Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
TDD-UNet: Transformer with double decoder UNet for COVID-19 lesions segmentation

Xuping Huang a,1, Junxi Chen b,1, Mingzhi Chen c, Lingna Chen a,∗, Yaping Wan a,∗

a Computer School, University of South China, Hengyang 421001, China
b Affiliated Nanhua Hospital, University of South China, Hengyang 421001, China
c College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China

ARTICLE INFO

Keywords:
COVID-19
UNet
Transformer
Dual decoding module
Self-attention

ABSTRACT

The outbreak of new coronary pneumonia has brought severe health risks to the world. Detection of COVID-19 based on the UNet network has attracted widespread attention in medical image segmentation. However, the traditional UNet model is challenging to capture the long-range dependence of the image due to the limitations of the convolution kernel with a fixed receptive field. The Transformer Encoder overcomes the long-range dependence problem. However, the Transformer-based segmentation approach cannot effectively capture the fine-grained details. We propose a transformer with a double decoder UNet for COVID-19 lesions segmentation to address this challenge, TDD-UNet. We introduce the multi-head self-attention of the Transformer to the UNet encoding layer to extract global context information. The dual decoder structure is used to improve the result of foreground segmentation by predicting the background and applying deep supervision. We performed quantitative analysis and comparison for our proposed method on four public datasets with different modalities, including CT and CXR, to demonstrate its effectiveness and generality in segmenting COVID-19 lesions. We also performed ablation studies on the COVID-19-CT-505 dataset to verify the effectiveness of the key components of our proposed model. The proposed TDD-UNet also achieves higher Dice and Jaccard mean scores and the lowest standard deviation compared to competitors. Our proposed method achieves better segmentation results than other state-of-the-art methods.

1. Introduction

COVID-19 has seriously disrupted people’s everyday life because of its high infectivity and irreversible damage to the human body. The diagnostic tests of COVID-19 mainly rely on the reverse transcription-polymerase chain reaction (RT-PCR) test for collected samples of nose and throat swabs. However, the RT-PCR diagnostic test occurs to a high false-negative rate (FNR) [1], and the test results are not immediately available. Chest computed tomography (CT) has a potential role and high sensitivity in the diagnosis of COVID-19 [2], and Chest X-ray is sensitive and moderately specific for the diagnosis of COVID-19 [3]. Chest CT has more utility in excluding COVID-19 than for differentiating other lung diseases. Chest CT scanning provides more detailed information than plain chest X-rays. CT scans and X-rays require experienced radiologists to analyze the slices one by one, which is time-consuming. Chest CT scans for COVID-19 patients often show bilateral patchy shadows or multiple areas of ground-glass opacity (GGO) [4], which are irregular and very small in the lungs. These cause difficulties in obtaining accurate segmentation results, as shown in Fig. 1. Therefore, deep learning improves automatic image segmentation of lesions in chest CT or X-ray images of COVID-19 patients [5].

For previous medical image segmentation, non-uniform gray distribution and blurred edges result in biases. Image segmentation lacks a large amount of labeled image data, which causes it to overfit easily for models with many parameters. Convolutional Neural Networks (CNNs) [6] had achieved excellent results in image segmentation. CNNs learn local features automatically from medical images. A fully convolutional neural network (FCN) [7] solves the semantic segmentation problem. FCN accepts any size of the input image and uses the deconvolution layer to recover the feature map resolution similar to the size of the input image. Dilated convolution [8] is used to design an Atrous Spatial Pyramid Pooling (ASPP) module, which solves the problem of segmentation scale. The pyramid pooling module in PSPNet [9] aggregates contextual information from different regions to obtain global information. UNet [10] focuses on fusing the global

* Corresponding author.
E-mail addresses: linda_cjx@163.com (L. Chen), 512829758@qq.com (Y. Wan).
1 Xuping Huang and Junxi Chen contributed equally to this study.

https://doi.org/10.1016/j.compbiomed.2022.106306
Received 22 May 2022; Received in revised form 22 October 2022; Accepted 6 November 2022
Available online 8 November 2022
0010-4825/© 2022 Elsevier Ltd. All rights reserved.
context information with local features at different scales. Therefore, UNet shows excellent potential in medical image segmentation.

UNet [10] network uses a unique encoding and decoding layer to fuse the high-level features with the low-level features of images. Many variants of UNet have emerged in recent years. Attention UNet [11] filters unwanted feature information by adding an attention gate at the skip connection. Unet++ [12] further integrates multi-level feature maps based on UNet to obtain more detailed image information by adding skip connection paths. However, the above approaches have two main limitations. Firstly, due to the locality of convolution operation, UNet uses convolution to extract features, making it challenging to capture global information. Secondly, the down-sampling process of the encoder layers causes information loss, which leads to the semantic gap between the feature maps of the encoder layers and decoder layers. Therefore, it is worth studying how to fuse feature maps more effectively.

Recently, transformers have been widely used in computer vision. Transformer [13] was initially applied to natural language processing (NLP) and mainly aimed at sequence data. Due to the self-attention mechanism of transformers, the training model fully captures the relationship between each data sequence. Transformers overcome the deficiency of traditional convolution operation in image segmentation. However, transformers are very computation-intensive by calculating the relationship between each sequence of data. Most Vision Transformers (ViT) [14] in image segmentation accurately captured position relationships when pre-trained on large datasets. To combine the advantages of convolution operation and transformers, we propose a transformer with a double decoder UNet to segment COVID-19 lesions. In summary, the main contributions of this method include the following:

1. We combine the encoder part of UNet with the multi-head self-attention (MHSA) in Transformer to obtain a larger receptive field and more feature information.
2. We adopt a dual decoder structure to promote the result of foreground segmentation by predicting foreground and background and applying deep supervision.
3. We evaluate the performance of our proposed method on four different datasets for the COVID-19 lesions segmentation task.

