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.

Computers in Biology and Medicine 149 (2022) 105981

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

HADCNet: Automatic segmentation of COVID-19 infection based on a hybrid attention dense connected network with dilated convolution

Ying Chen [a], Taohui Zhou [a,*], Yi Chen [b,**], Longfeng Feng [a], Cheng Zheng [a], Lan Liu [c,***], Liping Hu [c], Bujian Pan [d]

[a] School of Software, Nanchang Hangkong University, Nanchang, 330063, PR China
[b] Department of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, 325035, PR China
[c] Department of Radiology, Jiangxi Cancer Hospital, Nanchang, 330029, PR China
[d] Department of Hepatobiliary Surgery, Wenzhou Central Hospital, The Dingli Clinical Institute of Wenzhou Medical University, Wenzhou, Zhejiang, 325000, PR China

ARTICLE INFO

Keywords:
HADCNet
Dual hybrid attention strategy
COVID-19 infection
Segmentation
Deep learning

ABSTRACT

the automatic segmentation of lung infections in CT slices provides a rapid and effective strategy for diagnosing, treating, and assessing COVID-19 cases. However, the segmentation of the infected areas presents several difficulties, including high intraclass variability and interclass similarity among infected areas, as well as blurred edges and low contrast. Therefore, we propose HADCNet, a deep learning framework that segments lung infections based on a dual hybrid attention strategy. HADCNet uses an encoder hybrid attention module to integrate feature information at different scales across the peer hierarchy to refine the feature map. Furthermore, a decoder hybrid attention module uses an improved skip connection to embed the semantic information of higher-level features into lower-level features by integrating multi-scale contextual structures and assigning the spatial information of lower-level features to higher-level features, thereby capturing the contextual dependencies of lesion features across levels and refining the semantic structure, which reduces the semantic gap between feature maps at different levels and improves the model segmentation performance. We conducted fivefold cross-validations of our model on four publicly available datasets, with final mean Dice scores of 0.792, 0.796, 0.785, and 0.723. These results show that the proposed model outperforms popular state-of-the-art semantic segmentation methods and indicate its potential use in the diagnosis and treatment of COVID-19.

1. Introduction

Coronavirus disease 2019 (COVID-19), which has high rates of infectivity and lethality, has spread worldwide, resulting in an urgent health crisis across the world and necessitating the development of a rapid, robust and effective detection method to slow or halt the spread of the virus. In clinical practice, the reverse transcription-polymerase chain reaction (RT–PCR) test is considered the gold standard for detecting COVID-19 due to its high specificity [1]; however, the assessment of RT–PCR assays is time-consuming, insensitive and nondynamic, and more efficient alternatives are needed [2]. Computed tomography (CT) imaging is one of the most commonly used methods for detecting lung infections due to its high spatial resolution and the unique relationship between CT images and the air content in the lungs. Several studies have demonstrated the high sensitivity and rapid response of CT scans for detecting COVID-19 infections [3,4], with CT scans having the ability to accurately locate and dynamically reflect changes in the infected area of the lungs during treatment; thus, CT imaging is an effective alternative method for diagnosing COVID-19. As the manual examination of a large number of CT scans is time-consuming and yields subjective and biased results, this paper primarily focuses on the automatic segmentation of COVID-19 infections to assist physicians in rapid diagnosis and treatment assessment. However, due to differences in the location and extent of viral invasion, COVID-19 infections exhibit more complex features than common pneumonia, necessitating the development of segmentation networks with strong image semantic feature understanding and the

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: c_y2008@nchu.edu.cn (Y. Chen), 3156574420@qq.com (T. Zhou), kenyoncy2016@gmail.com (Y. Chen), flf1998@qq.com (L. Feng), ZasonCheng@163.com (C. Zheng), liulan6688@163.com (L. Liu), 158562940@qq.com (L. Hu), panbujian@126.com (B. Pan).

https://doi.org/10.1016/j.compbiomed.2022.105981
Received 3 July 2022; Received in revised form 3 August 2022; Accepted 14 August 2022
Available online 20 August 2022

ability to respond to various lesions to preserve fine-grained feature details and handle the complex diversity of lesion structures at the pixel level. The green, blue and red areas in Fig. 1 are characteristic representations of solid lung lesions, bilateral patchy shadows, and asymmetric gross glassy opacities (GGOs), respectively, with GGOs and solid lung lesions showing peripheral distributions. In addition, the low contrast and blurred boundaries increase the segmentation difficulty, with high interclass similarity and high intraclass variability being the main factors affecting the segmentation accuracy of COVID-19 infections.

To address the above issues, this paper proposes HADCNet, a novel automatic segmentation network for COVID-19 lesions that aggregates the contextual semantic information from images at the peer and cross levels to refine image features and enrich feature representations. The encoder hybrid attention module efficiently assembles semantic feature information from different scales at the peer level, while the decoder hybrid attention module integrates feature contextual dependencies across levels through improved skip connections, effectively balancing feature differences between the encoder and the decoder. The boundary details and refined structural information are preserved, resulting in richer fine-grained feature representations, and the sensitivity of the network is significantly improved, leading to an improved segmentation performance. The proposed method was tested on four public COVID-19 infection datasets and outperformed state-of-the-art segmentation networks in terms of the segmentation performance.

In summary, the main contributions of this study can be summarized as follows:

- This paper presents a COVID-19 lesion automatic segmentation network (HADCNet) based on the dual hybrid attention strategy.
- HADCNet uses the encoder hybrid attention module and the decoder hybrid attention module to capture context dependencies in images at the peer and across levels, reducing the semantic gap between feature maps from different levels and improving the segmentation performance of the model.
- To validate the segmentation performance of the proposed model, HADCNet was compared with popular state-of-the-art methods on four publicly available datasets.
- the effectiveness of the dual hybrid attention strategy was verified through extensive ablation experiments.

The remainder of this paper is structured as follows: Section 2 introduces related work on COVID-19 infection segmentation. Section 3 describes the proposed network in detail. The experimental results, analyses, conclusions, and a discussion of future work are presented in Sections 4-5.

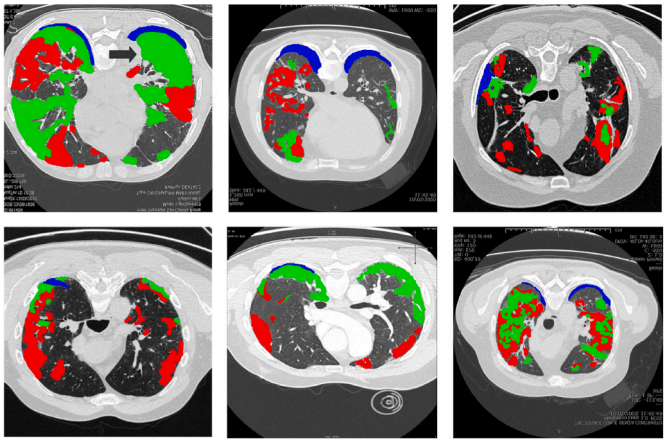

Fig. 1. Example images of different features of COVID-19 infections.

2. Related work

2.1. Segmentation in medical images

The segmentation of lesions from medical images can provide doctors with critical information for diagnosing and quantifying diseases. Traditional medical image segmentation algorithms typically extract features manually using physical information such as the texture, structure and location of the image, with a reliance on extensive preprocessing operations and an experienced manual design process [5]. Segmentation methods based on machine learning achieve the segmentation goal by focusing on algorithms and predicting personalized features. Recent studies on the automatic segmentation of medical images have concentrated on the combined use of traditional and machine learning to achieve this goal. Su et al. [6] efficiently improved the segmentation performance of COVID-19 lesions through a multilevel thresholding image segmentation method. Liu et al. [7] applied a multilevel segmentation model based on modified differential evolution to segment breast cancer images. References [8–11] implemented segmentation of lesion images based on swarm intelligence optimization algorithms.

