[15] Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3367-3375).

[16] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[17] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[18] Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters, 15(5), 749-753.