2. Related work

For medical image segmentation, many researchers proposed the encoder-decoder structure based on end-to-end architecture, such as FCN [7], UNet [10], DeepLab V3 [15] and Unet++ [12]. These approaches have been successful in many medical image analysis fields [16–18]. Recently, there have been a lot of improvements to the encoder-decoder structure to achieve better segmentation performance. For example, Gu et al. proposed CE-Net [19] as an extension of GoogleNet [20], which demonstrated good performance for organ segmentation tasks. Likewise, Qiu et al. proposed a lightweight model, MiniSeg [21], based on the Attentive Hierarchical Spatial Pyramid module, for COVID-19 lesions segmentation, which was inspired by the SpPnet [22]. While all these methods achieve good results in segmentation tasks, they are still limited to the locality of CNNs, which do not extract global features very well. The Transformer [13] compensates for this limitation by capturing long-range dependencies.

Recently, ViT [14] outperformed state-of-the-art CNNs on ImageNet classification. Medical image segmentation based on Transformers has been developed [23–26]. TransUNet [27] is the first Transformer-based framework for medical image segmentation. Valanarasu et al. proposed a gated axial-attention model (MedT [18]) for medical image segmentation to treat the low number of data samples. Also, an improved method of modeling global contexts was developed by Zhang et al. by combining CNNs and Transformers into a two-branch architecture called TransFuse [28]. However, the above methods mainly focus on the defects of convolution operation rather than the optimization design of the UNet structure. Despite Transformer’s strong capabilities in a global modeling context, fine-grained details of medical images are still difficult to capture.

Previously mentioned methods (i.e., UNet, Unet++, and MiniSeg) were also used for COVID-19 lesions segmentation [29–32], but most methods did not perform well. To enhance the learnability of the model, Fan et al. [33] proposed a semi-supervised segmentation scheme based on random selection propagation. Yang et al. [34] proposed a semi-supervised federal method for COVID-19 lesions segmentation to deal with differences in cross-center imaging equipment and data distribution. Similarly, Wang et al. [35] proposed a multi-task deep learning method, DeepSC-COVID, which simultaneously learns lesions segmentation and classification for 3D COVID-19. Yu et al. [36] utilized a novel way to capture complementary information by combining vertical and horizontal kernels in two distinct branches.

Although these methods achieved good results, they may cause structural redundancy. In this paper, we combined the self-attention mechanism in Transformer with UNet to solve the intrinsic locality of convolution. We also propose a novel double-branch decoder integrated with additional deep supervision [37] to improve the network’s ability to perceive details via foreground and background segmentation.

3. Method

3.1. Overview

In this research, we present a transformer with a double decoder UNet to segment the infected area of COVID-19, and our proposed method is called “TDD-UNet”. The network is utilized to improve the defect of convolutional locality and corrects the bias of traditional single-branch segmentation networks by introducing a background segmentation branch. Fig. 2 presents the architecture of the proposed “TDD-UNet”, which is mainly based on the UNet’s encoder-decoder architecture and integrates a multi-head self-attention mechanism, double-branch decoder, and deep supervision. Similar to UNet, the encoder of TDD-UNet is used to obtain the feature representations. Then, the feature representations from the last encoding layer are input into the multi-head self-attention, where they will be re-weighted pixel-wise and get our interested informative presentations. They are projected to the label space by our double-branch decoder. Finally, we obtain the segmentation result. The novel double-branch decoder module aims to effectively fuse the background and foreground features maps and suppress the existing bias predictions in the traditional single-branch decoder. Moreover, it can learn more feature information and semantic details of the image in the decoding layers; then refine the lesion’s boundary to yield more accurate segmentation results. In our proposed method, background and foreground branches have the same feature map as their input. Although learning features from different directions, their learning results should be consistent with their corresponding ground truth. Deep supervision is employed for the outputs.
Fig. 2. An overview of our proposed TDD-UNet, which is implemented based on a U-shape architecture and consists of three modules: MHSA, Double-Decoder, and Deep Supervision.

Fig. 3. The proposed double-branch decoder and calculation process of Multi-head self-attention.

from the two decoder branches to enhance this consistency further. The details of multi-head self-attention for the backbone, double-branch decoder and deep supervision are presented in the following subsections.

3.2. MHSA & Double-Decoder

MHSA focuses more on the long-range dependence on image patches to extract long-range structural information between patches. However, due to the intrinsic locality of convolution operation, the traditional UNet-like model generally demonstrates limitations in explicitly modeling long-range dependency. We introduce multi-head self-attention functions at the bottom of the UNet as described in [13], as shown in Fig. 3. We expect to connect every pixel in the feature map of the last encoding layer with each other, thus giving access to a wide receptive field. The prediction for one specific pixel can be influenced by any input pixel. The formula for self-attention is shown in Eq. (1):

\[
\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V
\]

As shown in Fig. 3, the input matrix \(a \in R(d, N)\) is multiplied by three different matrices \(W_q, W_k, W_v \in R(d, d)\) to obtain three intermediate matrices, a query matrix \(Q \in R(d, N)\), a key matrix \(K \in R(d, N)\), and a value matrix \(V \in R(d, N)\), which are the three inputs of the self-attention module. They have the same dimension. Take \(K\) transposed and multiply it by \(Q\) to get the attention matrix, representing the attention in pairs at each position. Then it takes a softmax operation to get \(\tilde{A} \in R(N, N)\), and finally, it is multiplied by \(V\) matrix to get the final output and combined with attention matrices calculated from other heads. In addition, a location embedding is added to the input feature to allow for context information. This is particularly relevant for medical image segmentation, where different anatomical structures follow fixed spatial positions. Thus, location embedding can be used to capture the absolute and relative positions between lesions.