With the development of computer hardware and technology, convolutional neural networks (CNNs) have achieved great success in various fields in image classification tasks, and some excellent CNN-based networks have been developed, including fully connected networks (FCNs) [12], U-Net [13], AlexNet [14], VGGNet [15], ResNet [16], DenseNet [17], and generative adversarial networks (GANs) [18]. These networks have also been applied in the field of medical image segmentation. For example, Guan et al. [19] proposed a multichannel progressive generative adversarial network based on texture constraints. The CNN-based FCN achieved the first end-to-end segmentation and performed better than traditional manual feature segmentation methods. Based on the FCN implementation, U-Net introduces a skip connection between the encoder and the decoder at the peer level to fuse high-level and low-level image features, effectively solving the problem of information loss during the decoding process. However, U-Net has several limitations; skip connections do not fully utilize encoder features, and feature information can be lost due to up-sampling (such as bilinear interpolation). In response to the shortcomings of U-Net, several studies have attempted to improve this network, for example, by improving the skip connections [20–22], introducing cascade structures [23,24], and modifying the U-Net encoder layer structure [25,26]. Zhou et al. [22] proposed the U-Net++ network, which uses a nested skip connection structure to integrate features across different semantic levels and thus enables highly flexible feature fusion by fully exploiting image features at different scales. In addition, the introduction of structures such as conditional random fields [27], Markov random fields [28], and spatial pyramid pools [29] enriches contextual feature representations while preserving image details. Nevertheless, the performance of many encoder-decoder-based segmentation algorithms remains limited because of problems such as restricted local acceptance domains, significant differences between encoder and decoder feature maps, and the inadequate use of contextual information.

2.2. Computer-aided COVID-19 image analysis

The automatic segmentation of infected regions or lesions in the lungs can assist physicians in evaluating and quantifying lung disease, which is crucial for the diagnosis of COVID-19 and follow-up treatments. Xie et al. [30] applied a relational approach with nonlocal neural network coding blocks (RTSU-Net) to capture structured information between convolutional features. Chen et al. [31] built an interactive attentional refinement network (RefNet) that enhances the discrimination of complex features by combining residual learning and attention.

Although the target lesion segmentation accuracy of the above methods is better than that of traditional FCNs or U-shaped networks, several unresolved issues remain. First, lesion segmentation relies on local fine-grained texture information; however, an encoder with only a single-scale receptive domain cannot fully capture the overall texture representation of the lesion and is prone to losing structural differences in the contextual feature information at the peer level. Second, due to the short-term context dependency of U-Net and its variants, only encoded features at the same level are integrated into the decoder, resulting in large feature differences between the encoder and the decoder and increasing the difficulty of accurately capturing detailed feature representations of lesion boundaries. Finally, due to the structure of the encoder in conventional pixel-level supervision, some detailed texture information is lost, and the refinement of complex and diverse lesion feature representations is intractable.

2.3. Attention mechanism

Attention, which uses top-level information to guide a bottom-up feedforward procedure for cyclically processing visual information, plays an essential role in human visual cognition [32,33]. Xiaohang et al. [34] deployed a deep learning framework that includes a multimodal spatial attention module to automatically learn the spatial regions of features and segment lung tumours using the generated spatial attention maps. As application scenarios become more diverse and robustness requirements increase, hybrid attention mechanisms have become more widely used. Zhao et al. [35] presented an integrated spatial and channel attention semantic segmentation network (SCAU-Net) that enhances locally relevant features at the spatial and channel levels while suppressing irrelevant features. Fu et al. [36] developed a dual-attention network that combines channel attention and positional attention to capture rich contextual dependencies for segmentation by integrating local features with global dependencies through a self-attention mechanism. These algorithms typically produce refined feature representations in the channel dimension, spatial dimension, or a combination of the two, enhancing the information region representation of the target structure while suppressing irrelevant features; thus, the network can learn more general visual structures.

2.3.1. Squeeze-and-excitation block

The squeeze-and-excitation (SE) module [37] is a typical channel attention mechanism that uses two operations, squeezing and excitation, to learn the importance of different channel features, thereby improving the classification accuracy of feature maps. Specifically, the squeezing operation is implemented through global average pooling (GAP), which ignores the spatial information of the feature map to ensure that more channel information and the global perceptual field of the feature map are obtained. The excitation operation captures the weight relationship between channels through two fully connected operations; these relationships are weighted in the feature maps as the channel attention, allowing the network to focus on more important feature information in the channel dimension.

2.3.2. Spatial attention module

The spatial attention module (SA) [38] is commonly used to focus on the spatial location information of feature maps. First, two feature maps with the same resolution and number of channels are obtained through maximum pooling and average pooling, thus emphasizing the spatial information of the features. Then, the feature maps of the two channel dimensions are concatenated and used as the input for the 7×7 convolution. Finally, a spatial weight value between 0 and 1 is obtained with the sigmoid function; this value is weighted in the original feature map as the spatial attention, which enables the network to learn more important feature information in the spatial dimension.

2.4. Dilated convolution

The main idea of dilation convolution is to insert "holes" (zeros) between pixels in the convolution kernel to prevent the loss of contextual information during down-sampling, resulting in a larger-scale receptive field with richer, denser feature representations. Yu et al. [39] introduced dilation convolution into dense prediction and systematically aggregated multiscale contextual information without a decrease in the resolution, demonstrating that dilation convolution can play an active role in semantic segmentation. Nonetheless, simply overlaying dilation convolution in a network can lead to grid effects and the loss of a significant number of feature representations, thereby severely degrading the performance. To alleviate the grid problem, Wang et al. [40] developed a hybrid dilation convolution framework that effectively increases the perceptual field size and improves the segmentation performance for objects of different sizes.

3. Proposed methods

3.1. HADCNet architecture

Although the proposed network is based on the U-Net architecture, in contrast to the original U-Net, HADCNet uses hybrid attention modules in the encoder and decoder to refine the feature information while effectively balancing the semantic differences between different levels of features to obtain fine-grained feature representations. First, we introduce the encoder hybrid attention module, which uses densely connected features to capture rich contextual information and expands the receptive domain through hybrid dilated convolutions without reducing the resolution. Furthermore, the SE operation effectively integrates different scale feature representations at the peer level and refines the complex structural information of the features. Second, a hybrid attention module is introduced in the decoding stage to embed the spatial information of the underlying features in high-level features through an improved skip connection. In addition, the semantic information of the higher-level features in the channel dimension is assigned to the underlying features, and the effective fusion of these two embedded features enables the segmentation target to be accurately located, with the boundary details and refined structural information preserved. The decoder hybrid attention module integrates feature representations from different hierarchies across various levels in the decoder, refining the structural information of the features during upsampling while reducing the semantic gap between the encoder and the decoder to generate improved dense predictions that effectively represent the high interclass similarity and intraclass variability of the lesion features. The overall structure of HADCNet is shown in Fig. 2, with the dashed lines indicating improved skip connections, the different coloured rectangles representing different network operations, and the number at the bottom of each rectangle indicating the number of channels after that operation.

The pseudocode of HADCNet is described below.

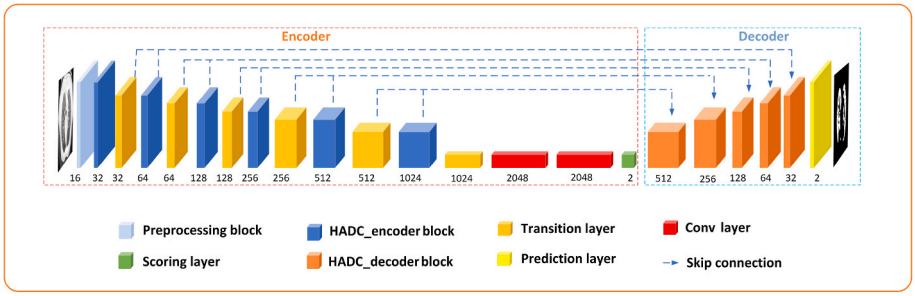

Fig. 2. The overall network structure of HADCNet.