Double-Branch Decoder is aimed to leverage the background-prediction task in improving the segmentation performance of low-contrast small lesions. It consists of two decoders corresponding to the background and foreground segmentation tasks, respectively. As shown in Fig. 3, the \(F_{in} \in R^{1024x32x32}\) denotes the output feature map of the last encoding layer, which has the biggest receptive field. To obtain the structural relationship between each pixel, we employ it as the input of the MHSA module to calculate the attention matrix. Finally, the final output \(F_{MHSA} \in R^{1024x32x32}\) of the MHSA is produced by the element-wise multiplication of \(\tilde{A}\) and \(V\) (see Fig. 3). Compared to the previous \(F_{in} \in R^{1024x32x32}\), the output feature map contains more accurate semantic information. Since we suppose to obtain two different
segmentation results with all the semantic information of the image in the decoding stage, we use two decoders to optimize for different target results. Similar to UNet, the decoder contains four consequent Conv3×3 and upsampling. We obtain the feature maps \( F_{\text{back}} \in \mathbb{R}^{1 \times 512 \times 512} \) and \( F_{\text{fore}} \in \mathbb{R}^{1 \times 512 \times 512} \) from the two decoders. The background foreground fusion strategy effectively combines full semantic information of images by merging feature maps from the two branches and improving the COVID-19 lesions segmentation performance. Specifically, the merging process of two feature maps is defined as follows:

\[
F_{\text{out}} = \frac{1 - \sigma (F_{\text{back}}) + \sigma (F_{\text{fore}})}{2}
\]

where \( \sigma \) denotes sigmoid function and \( F_{\text{out}} \in \mathbb{R}^{1 \times 512 \times 512} \) refers to the final segmentation result.

**Double-Branch Deep Supervision** is to learn hierarchical representations from the opposite layer aggregated feature maps. In our proposed method, considering that background and foreground branches have the same feature map as their input, although learning features from different directions, the learning results should be opposite. However, the multi-task learning model is difficult to train because tasks may have different learning curves. Our proposed method balances the training objectives by weighting the losses. Furthermore, we employ deep supervision on the two segmentation decoders by integrating the segmentation layers from the two branches to form the final network output, shown in Fig. 3. The calculation formula of loss \( L_{April} \), for deep supervision is as follows:

\[
L_{April} = aL_{\text{for}} + bL_{\text{bg}}
\]

where  \( L_{\text{for}} \) represents the loss of the feature map of the foreground segmentation branch (the \( F_{\text{fore}} \) in Fig. 3), and \( L_{\text{bg}} \) represents the loss of the feature map of the background segmentation branch (the \( F_{\text{back}} \) in Fig. 3). And \( a \) and \( b \) are the trade-off parameter for background and foreground loss. We set \( a = 0.7, b = 0.3 \) in this study experimentally.

### 3.3. Loss function

The proposedage UNet is a slice-level end-to-end network composed of background and foreground segmentation branches. Its loss function comprises two parts: deep supervision and final segmentation loss. The use of binary cross-entropy (BCE) loss [38] in the original UNet performs poorly on our dataset. Moreover, BCE loss is also unsuitable for our circumstance [39,40], in which the CT images of patients with COVID-19 are highly imbalanced data. The region of lesions is able for our circumstance [39,40], in which the CT images of patients with COVID-19 are highly imbalanced data. The region of lesions is

\[
L_{\text{BCE}} = -\frac{1}{N} \sum_{i} [y_{i} \log(p_{i}) + (1 - y_{i}) \log(1 - p_{i})]
\]

where \( y_{i} \) denotes the true label of the sample, \( p_{i} \) refers to the predicted label, and \( N \) is the total number of pixels in the image. Similarly, the Dice loss can be written as:

\[
L_{\text{Dice}} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} = 1 - \frac{\sum_{i} \hat{p}_{i} y_{i} + s}{\sum_{i} \hat{p}_{i} + \sum_{i} y_{i} + s}
\]

where \( X \) denotes the ground truth, \( Y \) denotes the predicted result, and \( \hat{p}_{i}, y_{i} \) represents the value of the \( i \)th pixel of the expected result and ground truth, respectively. The smooth parameter \( s \) prevents division by 0 and is set to 1. To solve the problem of class imbalance, we use the weighted Dice loss and BCE loss for the deep supervision loss and final segmentation:

\[
L_{\text{final}} = \lambda L_{\text{BCE}} + (1 - \lambda) L_{\text{Dice}}
\]

\[
= \lambda \left( -\frac{1}{N} \sum_{i} [y_{i} \log(p_{i}) + (1 - y_{i}) \log(1 - p_{i})] \right) + (1 - \lambda) \left( 1 - \frac{\sum_{i} \hat{p}_{i} y_{i} + s}{\sum_{i} \hat{p}_{i} + \sum_{i} y_{i} + s} \right)
\]

where \( \lambda \) denotes the trade-off parameter for the two losses, and we set \( \lambda = 0.3 \) in this study. The \( L_{\text{final}} \) refers to the loss of final segmentation (the \( F_{\text{out}} \) in Fig. 3). Similarly, we calculate the loss for the previously mentioned \( L_{\text{for}} \) and \( L_{\text{bg}} \) in this way. The total loss function can be written as:

\[
L_{\text{total}} = L_{\text{final}} + L_{\text{bg}} = L_{\text{final}} + \left( a \times L_{\text{for}} + b \times L_{\text{bg}} \right)
\]

### 4. Experiment

In this section, we will detail our experimental details, datasets, experimental benchmarks, and results on four different COVID-19 datasets. We also compare some classic segmentation models and some other methods. Moreover, our segmentation results are visualized to show the effectiveness of our proposed method more intuitively. To study the specific effects of different modules and parameter combinations on our proposed method, we conduct ablation experiments on the modules and parameters involved in the method.

#### 4.1. Datasets

To demonstrate our proposed method’s effectiveness in segmenting COVID-19 lesions, we selected four challenging COVID-19 datasets from the Kaggle competition, including two modal, CT, and X-ray. Experienced doctors label all datasets. CT image is 3D data. We slice them manually and resize them to 512 × 512 in training. The X-ray dataset was not resized in training; its original resolution is 224 × 224. According to the quantity after slicing and image modal, we rename them as COVID-19-CT-298, COVID-19-CT-505, COVID-19-CT-1650, and COVID-19-X-ray-2951. Corresponding quantities are 298, 505, 1650, and 2951 respectively, as shown in Table 1. In the training process, we randomly divided each dataset into the training set, validation set, and test set according to the ratio of 6:2:2. But MiniSeg [21] was trained using the five-fold cross-validation method in the original paper. We loaded MiniSeg model structure for a fair comparison between this research.

#### 4.2. Implementation details

We implement our proposed method with PyTorch on NVIDIA P100 GPU with 16 GB memories. We set the batch size of 4 and the initial learning rate of 0.01 with a polynomial learning rate scheduler. The model is trained using an SGD optimizer, with a weight decay of 1e-4.

#### Table 1

| Dataset | Quantity | Train-val-test split | Resolution | Modal |
|---------|----------|----------------------|------------|-------|
| COVID-19-CT-505 | 505 | 303-101-101 | 512×512 | CT |
| COVID-19-CT-1650 | 1650 | 990-330-330 | 512×512 | CT |
| COVID-19-CT-298 | 298 | 178-60-60 | 512×512 | CT |
| COVID-19-X-ray-2951 | 2951 | 1770-590-591 | 224×224 | X-ray |

2: https://medicationsegment.com/covid19/
3: https://academic torrents.com/details/f2175c467e0416a65568bb70c2bc d15c7325fd2/tech&hit=1&filelist=1
4: https://www.kaggle.com/datasets/aysendegerli/qatacov19-dataset?resource=download
5: https://www.kaggle.com/datasets/aysendegerli/qatacov19-dataset?resource=download
4.3. Evaluation metrics

To evaluate the effectiveness of our proposed method more objectively, we adopt five evaluation indicators commonly used in medical imaging analysis to measure the performance of the proposed method of the segmentation of COVID-19 lesions, such as Accuracy (ACC), Sensitivity, Specificity, Jaccard, and Dice. Note that the higher these indicators are, the better the segmentation effect is. Due to the extremely small and irregular lesions of COVID-19, there is an extremely class imbalance phenomenon. We mainly evaluate the results of model segmentation from Dice and Jaccard, supplemented by other indicators. Here is the calculation formula for each indicator:

\[
\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}
\]

\[
Jaccard = \frac{TP}{TP + FP + FN}
\]

\[
\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]

where \(TP\) and \(FP\) are true-positive and false-positive variables, representing the number of pixels of COVID-19 lesions correctly segmented and the number of background pixels incorrectly segmented by the model, respectively. Correspondingly, \(TN\) is the true negative variable, representing the number of correctly segmented background pixels. \(FN\) is the false-negative variable, representing the COVID-19 lesions pixels incorrectly labeled as background pixels.

4.4. Results

We evaluate our proposed method on four COVID-19 datasets, where three CT datasets are manually sliced from 3D volumes. We manually remove some images without lesions to alleviate the training difficulty from class imbalance to obtain the final dataset. For four COVID-19 datasets, we adopt the same indicators to analyze the segmentation results quantitatively.

4.4.1. Results of COVID-19-CT-505 segmentation

Compared with the baseline model (UNet) and other segmentation methods, our proposed method performs the best on the COVID-19-CT-505 dataset. Table 2 shows the comparison results of different methods. Compared with the baseline model, Dice increases from 0.7524 ± 0.1553 to 0.7978 ± 0.1423. Meanwhile, four indicators significantly improved, including Jaccard, ACC, Sensitivity, and Specificity. Our proposed method achieves some improvements on Dice compared with Unet++ [12], Channel Unet [42], Attention Unet [11], MiniSeg [21], PSPNet [9], DeepLab V3 [15], CEnet [19], and Dualnorm Unet [43], respectively, which demonstrates the effectiveness of the proposed method. It is noted that the Sensitivity is slightly lower than the other methods due to a trade-off between Specificity and Sensitivity.

As seen in Fig. 4, the lesions of COVID-19 are irregular and very small. Our model clearly distinguishes the adhesion, cavity, and boundary of lesions, which validates the effectiveness of our proposed method. However, other models cannot adapt well to the complexity of COVID-19 images, which generates some missing or incorrect predictions of target areas. Due to the double-decoder structure, the background segmentation branch promotes the foreground target prediction. Our model outperforms the other state-of-the-art methods due to the combination of MHSA and double-decoder.

4.4.2. Results of COVID-19-X-ray-2951 segmentation

Table 3 shows our segmentation results on the COVID-19-X-ray-2951 dataset. Different from the three CT datasets, the image resolution is 224 × 224, TDD-Unet model is used to solve the segmentation problem of very-low-contrast X-ray and the indistinct boundary between the infected area and normal tissues. We observe that our proposed method excels with the other nine models for four indicators, including Dice, Jaccard, ACC, and Specificity. However, our proposed method has slightly less Sensitivity. We visualize the segmentation results of the comparable models in Fig. 5. The proposed method gains better segmentation results, which are more similar to the ground truth than the results of other models. Experimental results show superior performance against other competitors, confirming the effectiveness of our proposed method.

4.4.3. Results of COVID-19-CT-1650 segmentation

Table 4 shows the compared results on the COVID-19-CT-1650 dataset. We can observe that the segmentation performance of COVID-19-CT-1650 is better than that of COVID-19-CT-505 and COVID-19-X-ray-2951 because the Dice value of Unet reaches 0.8656 ± 0.0699. The image number of the COVID-19-CT-1650 is more than that of the COVID-19-CT-505 dataset. The X-ray is the observation of overlapping images of the inspection site, which is challenging to provide high resolution comparable to CT images and is easily interfered with by other structures or organs. CT scans are often used for cross-sectional imaging to provide more three-dimensional information and observe lesions that cannot be seen with X-rays. Although X-rays do less damage to the human body, CT images are more sensitive than others, along with more accurate segmentation results.

Fig. 6 compares our model with other Unet models without multi-head self-attention. From Fig. 6, we find that our proposed method has a powerful ability to capture the global dependence and the missing parts. These results verify that our model with the multi-head self-attention provides more detailed information on boundary segmentation, resulting in better performance than other methods.