The pseudocode of the proposed HADCNet

```
Image: Input image of the network;
While (stop condition is not met) do
    Converting image to the feature map F;
    Generating spatial attention using Eq.(1);
    Generating channel attention using Eq.(2);
    Obtaining code-fused attention using Eq.(3);
    If condition is met    //The transition layer in the network calculates attention
      Generating spatial attention;
    End if
    Passing feature maps through the network;
    If condition is not met    //Feature map has not been decoded
      Continue with steps 4, 5, 6, 7, 8, 9;
    Else
      Generating channel attention using Eq.(4);
      Generating spatial attention using Eq.(5);
      Obtaining the final fused attention using Eq.(6);
      If condition is not met    //The prediction layer has not yet processed the feature map
        Continue with steps 14, 15, 16;
      Else
        Obtaining the final output feature map;
      End if
    End if
  End while
```

3.2. Encoder stage of the proposed method

This stage consists of 1 pre-processing block, 6 encoder hybrid attention blocks (HADC_encoder module), 6 transition layers, 2 convolutional (Conv) layers, and 1 scoring layer. The pre-processing block is composed of a Conv layer, a batch normalization (BN) layer, and the rectified linear unit (ReLU) function. This combination forms the Conv3 × 3-BN-ReLU structure, which coarsely extracts the features of the input image, with the ReLU function suppressing the output of some neurons, effectively alleviating the overfitting problem, and the BN layer adjusts the semantic data distribution in the feature map in real time to enhance the robustness of the network. Each HADC_encoder module captures rich contextual semantic information by fusing semantic representations between feature maps through dense connections. Furthermore, the multiscale patterns in the encoder feature maps are obtained using hybrid dilation convolution, and the SE operation integrates the feature

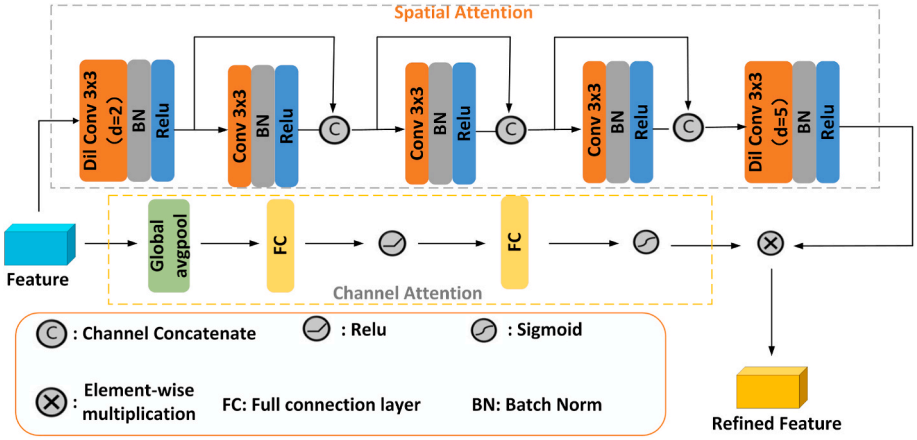

Fig. 3. The structure of the HADC_encoder module.

information across different scales at the peer level. The detailed structure of the HADC_encoder module is shown in Fig. 3, where Dil Conv 3 × 3(d = 2) and Dil Conv3 × 3(d = 5) represent the dilation convolution with expansion rates of 2 and 5, respectively, and the dense connection consists of the middle three Conv3 × 3-BN-ReLU structures. The transition layer has a Conv1 × 1-BN-ReLU structure, an average pooling layer, and a spatial attention (SA) layer, where the Conv1 × 1 layer is used to increase the nonlinearity of the network and the spatial attention value is calculated and saved before each pooling operation in the transition layer as one of the input values for the skip connection. The last transition layer passes through 2 convolutional layers and the scoring layer. The convolutional layers have convolutional kernel sizes of 7 × 7 and 3 × 3, while the output of the scoring layer has a channel dimension of 2.

The HADC_encoder module combines spatial attention and channel attention and uses the output of the pre-processing block or the previous transition layer as its input value. We use $F \in \mathbb{R}^{C\times H\times W}$ to denote the input feature map, $M_s(F) \in \mathbb{R}^{1\times H\times W}$ to denote the spatial attention, $M_c(F) \in \mathbb{R}^{C\times 1\times 1}$ to denote the channel attention, and $M(F) \in \mathbb{R}^{C\times H\times W}$ to denote the final fused attention. The HADC_encoder function is thus computed as follows:

$$M_s(F) = f_{3\times3}^{d=5}(H_l([x_{l-1}, x_{l-2}, ..., x_0])), \left(x_0 = f_{3\times3}^{d=2}(F), l = 4\right) \quad (1)$$

We design the spatial attention $M_s(F)$ with dilated convolution, combining the 3 × 3 convolution kernel and dilated convolutions with expansion rates of [1,2,5], replacing the dilated convolution with an expansion rate of 1 with a densely connected block, as shown in Equation (1), where $f_{3\times3}^{d=2}$ and $f_{3\times3}^{d=5}$ represent dilated convolutions with expansion rates of 2 and 5, respectively. $H_l(.)$ is a densely connected block, which includes 3 × 3 three convolution layers, H_l represents the Conv3 × 3-BN-ReLU structure, and the value after combining the three convolution layers is the final output of the dense connection. The equivalent kernel sizes for the dilated convolution with a convolution kernel size of 3 × 3 and dilation rates of [1,2,5] are 3, 5, and 11, respectively. According to the definition of the receptive domain, a dilated convolution with this stacked structure has a 17 × 17 receptive domain and can thus capture the global information. Furthermore, the combined hybrid dilation convolution and dense connection captures rich multiscale semantic information, which solves the problem of fixed-size convolutional layers having a local field of perception with only a single-grain size that responds to diverse feature maps.

$$M_c(F) = \sigma\left(MLP\left(P_{avg}(F)\right)\right) = \sigma\left(W_1\left(\delta\left(W_0\left(P_{avg}(F)\right)\right)\right)\right) \quad (2)$$

where σ, P_{avg} and δ represent the sigmoid function, global average pooling, and the ReLU operation, respectively, and $W_1 \in \mathbb{R}^{C\times C/r}$ and $W_0 \in \mathbb{R}^{C/r\times C}$ represent the incremental and decremental weight parameters in the multilayer perceptron (MLP), respectively, with the weight decay rate r taken as 16.

$$M(F) = M_s(F) \circ M_c(F) + F \quad (3)$$

where ∘ and + represent elementwise multiplication and elementwise addition, respectively, and $M_s(F) \circ M_c(F)$ is equivalent to integrating multiscale contextual semantic information at the peer level in the channel dimension to refine the encoder feature information by supressing irrelevant feature representations. In addition, the fused attention feature map is subjected to an elementwise addition operation with the initial input feature map to prevent gradient disappearance.

3.3. Decoder stage of the proposed method

This stage includes 5 decoder hybrid attention blocks (HADC_decoder module) and 1 prediction layer. Three feature maps are fed into the HADC_decoder module, and the source of the guiding information is the output of the scoring layer in the encoder or the previous

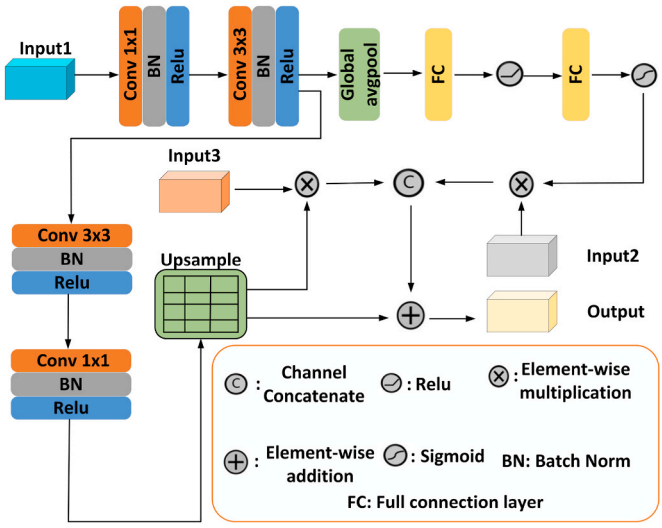

Fig. 4. The structure of the HADC_decoder module.