4.4.4. Results of COVID-19-CT-298 segmentation

As we can see from Table 5, our proposed method outperforms other models in terms of key metrics. Dice increases from 0.7556 ± 0.2340 to 0.7680 ± 0.2154 in our model compared with the baseline Unet model, and Jaccard increases from 0.6522 ± 0.2427 to 0.6633 ± 0.2309. Sensitivity in our model has an increase of 0.0217. There is little difference between the two indicators with Accuracy and Specificity for all compared models, which are not able to perform well on the training 298 images. Despite large lesions, the object area of segmentation is more irregular and scattered than in the COVID-19-CT-505 dataset and COVID-19-CT-1650 dataset. From Fig. 7, we can observe that other competitors cannot segment small objects of COVID-19 lesions very well because they are unable to extract effective features. In contrast, our method can accurately detect tiny objects due to the combination of the MHSA module with double decoder layers. Moreover, our proposed TDD-Unet can reduce the area of false-negative regions while limiting the expansion of false-positive regions compared with other baseline models. We note that the segmentation results given by other models are not satisfactory, and they incorrectly segment most of the small disconnected regions. The results of this experiment show that ordinary Unet architecture is deficient in capturing the complex global feature information on chest CT images. Our TDD-Unet tackles these shortcomings by restructuring Unet architecture and collaborating with MHSA and double decoder.
Table 2
Results on COVID-19-CT-505.

| Model          | Dice      | Jaccard   | Acc        | Sensitivity | Specificity     |
|----------------|-----------|-----------|------------|-------------|-----------------|
| U-Net          | 0.7524 ± 0.1553 | 0.6247 ± 0.1774 | 0.9966 ± 0.0026 | 0.8124 ± 0.1833 | 0.9981 ± 0.0016 |
| U-Net++        | 0.7344 ± 0.2162 | 0.6168 ± 0.2196 | 0.9970 ± 0.0024 | 0.7485 ± 0.2446 | 0.9987 ± 0.0011 |
| Channel_U-Net | 0.7800 ± 0.1853 | 0.6683 ± 0.1963 | 0.9974 ± 0.0021 | 0.7906 ± 0.2155 | 0.9988 ± 0.0009 |
| Attention_U-Net | 0.7498 ± 0.1467 | 0.6189 ± 0.1668 | 0.9965 ± 0.0024 | 0.8304 ± 0.1714 | 0.9978 ± 0.0018 |
| MiniSeq        | 0.7788 ± 0.1437 | 0.6567 ± 0.1652 | 0.9971 ± 0.0022 | 0.8205 ± 0.1713 | 0.9984 ± 0.0014 |
| PSPNet         | 0.7408 ± 0.1616 | 0.6100 ± 0.1734 | 0.9966 ± 0.0024 | 0.7748 ± 0.1797 | 0.9983 ± 0.0012 |
| DeepLabV3      | 0.7181 ± 0.1450 | 0.5779 ± 0.1599 | 0.9962 ± 0.0027 | 0.7680 ± 0.1733 | 0.9980 ± 0.0013 |
| CEnet          | 0.7711 ± 0.1424 | 0.6461 ± 0.1644 | 0.9970 ± 0.0021 | 0.8044 ± 0.1649 | 0.9985 ± 0.0011 |
| DualNorm_U-Net | 0.7653 ± 0.1615 | 0.6429 ± 0.1810 | 0.9971 ± 0.0025 | 0.7838 ± 0.1914 | 0.9986 ± 0.0011 |
| TDD-U-Net(Ours)| **0.7978 ± 0.1423** | **0.6827 ± 0.1659** | **0.9975 ± 0.0019** | **0.7918 ± 0.1578** | **0.9991 ± 0.0009** |

Fig. 4. Qualitative results of comparison with the proposed method and other models on COVID-19-CT-505.

Table 3
Results on COVID-19-X-ray-2951.

| Model          | Dice      | Jaccard   | Acc        | Sensitivity | Specificity     |
|----------------|-----------|-----------|------------|-------------|-----------------|
| U-Net          | 0.7423 ± 0.2143 | 0.6300 ± 0.2374 | 0.9496 ± 0.0399 | 0.8339 ± 0.1956 | 0.9646 ± 0.0362 |
| U-Net++        | 0.7347 ± 0.2266 | 0.6232 ± 0.2416 | 0.9510 ± 0.0394 | 0.7935 ± 0.2240 | 0.9710 ± 0.0328 |
| Channel_U-Net | 0.7419 ± 0.2192 | 0.6311 ± 0.2417 | 0.9479 ± 0.0430 | 0.8438 ± 0.1788 | 0.9624 ± 0.0440 |
| Attention_U-Net | 0.7422 ± 0.2135 | 0.6299 ± 0.2375 | 0.9506 ± 0.0412 | 0.8096 ± 0.2097 | 0.9686 ± 0.0337 |
| MiniSeq        | 0.7280 ± 0.2226 | 0.6136 ± 0.2394 | 0.9488 ± 0.0444 | 0.7845 ± 0.2214 | 0.9699 ± 0.0378 |
| PSPNet         | 0.7320 ± 0.2235 | 0.6197 ± 0.2450 | 0.9493 ± 0.0442 | 0.7977 ± 0.2161 | 0.9673 ± 0.0395 |
| DeepLabV3      | 0.7222 ± 0.2280 | 0.6084 ± 0.2460 | 0.9477 ± 0.0451 | 0.7861 ± 0.2257 | 0.9671 ± 0.0388 |
| CEnet          | 0.7373 ± 0.2266 | 0.6275 ± 0.2468 | 0.9484 ± 0.0434 | 0.8415 ± 0.2120 | 0.9599 ± 0.0446 |
| DualNorm_U-Net | 0.7411 ± 0.2206 | 0.6303 ± 0.2416 | 0.9512 ± 0.0416 | 0.8141 ± 0.2098 | 0.9698 ± 0.0355 |
| TDD-U-Net(Ours)| **0.7506 ± 0.2151** | **0.6415 ± 0.2401** | **0.9536 ± 0.0407** | **0.8124 ± 0.2030** | **0.9737 ± 0.0296** |
4.4.5. Statistical evaluation

We further incorporate ROC curves and AUC criteria to evaluate the capability of different models on four datasets. Based on ROC curves and AUC values, as shown in Fig. 8, we can see that our proposed method’s ROC and AUC values are better than other competitors. Moreover, the Dice score and Jaccard boxplots on four datasets are shown in Fig. 9. Higher Dice and Jaccard mean scores and the lowest standard deviation occur compared to the other segmentation models, indicating our proposed method’s generality.

4.5. Ablation study of parameter

In this section, we will study the effect of the parameter combination of the loss function on the segmentation performance. Considering that the dual-decoding branch structure is adopted in our method, deep supervision is used to supervise the segmentation results obtained from the foreground and background decoding branches. The weight distribution method of the loss function is worth exploring. For a single decoder branch, we use BCELoss+DiceLoss for training. We use parameters $a$ and $b$ to represent the proportion of the loss function of the foreground branch and the background branch in the total loss function, respectively. The step size is set to 0.1, and the experiment is repeated. The experimental results are shown in Table 6. We observe that when the weight of the loss function of $a$ in the foreground branch is 0.7 and $b$ in the background branch is 0.3, the model achieves a better result, which shows that the accuracy of foreground segmentation is higher than the prediction of background segmentation.

4.6. Ablation study of module

To investigate the effectiveness of each module, we designed four sets of controlled trials, denoted as UNet, UNet+MHSA, UNet+DD, and our proposed TDD-Unet. The results of four sets of controlled trials are summarized in Table 7. Moreover, we compare the experimental control group from multiple indicators in Fig. 10, and the results reveal that our proposed TDD-Unet performs much better than other experimental control groups.

**Effectiveness of MHSA.** From Table 7, we observe that the model’s performance significantly improves due to the introduction of MHSA, surpassing all five indicators of the baseline UNet. The Dice indicator improves from 0.7524 ± 0.1553 to 0.7816 ± 0.1598, demonstrating the effectiveness of MHSA in capturing long-range relationships. The encoding layer gets more precise feature information, which gains more effectiveness of MHSA in capturing long-range relationships. The experimental control groups.

| Model        | Dice       | Jaccard     | Acc         | Sensitivity | Specificity |
|--------------|------------|-------------|-------------|-------------|-------------|
| UNet         | 0.7556 ± 0.2340 | 0.6522 ± 0.2427 | 0.9959 ± 0.0045 | 0.7618 ± 0.2506 | 0.9977 ± 0.0030 |
| UNet++       | 0.7487 ± 0.2362 | 0.6406 ± 0.2474 | 0.9959 ± 0.0047 | 0.7617 ± 0.2474 | 0.9977 ± 0.0032 |
| Channel-Unet | 0.7616 ± 0.2306 | 0.6594 ± 0.2409 | 0.9959 ± 0.0046 | 0.7548 ± 0.2566 | 0.9976 ± 0.0031 |
| Attention-Unet | 0.7434 ± 0.2449 | 0.6405 ± 0.2539 | 0.9960 ± 0.0043 | 0.7340 ± 0.2633 | 0.9978 ± 0.0028 |
| MiniSeg      | 0.7484 ± 0.2327 | 0.6421 ± 0.2402 | 0.9956 ± 0.0048 | 0.7644 ± 0.2468 | 0.9970 ± 0.0037 |
| PSPNet       | 0.7447 ± 0.2184 | 0.6328 ± 0.2294 | 0.9954 ± 0.0051 | 0.7654 ± 0.2238 | 0.9973 ± 0.0031 |
| DeepLabV3    | 0.6912 ± 0.2630 | 0.5778 ± 0.2487 | 0.9941 ± 0.0066 | 0.6976 ± 0.2791 | 0.9967 ± 0.0041 |
| CNet         | 0.7615 ± 0.2202 | 0.6563 ± 0.2354 | 0.9961 ± 0.0042 | 0.7649 ± 0.2434 | 0.9977 ± 0.0028 |
| Dualnorm-Unet | 0.7505 ± 0.2430 | 0.6485 ± 0.2494 | 0.9960 ± 0.0043 | 0.7637 ± 0.2666 | 0.9974 ± 0.0030 |
| TDD-Unet(Ours) | 0.7680 ± 0.2154 | 0.6633 ± 0.2309 | 0.9959 ± 0.0047 | 0.7835 ± 0.2383 | 0.9974 ± 0.0031 |
Fig. 6. The result of our proposed method (with MHSA) compared to other attention models on COVID-19-CT-1650.

Fig. 7. Visual comparison between different models in our system for COVID lesions segmentation on COVID-19-CT-298. The colors red, blue, and green represent the correct segmentation, the under-segmentation, and the over-segmentation, respectively.
Fig. 8. ROC curves of different models for COVID-19 lesions segmentation. (a) COVID-19-CT-505, (b) COVID-19-X-ray-2951, (c) COVID-19-CT-1650. (d) COVID-19-CT-298.

Table 6
Ablation study for parameter on COVID-19-CT-505 (a-fore; b-bg).

| Model (TDD-UNet) | Dice       | Jaccard    | Acc      | Sensitivity | Specificity |
|------------------|------------|------------|----------|-------------|-------------|
| (a=0.1, b=0.9)   | 0.8034 ± 0.1477 | 0.6647 ± 0.1749 | 0.9972 ± 0.0022 | 0.8165 ± 0.1729 | 0.9986 ± 0.0011 |
| (a=0.2, b=0.8)   | 0.7917 ± 0.1411 | 0.6744 ± 0.1682 | 0.9973 ± 0.0022 | 0.8102 ± 0.1745 | 0.9988 ± 0.0010 |
| (a=0.3, b=0.7)   | 0.7873 ± 0.1548 | 0.6717 ± 0.1802 | 0.9972 ± 0.0021 | 0.8493 ± 0.1686 | 0.9983 ± 0.0014 |
| (a=0.4, b=0.6)   | 0.7780 ± 0.1694 | 0.6620 ± 0.1886 | 0.9971 ± 0.0024 | 0.8117 ± 0.1863 | 0.9986 ± 0.0013 |
| (a=0.5, b=0.5)   | 0.7692 ± 0.1409 | 0.6806 ± 0.1672 | 0.9973 ± 0.0021 | 0.8468 ± 0.1509 | 0.9985 ± 0.0011 |
| (a=0.6, b=0.4)   | 0.7645 ± 0.1507 | 0.6666 ± 0.1741 | 0.9973 ± 0.0021 | 0.8053 ± 0.1805 | 0.9988 ± 0.0010 |
| (a=0.7, b=0.3)   | 0.7978 ± 0.1423 | 0.6862 ± 0.1659 | 0.9975 ± 0.0019 | 0.7918 ± 0.1578 | 0.9991 ± 0.0009 |
| (a=0.8, b=0.2)   | 0.7895 ± 0.1539 | 0.6821 ± 0.1696 | 0.9975 ± 0.0021 | 0.8018 ± 0.1804 | 0.9990 ± 0.0009 |
| (a=0.9, b=0.1)   | 0.7819 ± 0.1513 | 0.6636 ± 0.1787 | 0.9973 ± 0.0020 | 0.8136 ± 0.1738 | 0.9986 ± 0.0012 |