HADC_decoder module. Feature 1 is the output of the corresponding HADC_encoder module in the encoder, and feature 2 is the spatial attention value corresponding to the transition layer, which is calculated and saved in the encoder. In the HADC_decoder module, the guiding information first passes through the Conv1 × 1-BN-ReLU-Conv3 × 3-BN-ReLU structure to introduce more nonlinearity; then, the two steps of the improved skip connection are performed. The first step is to obtain the channel semantic information of the guiding signal using the SE operation; however, in this case, the value is not multiplied by the original input feature map. Instead, only a weight between 0 and 1 is retained and multiplied with feature 1, which is equivalent to embedding the channel position information with rich high-level features into low-level features to integrate multiscale contextual information. Then, feature 2, which is the spatial attention value that preserves the low-level features, is multiplied by the guiding information after the Conv3 × 3-BN-ReLU-Conv1 × 1-BN-ReLU and up-sampling operations. This step corresponds to assigning the spatial semantic information of the low-level features to the high-level features. Finally, the low-level features containing the semantic information of the high-level features obtained during the first step are fused with the high-level features containing the spatial information of the low-level features obtained during the second step in the channel dimension, which is equivalent to integrating semantic information of different scales across levels to generate effective fused features. The fused feature representations are then elementwise added with the guiding information after up-sampling. The concrete structure of the HADC_decoder module is illustrated in Fig. 4, where input 1, input 2, input 3 and output represent the guiding signal, feature 1, feature 2 and final fused feature, respectively. The prediction layer samples the output of the final HADC_decoder module into a feature map with the same resolution as the original image with 2 channels (indicating the two categories, namely, infected and noninfected) and obtains the final segmentation result by classifying all pixel positions in the feature map.

The HADC_decoder module merges the high-level and low-level features in the channel and spatial dimensions, thus capturing multiscale contextual information across levels and refining the feature map during up-sampling. We represent the guiding information in the HADC_decoder module as $G \in \mathbb{R}^{C_g\times H_g\times W_g}$, feature 1 as $F_1 \in \mathbb{R}^{C_{f_1}\times H_{f_1}\times W_{f_1}}$, feature 2 as $F_2 \in \mathbb{R}^{C_{f_2}\times H_{f_2}\times W_{f_2}}$, the low-level features with rich high-level semantic information obtained in the first step as $Z_c \in \mathbb{R}^{C\times H\times W}$, the high-level features with low-level feature spatial information obtained in the second step as $Z_s \in \mathbb{R}^{C\times H\times W}$, and the fused features as $Z \in \mathbb{R}^{C\times H\times W}$. The entire calculation process of the HADC_decoder block is as follows:

$$Z_c = \sigma\left(\mathrm{MLP}\left(P_{avg}(\mathrm{H}(\mathrm{G}))\right)\right)\circ F_1 = \sigma\left(W_{C_{f_1}}\left(\delta\left(W_{C_g/r}\left(P_{avg}(\mathrm{H}(\mathrm{G}))\right)\right)\right)\right)\circ F_1 \quad (4)$$

where H(.), ∘, σ, P_{avg}, and δ represent the Conv1 × 1-BN-ReLU-Conv3 × 3-BN-ReLU structure, elementwise multiplication operation, sigmoid function, global average pooling, and ReLU operation, respectively, and and are the weight parameters, with r taken as 16. In contrast to the channel attention in the HADC_encoder module, the input size of the multilayer perceptron (MLP) is larger than the output size, which further suppresses irrelevant feature representations. The final operation is equivalent to adding the semantic information encoded by G at a deep level in the channel dimension to.

$$Z_s = P_{up}(H_2(H_1(G)))\circ F_2 \quad (5)$$

where H_1 and H_2 denote the Conv1 × 1-BN-ReLU-Conv3 × 3-BN-ReLU and Conv3 × 3-BN-ReLU-Conv1 × 1-BN-ReLU operations, respectively, resulting in a bottleneck-like structure that extracts the feature map semantic information while introducing additional nonlinearity and reducing the number of parameters. P_{up} denotes the use of transposed convolution for up-sampling, while the ∘ operation embeds the spatial attention values saved in F_2 into the feature representation guided by G.

$$Z = [Z_c, Z_s] + \alpha F \quad (6)$$

where α is a trainable parameter that begins at 0 and assigns greater weights, $[\cdot]$ denotes the concatenation operation, + represents elementwise addition, F denotes G after up-sampling, and Z is the effective fusion feature map at different levels. In addition, the combination of long-short residual connections in the HADC_decoder module prevents performance degradation caused by long residual connections, effectively solving the problem of gradient disappearance.

4. Experimental description, results and discussion

4.1. Dataset description and hardware environment

In this study, we conducted experiments with four datasets, which are described as follows.

1) *Dataset-I*: This dataset is a publicly available COVID-19 CT dataset [41] that includes 100 axial lung CT slices, each with detailed multiclass annotations by radiologists, including GGOs, consolidation lesions, and pleural effusions, and a slice resolution of 512 × 512 pixels.
2) *Dataset-*II*:* This dataset contains 829 CT slices [42], with 373 COVID-19-positive slices segmented by radiologists using two labels, GGOs and consolidation lesions, and a raw image resolution of 630 × 630 pixels.
3) *Dataset-III:* This dataset is derived from a database with 4695 pneumonia CT images [43], of which 752 annotated CT slices from 150 COVID-19 patients were used as the third COVID-19 lesion segmentation dataset.
4) *Dataset IV:* Ma et al. [44] published 20 COVID-19 CT scans from the Coronacases Initiative and Radiopaedia. This dataset contains a total of 1844 annotated COVID-19 infection axial images with a resolution of 512 × 512 pixels.

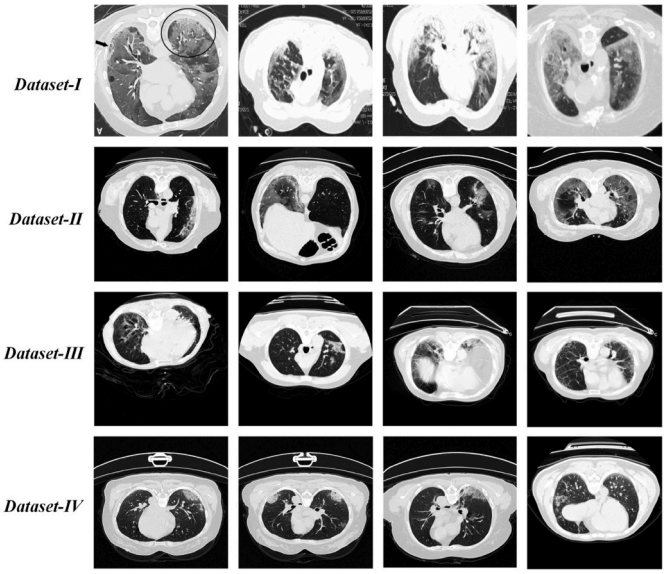

Fig. 5. Example images from four COVID-19 infection benchmark datasets.

Dataset-I contains the largest proportion of COVID-19 infected areas per image, but the background of the images is cluttered, and the contrast in the images varies significantly; thus, the image quality is the worst among the four datasets. *Dataset-II* has the best image quality but a smaller proportion of COVID-19-infected areas. *Dataset-III* has a poorer image quality than *Dataset-II* and a smaller proportion of COVID-19 infected areas than *Dataset-II*. *Dataset-IV* has a better image quality than *Dataset-I* but the smallest proportion of COVID-19 infected areas. These results are illustrated in Fig. 5, which shows example images from the four datasets.