Table 7
Ablation study for module on COVID-19-CT-505.

| Model (TDD-UNet) | Dice       | Jaccard    | Acc      | Sensitivity | Specificity |
|------------------|------------|------------|----------|-------------|-------------|
| UNet (Baseline)  | 0.7524 ± 0.1553 | 0.6247 ± 0.1774 | 0.9966 ± 0.0026 | 0.8124 ± 0.1833 | 0.9981 ± 0.0016 |
| UNet (+MHSA)     | 0.7816 ± 0.1598 | 0.6637 ± 0.1741 | 0.9972 ± 0.0021 | 0.8189 ± 0.1734 | 0.9986 ± 0.0010 |
| UNet (+DD)       | 0.7690 ± 0.1640 | 0.6373 ± 0.1879 | 0.9970 ± 0.0023 | 0.7828 ± 0.2108 | 0.9986 ± 0.0012 |
| TDD-UNet (Ours)  | 0.7978 ± 0.1423 | 0.6827 ± 0.1659 | 0.9975 ± 0.0019 | 0.7918 ± 0.1578 | 0.9991 ± 0.0009 |

Effectiveness of Double-Decoder. To verify the effectiveness of the double-decoder, we re-conduct the experiment without MHSA and adopt the optimal parameter. From Table 7, we can see that compared with the baseline model, the segmentation results from the double-decoder branch significantly improve, and the Dice indicator increases from 0.7524 ± 0.1553 to 0.7600 ± 0.1640. Moreover, the segmentation performance with a double-decoder is lower than that with MHSA, which proves the limitations of convolution.

Effectiveness of MHSA & Double-Decoder. The model with MHSA and double-decoder is tested to evaluate the performance. From Table 7, we observe that the model’s performance further improves compared with UNet with MHSA (UNet + MHSA) and UNet with double-decoder (UNet + DD). The Dice indicator increases from 0.7524 ±
Fig. 9. Boxplots of Dice coefficient (DSC) and Jaccard scores for all test samples of: (a) COVID-19-CT-505, (b) COVID-19-X-ray-2951, (c) COVID-19-CT-1650, and (d) COVID-19-CT-298. Different boxes indicate the score ranges of several methods. The red line inside each box represents the median value. All values outside the whiskers are considered outliers, which are marked with the ♢ symbol.

Fig. 10. Visualization results of module ablation. All metrics are the higher the better. Sen represents Sensitivity and Jac represents Jaccard. Our proposed TDD-UNet is better than others.

0.1553 for the UNet baseline to 0.7978 ± 0.1423 for our method, which indicates that MHSA and double-decoder promote each other. MHSA enables the model to obtain more detailed feature information on the encoding stage, which improves the segmentation results of two decoding branches. The segmentation result of the background branch also enhances foreground prediction, which validates the effectiveness of the combination of MHSA and double-decoder.

5. Discussion

Image segmentation is a crucial task in medical imaging. Accurate segmentation of lesions can help doctors make clinical diagnoses and follow-up treatment plans. Recently, Transformer has shown promising results from sequence data and capturing global relationships, widely used in computer vision. However, the computational overhead of calculating relationships between individual chunks is not negligible. Moreover, Transformer needs a lot of data training to learn the position relationship between patches. We propose an efficient double decoder UNet with Transformer for COVID-19 lesions segmentation in this work. Our proposed method segments four different medical image tasks, indicating that the generalization and robustness of our proposed method outperform other competitors.

Our proposed method uses the MHSA module to reduce the computing overhead of the complete Transformer. We add the MHSA module after the last encoding layer to maximize the benefits of long-range dependence extraction from multiple self-attention. To further improve the segmentation performance, we propose a foreground–background double decoding module, which achieves good results. The encoding layer of UNet is mainly used for feature extraction. The parallel double decoder structure is the kernel component of our proposed method to improve the performance, which surpasses the previous methods of the COVID-19 lesions segmentation task.

Our proposed method is effective, but there are still some shortcomings. The proposed foreground–background doubled decoding branch adopts the same convolution, which ignores the imbalance of foreground and background pixels. Moreover, our proposed method requires large-scale datasets to achieve better performance. In future work, we will investigate that a more effective double decoding branch may be designed and used to achieve better results.

6. Conclusion

Computer-aided medical diagnosis is of great significance in rapid screening for COVID-19. In this paper, we propose a new architecture, TDD-UNet, for automatically segmenting lesions from lung images of COVID-19 patients. To demonstrate the effectiveness and robustness of our proposed method, we conduct training and testing on four challenging and heterogeneous COVID-19 datasets with two modes. We segment foreground and background, respectively, which promotes the prediction of foreground through the result of background segmentation. Moreover, we introduce the MHSA modules of the Transformer to extract long-range dependencies to solve the local defects of traditional convolutional feature extraction and decrease parameter consumption. We prove that the combination of MHSA with the UNet branch improves the segmentation effect by performing the ablation experiment.
Experimental results show that the proposed method is significantly superior to the baseline model (UNet) and exceeds other state-of-the-art methods. In the future, we will focus on the relationship between space and channel level by exploring the relationship between foreground and background from the image level. TDD-UNet network will be improved to adapt to more segmentation tasks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (grant numbers 61504055 and 61701218), and the Natural Science Foundation of Hunan Province of China (grant numbers 61504055 and 61701218), and the Natural Science Foundation of Hunan Province of China (grant numbers 2020JJ4514 and 2020JJ4519). This study was also funded by Hunan provincial base for scientific and technological innovation cooperation.

Appendix

As a further test of the robustness of our method, we added 1000 images to COVID-19-X-ray-2951 containing images of healthy people and patients with other lung diseases. The same experimental settings were applied to a combined dataset and re-experimented. The recently added images are also from the Kaggle competition,

and here we provide the download link of the combined dataset. Our experimental results on the combined datasets are shown in Table 8 and Fig. 11. Our proposed method still achieved relatively high performance, Dice, Jaccard, Acc, Sensitivity, and Specificity with 0.7894 ± 0.2559, 0.7111 ± 0.2834, 0.9634 ± 0.0429, 0.8246 ± 0.2392, and 0.9781 ± 0.0399, respectively, among which Dice, Jaccard, Acc, Sensitivity, and Specificity were all the highest. Our proposed method maintains accurate segmentation results when we incorporate some control images, as shown in Fig. 11. Furthermore, our segmentation results outperform other compared models at the lesion boundary, demonstrating the robustness of our proposed method.

References

[1] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, W. Tan, Detection of SARS-CoV-2 in different types of clinical specimens, JAMA 323 (18) (2020) 1843–1844.
[2] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, Radiology 296 (2) (2020) E32–E40.
[3] N. Islam, S. Ebrahimzadeh, J.P. Salameh, S. Kazi, N. Fabiano, L. Treanor, M. Absi, Z. Hallgrimsson, M.M. Leeflang, L. Hooft, et al., Thoracic imaging tests for the diagnosis of COVID-19, Cochrane Database Syst. Rev. (3) (2021).
[4] Z. Ye, Y. Zhang, V. Wang, Z. Huang, B. Song, Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review, Eur. Radiol. 30 (8) (2020) 4381–4389.
[5] V. Rajinikanth, N. Dey, A.N.J. Raj, A.E. Hassani, K. Santosh, N. Raja, Harmony-search and otu based system for coronavirus disease (COVID-19) detection using lung CT scan images, 2020, arXiv preprint arXiv:2004.03431.
[6] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012) 1097–1105.
[7] https://www.hxp2396.xyz:10009/index.php?share/file&user=1&sid=6auN9q4i
[24] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: A multitask edge-aware learning, Med. Phys. 48 (4) (2021) 1771–1780.
[25] Z. Gao, B. Hong, X. Zhang, Y. Li, C. Jia, J. Wu, C. Wang, D. Meng, C. Li, Instance-based vision transformer for subtyping of papillary renal cell carcinoma in histopathological image, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 299–308.
[26] O. Petit, N. Thome, L. Rambour, L. Theymy, T. Collins, L. Soler, U-net transformer: Self and cross attention for medical image segmentation, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2021, pp. 267–276.
[27] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transnet: Transformers make strong encoders for medical image segmentation, 2021, arXiv preprint arXiv:2102.04306.
[28] Y. Zhang, H. Liu, Q. Hu, Transfuse: Fusing transformers and cnns for medical image segmentation, 2021, arXiv preprint arXiv:2102.08005.
[29] X. Chen, L. Yao, Y. Zhang, Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images, 2020, arXiv preprint arXiv:2004.05645.
[30] T. Zhou, S. Canu, S. Ruan, An automatic covid-19 ct segmentation based on u-net with attention mechanism, 2020, arXiv preprint arXiv:2004.06673.
[31] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography, Sci. Rep. 10 (1) (2020) 1–11.
[32] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Gao, M. Cai, J. Yang, Y. Li, X. Meng, et al., A deep learning algorithm using CT images to screen for Corona Virus disease (COVID-19), Eur. Radiol. (2021) 1–9.
[33] D.-P. Fan, T. Zhou, G.P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Inf-net: Automatic covid-19 lung infection segmentation from ct images, IEEE Trans. Med. Imaging 39 (8) (2020) 2626–2637.
[34] D. Yang, Z. Xu, W. Li, A. Myronenko, H.R. Roth, S. Harmon, S. Xu, B. Turkbey, E. Turkbey, X. Wang, et al., Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan, Med. Image Anal. 70 (2021) 101992.
[35] X. Wang, L. Jiang, L. Li, M. Xu, X. Deng, L. Dai, X. Xie, T. Li, Y. Guo, Z. Wang, et al., Joint learning of 3d lesion segmentation and classification for explainable COVID-19 diagnosis, IEEE Trans. Med. Imaging (2021).
[36] Q. Yu, L. Qi, Y. Yao, W. Wang, Y. Shi, Crosslink-net: double-branch encoder network via fusing vertical and horizontal convolutions for medical image segmentation, IEEE Transactions on Image Processing 31 (2022) 5893–5908.
[37] F. Isemee, P. Kickingereder, W. Wick, M. Bendzus, K.H. Maier-Hein, Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge, in: International MICCIA Brainlesion Workshop, Springer, 2017, pp. 287–297.
[38] H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, IEEE Trans. Comput. Imag. 3 (1) (2016) 47–57.
[39] F. Milletari, N. Navab, S.A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV, IEEE, 2016, pp. 565–571.
[40] C.H. Sudre, M. Li, T. Vercauteren, S. Ourselin, M.J. Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp. 240–248.
[41] Q. Qiu, Z. Yang, S. Wu, D. Qian, J. Wei, G. Gong, L. Wang, Y. Yin, Automatic segmentation of hippocampus in hippocampal sparing whole brain radiotherapy: A multitask edge-aware learning, Med. Phys. 48 (4) (2021) 1771–1780.
[42] Y. Chen, K. Wang, X. Liao, Y. Qian, W. Zhang, Z. Yuan, P.A. Heng, Channel-unet: a spatial channel-wise convolutional neural network for liver and tumors segmentation, Front. Genet. 10 (2019) 1110.
[43] J. Xiao, L. Yu, L. Xing, A. Yuille, Y. Zhou, DualNorm-UNet: Incorporating global and local statistics for robust medical image segmentation, 2021, arXiv preprint arXiv:2103.15856.