For the four COVID-19 segmentation datasets, we chose to discard all slices that did not contain infections and kept only the annotated images; however, we did not differentiate the type of lesion and processed the slices as binary type images. The data expansion operation was not performed on the datasets, as this operation may cause the experimental data to leak during training and testing, resulting in an inflated model segmentation performance. Additionally, we refrained from using pretrained models during our experiments and trained the proposed model from scratch using the experimental datasets. The relatively small size of the segmentation datasets and the extremely unbalanced distribution of COVID-19 infections had a significant impact on the smaller datasets, leading to a lack of robustness in the model training results. To obtain the most accurate and objective assessment results, we performed 5-fold cross-validations on each of the four datasets, with 80% training and 20% testing (unseen) sets, and 20% of the training data were used for validation.

We used PyTorch 3.6.8 with the following specific parameters as our deep learning framework:

CPU: Intel (R) Core (TM) i9-7900X CPU @ 3.30 GHz 3.31 GHz.
GPU: NVIDIA GTX 1080ti GPU.
Memory: 64.0 GB.

4.2. Loss function

During the training process of the model, the cross-entropy function and Dice score are combined as the loss function, and the specific process for calculating the cross-entropy function is shown in Equation (7).

$$L_c(u,\hat{u}) = -\frac{1}{N}\left(\sum_{i=1}^{N}(1-w_i)u_i\log u_i + w_i(1-\hat{u}\log(1-\hat{u}))\right) \quad (7)$$

where u_i denotes the ground truth of voxel i, w_i is the weight, and $\hat{u}_i$ and $1-\hat{u}_i$ represent the probability that voxel i belongs to the background or the target, respectively.

The Dice score is a commonly used metric for medical segmentation that is calculated as follows:

$$L_d(u,\hat{u}) = 1 - \frac{2\sum_{i=1}^{N}u_i\hat{u}_i + \theta}{\sum_{i=1}^{N}(u_i+\hat{u}_i) + \theta} \quad (8)$$

where θ, which ensures a nonnegative value and smooths the loss as well as gradient, was set to 1e-4.

The final weighted loss function is calculated as follows:

$$loss = (1 - w) \times L_d + w \times L_c \tag{9}$$

where w is a trainable parameter that is learned from 0 to the best-assigned weight, and the weighted loss function becomes the Dice score when $w = 0$. Because the gradient of the cross-entropy loss function for weights of the last layer is only proportional to the difference between the predicted and true values, and the convergence rate is faster at this time due to the back propagation, the rate of updating the entire weight matrix is improved, which improves the training efficiency of the network. However, as the region of lesions in the datasets used in this study is much smaller than the background region, the cross-entropy loss function at this point will cause the model to be heavily biased towards the background region, leading to a decrease in the segmentation performance. To alleviate this problem, we introduced the dice loss function to balance the problem of positive and negative sample disequilibrium. The combination of the cross-entropy loss and the dice loss function can exploit the advantages of both, effectively equalizing the sample imbalance problem while accelerating the training speed of the model. Further, the use of the parameter w enables the network to iterate towards the optimal weight ratio of the two.

4.3. Evaluation metrics

We used five metrics to evaluate the segmentation results, including the Dice similarity coefficient (DSC), Jaccard similarity coefficient (JS), accuracy (Acc), sensitivity (Sen), and specificity (Spec). Acc indicates the proportion of correctly segmented samples to the total number of samples and is used to assess the pixel-level classification accuracy. DSC and JS represent the overlap ratio between the segmentation results and the ground truth, reflecting the variability between the segmentation results and the marker values; these metrics are used to estimate the overall performance of the segmentation task. Sen and Spec are used to measure the ability of the algorithm to segment regions of interest and background regions. These evaluation metrics all have values in the range of 0–1, with larger values indicating more desirable segmentation results.

4.4. Experimental settings

During the experiments, no image pre-processing was performed, and the original lung CT images were used as the input to train the HADCNet model. For the model to fully learn the features of COVID-19 infection, we performed a total of 100 epochs of training on *Dataset-I* and *Dataset-II* in the experiments; the initial learning rate was set to 0.0001, and the learning rate decreased by 1/10 every 20 epochs. Due to the larger volume of data in *Dataset-III* and *Dataset-IV*, we terminated the training early in the training phase to prevent overfitting; thus, we conducted a total of 40 and 60 epochs of training on *Dataset-III* and *Dataset-IV*, respectively, which allowed the model to better learn the feature information. The mini-batch size was set to 6 for each iteration, and the combined loss function was used to train the network by calculating the error between the predicted and labelled classifications for all pixels in the input image within the mini-batch size, with a loss value closer to 0 indicating better training results.

Table 1
Comparison of the details of each method and the initial values of the hyperparameters – indicates that the method has no relevant structure compared to the proposed model and therefore the corresponding hyperparameters are absent.

Model	Detail	Initial value(r)	Initial value(α)	Initial value(w)
IAEUnet	improved U-Net with HADC_encoder module	16	0	0
IDDUnet	improved DenseUnet with HADC_decoder module	16	0	0
IADUnet	improved DenseUnet with HADC_encoder module lacking dilated convolution	16	0	0
IDCUnet	improved U-Net with normal convolutional layer instead of dense connection in HADC_encoder module	16	0	0
DenseUnet	improved 2-D UNet	–	–	0
HADCNet	**proposed model**	**16**	**0**	**0**

To demonstrate the effectiveness of the proposed structure, we conducted ablation experiments on the four datasets using the following methods: IAEUnet (improved 2D U-Net with an HADC_encoder module), IDDUnet (improved DenseUnet with an HADC_decoder module), IADUnet (improved DenseUnet, in which the encoder includes the HADC_encoder module without dilated convolution), IDCUnet (improved U-Net, in which the encoder contains the HADC_encoder module with the dense connection replaced with a normal convolutional layer), and DenseUnet (improved 2-D UNet) [17]. In the experiments, IAEUnet, IDDUnet, IADUnet, IDCUnet, DenseUnet, and HADCNet were trained in the same experimental environment, and the segmentation results of each network were analysed and compared. Table 1 summarizes the details of each method and the use of the three hyperparameters in the proposed model for each method. Among them, the hyperparameter r is the weight parameter used in the HADC_encoder module and HADC_decoder module to suppress irrelevant feature representations. During the experiments, an r value between 8 and 24 had little effect on the results, with the best results obtained with a value of 16, while the segmentation performance decreased significantly outside this range, and adjusting the r value had only a slightly positive impact on the experimental results, so the default value of 16 was used for r. The hyperparameters α and w, which are trainable parameters in the model and the combined loss function, respectively, were both initialized to 0 during the experiments, with the optimal value sought by iteration of the relevant model.

4.5. Experimental results

The *Dataset-I* section in Fig. 6 shows sample segmentation images of the HADCNet model on *Dataset-I* in the fivefold cross-validation, where Fold1 to Fold5 represent the results of five fivefold cross-validation experiments, and columns a, b, c, and d represent the original image, the corresponding ground truth (GT) image, the mask image for HADCNet segmentation, and the resultant HADCNet segmentation image, respectively. The difference in the position between the mask image after HADCNet segmentation and the corresponding GT image is marked in green and red. The green areas indicate false positive pixels, which are over-segmented areas, while the red areas represent false negative pixels, which are under-segmented areas. The *Dataset-II*, *Dataset-III*, and *Dataset-IV* sections in Fig. 6 show example images of the HADCNet segmentation results for the fivefold cross-validation on *Dataset-II*, *Dataset-III*, and *Dataset-IV*, respectively, with the same meaning as the *Dataset-I* section in Fig. 6. We found that HADCNet could segment the entire lesion target, generate clear boundaries, and effectively handle complex lesion regions, especially for small- and medium-sized lesions. In addition, the segmentation results highlight the main locations and detailed information about the lesions, indicating the ability of HADCNet to competently identify diverse lesions in a balanced manner.

Table 2 summarizes the values of each evaluation metric for five fivefold cross-validation experiments on the four benchmark datasets. Although *Dataset-I* had the smallest amount of data and the worst image quality, this dataset had the largest and most balanced distribution of infected regions and achieved the second-highest DSC value (0.792). Although *Dataset-II* had more data than *Dataset-I*, the distribution of

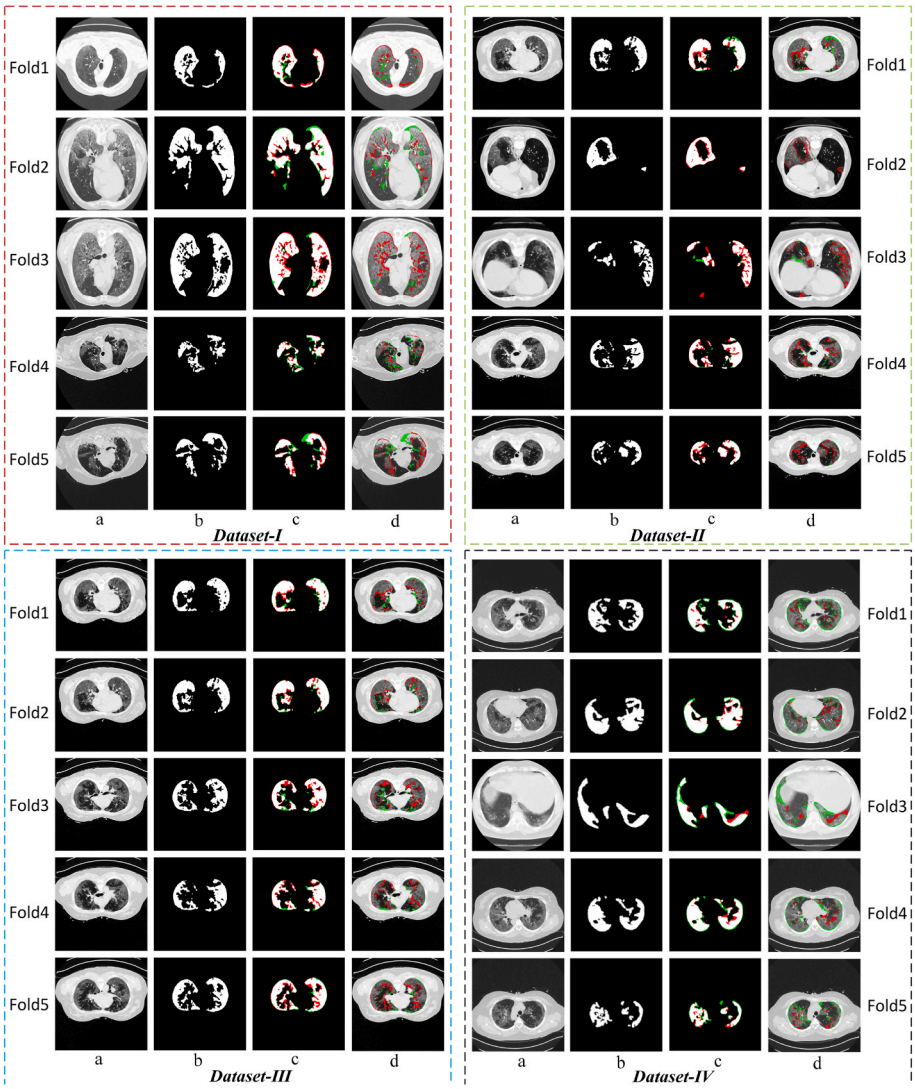

Fig. 6. Example images of the fivefold cross-validation segmentation results on the four benchmark datasets.

infected regions in the lung was more complex, and the segmentation results were not significantly different from those of *Dataset-I*. Except for a significant drop in the Sen value, the segmentation results of *Dataset-III* were similar to those of *Dataset-I* and *Dataset-II*. Although the data volume of *Dataset-III* was larger than that of the other two datasets, this dataset contained fewer COVID-19 infection features, resulting in difficulties for the model in learning more lung infection features and thereby limiting the segmentation performance of the model. The overall assessment on *Dataset-IV* was worse than that of the other three datasets. Although *Dataset-IV* had the largest amount of data among the four datasets, the quality of the images was poor, and the dataset contained the most unbalanced and smallest distribution of lung infections, which limited the model segmentation performance. In addition, the values of the evaluation metrics for all five fivefold cross-validation experiments fluctuated within a small range for each of the four datasets, indicating that the robustness of the model was good on these four datasets.

An analysis of the results shown in Fig. 6 and Table 2 demonstrates that the proposed HADCNet model is robust on the four benchmark datasets. To further evaluate the COVID-19 infection segmentation performance of HADCNet, we compared the HADCNet segmentation results on the four COVID-19 infection datasets with the results of several state-of-the-art segmentation algorithms. The findings are summarized in Table 3, which includes the evaluation results from previous studies and those of HADCNet on the four benchmark datasets. On *Dataset I*, HADCNet obtained the highest values for each evaluation metric, with DSC values that were 1.9% and 1.3% better than those of the MiniSeg [46] and Wang et al. [52] models, respectively, while the Sen value was nearly 3% higher than that of Zhang et al. [50]. Similarly, HADCNet achieved the highest values for each evaluation metric on *Dataset-II*; the DSC value was almost 3.2% higher than that of SD-UNet [54], and the Sen value was significantly higher by almost 14% compared to that of U-Net [13], which had the second-best performance and a Sen value of 0.772. Although the highest JS value (0.652) was obtained by the Wang et al. [52] model on *Dataset-III* and the second-highest value (0.646) was obtained by HADCNet, HADCNet achieved the highest values for the other evaluation metrics, with a 2.7% improvement in the DSC value compared to the model of Wang et al. [52]. On *Dataset-IV*, the HADCNet model achieved the highest values for each evaluation metric. The experimental results show that the COVID-19 infection segmentation performance of HADCNet was generally better than that of the other segmentation algorithms. Furthermore, we found that most of the segmentation algorithms used for comparison were based on the conventional encoder-decoder structure, while HADCNet, which depends on a dual hybrid attention strategy to efficiently extract fine-grained details and semantic features at both the peer and cross levels, achieved a considerably better segmentation performance, demonstrating the improved segmentation

Table 2
Evaluation results of the fivefold cross-validation on the four benchmark datasets.

Dataset	Model	DSC	JS	Acc	Sen	Spec
Dataset-I	Fold1	0.789	0.650	0.970	0.849	0.987
	Fold2	0.795	0.659	0.969	0.878	0.982
	Fold3	0.792	0.655	0.970	0.883	0.986
	Fold4	0.789	0.650	0.970	0.865	0.988
	Fold5	0.792	0.656	0.969	0.881	0.981
	Avg	**0.792 ± 0.03**	**0.654 ± 0.04**	**0.970 ± 0.01**	**0.871 ± 0.03**	**0.985 ± 0.03**
Dataset-II	Fold1	0.790	0.654	0.991	0.904	0.993
	Fold2	0.791	0.653	0.991	0.914	0.992
	Fold3	0.793	0.674	0.994	0.909	0.995
	Fold4	0.809	0.669	0.990	0.916	0.996
	Fold5	0.799	0.671	0.993	0.918	0.995
	Avg	**0.796 ± 0.03**	**0.664 ± 0.05**	**0.991 ± 0.03**	**0.912 ± 0.06**	**0.994 ± 0.02**
Dataset-III	Fold1	0.782	0.642	0.994	0.745	0.997
	Fold2	0.777	0.636	0.993	0.725	0.998
	Fold3	0.791	0.655	0.993	0.791	0.997
	Fold4	0.786	0.648	0.994	0.760	0.997
	Fold5	0.785	0.646	0.994	0.732	0.998
	Avg	**0.785 ± 0.02**	**0.646 ± 0.02**	**0.993 ± 0.01**	**0.751 ± 0.31**	**0.997 ± 0.01**
Dataset-IV	Fold1	0.740	0.588	0.988	0.681	0.997
	Fold2	0.716	0.558	0.987	0.679	0.997
	Fold3	0.736	0.582	0.987	0.710	0.995
	Fold4	0.720	0.562	0.987	0.694	0.997
	Fold5	0.702	0.541	0.986	0.708	0.997
	Avg	**0.723 ± 0.18**	**0.566 ± 0.14**	**0.987 ± 0.01**	**0.694 ± 0.21**	**0.997 ± 0.02**

performance of the proposed model. Due to the limitations of the hardware environment, the training data in our experiments could not be extended in the same way as in other experiments.

To further confirm the performance and effectiveness of the proposed network, we compared the number of parameters and the computational cost of HADCNet on *Dataset-I* with those of the other state-of-the-art network models. *Dataset-I* contains the largest proportion of COVID-19-infected regions, but the image background is cluttered, and the contrast varies significantly. Thus, although the image quality of this dataset is the worst, it is the most representative among the four datasets given the possible duplicate comparison issues; therefore, only *Dataset-I* is discussed here. We selected DSC, JS, and Sen as the main segmentation effectiveness evaluation metrics, and the comparison results are shown in Table 4. The HADCNet model had the second smallest number of parameters (20.7 M), with the smallest number of parameters (8.95 M) obtained by the D2A U-Net + VGG [49] model; however, the DSC value of HADCNet was significantly improved by 8.7% compared to the DSC value of the D2A U-Net + VGG model. Compared with the model proposed by Wang et al. [52], which achieved the best segmentation performance among the other networks, HADCNet had a lower number of parameters and a lower computational cost. Despite the relative complexity of the proposed network structure, the number of parameters and the computational cost are smaller than those of many state-of-the-art methods, mainly due to the use of 1×1 convolutions to extract and integrate the features, as well as the use of hybrid dilated convolution in the proposed encoder hybrid attention module, which expands the perceptual field without increasing the number of parameters. Thus, the experimental results demonstrate the effectiveness of the proposed structure.

Table 3
Comparison results on the four benchmark datasets, – indicates that the data were not provided by the author of the corresponding paper.

Dataset	Model	DSC	JS	Acc	Sen	Spec
Dataset-I	Semi-Inf-Net [45]	0.739	–	–	0.725	0.960
	MiniSeg [46]	0.773	–	–	0.836	0.974
	CB-PL [47]	0.730	–	–	0.820	0.920
	Yu et al. [48]	0.779	–	–	0.791	0.983
	D2A U-Net + VGG [49]	0.705	–	0.968	0.663	–
	D2A U-Net + ResNet [49]	0.730	–	0.969	0.707	0.955
	Zhang et al. [50]	0.765	–	–	0.839	–
	CE-Net [51]	0.742	0.605	0.941	–	**0.985**
	Wang et al. [52]	0.779	0.648	0.944	–	
	Proposed HADCNet	**0.792**	**0.654**	**0.970**	**0.871**	
Dataset-II	U-Net [13]	0.737	0.617	0.978	0.773	0.992
	Gated-UNet [53]	0.738	0.615	0.982	0.693	0.994
	UNet++ [22]	0.759	0.634	0.981	0.739	0.993
	SD-UNet [54]	0.764	0.643	0.981	0.772	0.991
	Proposed HADCNet	**0.796**	**0.664**	**0.991**	**0.912**	**0.994**
Dataset-III	U-Net [13]	0.431	0.312	0.988	–	–
	U-Net++ [22]	0.579	0.459	0.991	–	–
	MR-UNET [55]	0.703	0.586	0.993	–	–
	Gated-UNet [53]	0.616	0.498	0.993	–	–
	CE-Net [51]	0.706	0.595	0.994	–	–
	Inf-Net [41]	0.650	0.520	0.992	–	–
	Wang et al. [52]	0.758	**0.652**	0.993	–	–
	Proposed HADCNet	**0.785**	0.646	**0.993**	**0.751**	**0.997**
Dataset-IV	Zhang et al. [50]	0.703	–	–	–	–
	3D nnUNet [56]	0.673	–	–	–	–
	Gated-UNet [53]	0.623	–	–	0.658	0.926
	Inf-Net [41]	0.682	–	–	0.692	0.943
	U-Net++ [22]	0.581	–	–	0.672	0.902
	Proposed HADCNet	**0.723**	**0.566**	**0.987**	**0.694**	**0.997**

Table 4
Comparison results on *Dataset*-I, – indicates that the data were not provided by the author of the corresponding paper.

Dataset	Model	Param.	FLOPs	DSC	JS	Sen
Dataset-I	Semi-Inf-Net [45]	33.12 M	13.92 G	0.739	–	0.725
	D2A U-Net + VGG [49]	**8.95 M**	53.19 G	0.705	–	0.663
	D2A U-Net + ResNet [49]	90.05 M	149.97 G	0.730	–	0.707
	CE-Net [51]	146.35 M	35.71 G	0.742	0.605	–
	Wang et al. [52]	39.75 M	46.71 G	0.779	0.648	–
	Proposed HADCNet	20.7 M	**13.2 G**	**0.792**	**0.654**	**0.871**

4.6. Ablation experiment results

A series of ablation experiments were conducted with the four benchmark datasets to verify the validity of the HADCNet structure. We analysed the segmentation results of the other networks (IAEUnet, IDDUnet, IADUnet, IDCUnet, and DenseUnet, as specified in Section 4.4) and of HADCNet in detail.

Fig. 7 shows example images of the segmentation results of each network on the four benchmark datasets in the ablation experiments, with columns 1 and 2 representing the original image and the corresponding GT image, respectively, while columns 2–7 indicate the segmentation results of the corresponding network. Each row shows the segmentation results of the network on the corresponding dataset. The difference in the position between the masked image after correlation

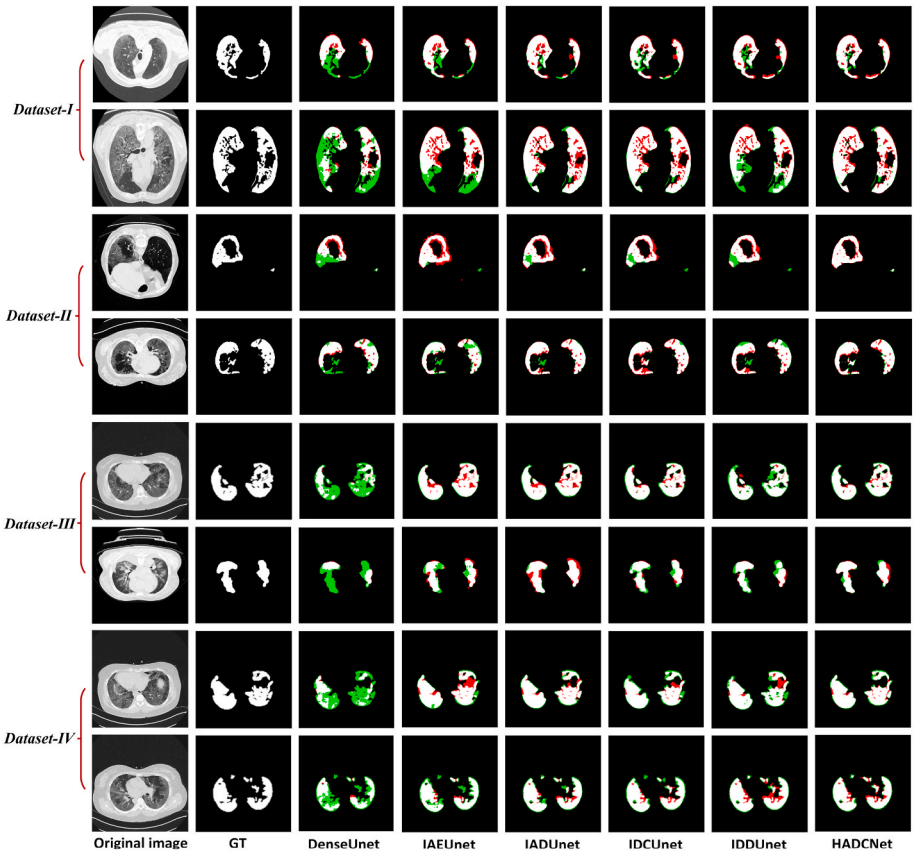

Fig. 7. Example images of the segmentation results of each network on the four benchmark datasets.

Table 5
Evaluation of the segmentation results of each network on the four benchmark datasets.

Dataset	Model	DSC	JS	Acc	Sen	Spec
Dataset-I	IAEUnet	0.728	0.572	0.963	0.654	0.980
	IDDUnet	0.741	0.589	0.962	0.719	0.982
	IADUnet	0.768	0.624	0.967	0.731	0.976
	IDCUnet	0.767	0.622	0.967	0.732	0.981
	DenseUnet	0.682	0.518	0.960	0.574	0.991
	HADCNet	**0.792**	**0.654**	**0.970**	**0.871**	**0.985**
Dataset-II	IAEUnet	0.739	0.586	0.979	0.655	0.991
	IDDUnet	0.718	0.560	0.984	0.595	0.991
	IADUnet	0.769	0.625	0.988	0.714	0.985
	IDCUnet	0.763	0.617	0.987	0.679	0.984
	DenseUnet	0.604	0.433	0.990	0.478	0.945
	HADCNet	**0.796**	**0.664**	**0.991**	**0.912**	**0.994**
Dataset-III	IAEUnet	0.730	0.575	0.990	0.735	0.994
	IDDUnet	0.740	0.594	0.989	0.739	0.991
	IADUnet	0.768	0.623	0.990	0.746	0.993
	IDCUnet	0.755	0.606	0.989	0.689	0.990
	DenseUnet	0.672	0.506	0.983	0.964	0.983
	HADCNet	**0.785**	**0.646**	**0.993**	**0.751**	**0.997**
Dataset-IV	IAEUnet	0.646	0.477	0.984	0.525	0.993
	IDDUnet	0.670	0.504	0.984	0.594	0.995
	IADUnet	0.698	0.536	0.986	0.579	0.991
	IDCUnet	0.655	0.487	0.984	0.532	0.997
	DenseUnet	0.533	0.463	0.980	0.409	0.932
	HADCNet	**0.723**	**0.566**	**0.987**	**0.694**	**0.997**

network segmentation and the corresponding GT image is marked in green and red, with the green region indicating the over-segmented region and the red region indicating the under-segmented region. The HADCNet segmentation result is significantly better than that of the other networks. HADCNet segmented complex and diverse lung infection regions more effectively than the other networks, with more sensitivity to the segmentation of ambiguous or microscopic lesions than the other models. This result is due to the dual mixed attention module, which enabled the model to detect infected lungs more accurately while reducing false positives. The IADUnet and IDCUnet segmentation results are better than those of the other three comparison networks but worse than those of the HADCNet model, suggesting that the introduction of dense connections and hybrid dilated convolution can help to segment infected lesions of different sizes. We also note that the improvement in the segmentation effectiveness of IAEUnet and IDDUnet over DenseUnet is greater on *Dataset-I* and *Dataset-II* than on *Dataset-III* and *Dataset-IV*, suggesting that when a single hybrid attention module is applied, the results of some lung infection images are improved when more data are used.

Table 5 summarizes the evaluation results of the five comparison networks and HADCNet on the four benchmark datasets. HADCNet achieved the best values for each evaluation metric on the four benchmark datasets; IADUnet and IDCUnet obtained the second-best evaluation results; IAEUnet and IDDUnet, which have only one hybrid attention module, obtained the second-worst evaluation results; and DenseUnet, which does not employ a hybrid attention strategy, obtained the worst evaluation results. The experimental results show that the proposed dual hybrid attention module can effectively highlight potentially infected regions and improve the segmentation performance, resulting in better evaluation metric values. In HADCNet, the encoder hybrid attention module integrates feature information across different scales at the peer level to refine the feature map in the encoding stage, while the decoder hybrid attention module embeds high-level feature information into low-level features that integrate multi-scale contexts through an improved skip connection and assigns the spatial information of low-level features to high-level features, reducing the semantic

gap between feature maps at different levels in the encoding-decoding stage and producing a richer feature representation to effectively segment lesion regions. In addition, the use of the dual hybrid attention module improved the segmentation performance of the network more than the use of either the encoder attention module or the decoder attention module alone.

4.7. Discussion

CT screening is an important tool for diagnosing COVID-19 in the face of large-scale COVID-19 infections and possible viral variants. The proposed method can effectively locate the position and structural information of a lesion to assist clinicians in dynamically assessing the severity of an infection based on inflammatory changes due to targeted therapeutic measures. Thus, the proposed method could play an important role in the diagnosis, treatment, and prognostic assessment of COVID-19 cases. Despite the excellent performance of the HADCNet model on several public datasets, some issues arose. First, based on an analysis of the experimental results, the edge segmentation of the more complicated lesion regions in the four public datasets was relatively poor, with more serious over-segmentation or under-segmentation problems occurring to varying degrees, indicating that capability of the proposed model to segment complex edges is still lacking, and an edge focus-based mechanism needs to be introduced to improve the segmentation of edge regions. In addition, when the encoder hybrid attention module or decoder hybrid attention module was used alone, the effectiveness was only enhanced when the amount of data was larger, so the use of the hybrid attention module alone to effectively enhance the segmentation performance in a small amount of data is an that needs to be investigated. Second, due to the lack of finely labelled datasets, extending the proposed segmentation model to other finer-scale lesions that have been observed in COVID-19 infections, such as fine reticular opacities, subpleural parenchymal bands, fibrous streaks, and diffuse distributions, is difficult. To account for dataset limitations, we can introduce unsupervised or semi-supervised techniques, which use a large amount of unlabelled data to train the network, into the proposed model to improve the dense prediction performance. Furthermore, because non-COVID-19 infected lesions have features and textures that are similar to COVID-19 infected lesions, we can perform migration learning by combining the multi-lesion manifestations of non-COVID-19 infection features with COVID-19 infection features to further enhance the generalizability and robustness of the network. Then, because we did not differentiate between COVID-19 lesion types when segmenting inflammatory lung regions and performed only single-class segmentation, which may limit the potential application of the model in clinical diagnosis and treatment processes, multiclass segmentation should be implemented to enhance the performance of the model. Finally, to reduce the training time and computational cost of the model, we used a 2D network and 2D images for training in our experiments; however, because the training time and the number of parameters are still large, a lightweight COVID-19 lesion segmentation algorithm should be designed to ensure that the proposed model is useful in clinical practice in the future.

5. Conclusion

This paper presents a deep learning-based COVID-19 infection segmentation algorithm (HADCNet) with an encoder-decoder architecture. HADCNet refines the feature map and improves the segmentation performance through a dual hybrid attention strategy with encoder and decoder hybrid attention modules. The encoder hybrid attention module captures the rich semantic information of lesion features by using hybrid dilated convolutions, dense connections, and the SE operation to integrate multiscale contextual dependencies at the peer level. The decoder hybrid attention module introduces an improved skip connection to embed the semantic information of the high-level features in the low-level features, while the spatial information of the low-level features is embedded in the high-level features, thus refining the feature map during up-sampling. Finally, the two embedded features are combined in the channel dimension to form an effective fused feature that integrates the contextual semantic information of the feature map across levels to accurately obtain the location and structural information of the lesion, enhancing the ability of the HADCNet model to discriminate COVID-19 lesions. Extensive experiments on four public COVID-19 benchmark datasets demonstrated the generalizability and robustness of the HADCNet model. In future work, we will aim to further improve the framework by addressing existing issues and improving the flexibility of the model in segmenting COVID-19 infections.

Declaration of competing interest

The work described has not been submitted elsewhere for publication,in whole or in part, and all the authors listed have approved the manuscript that is enclosed.

Acknowledgements

This work was supported by the following foundations: the National Natural Science Foundation of China (grant no.61762067), the Natural Science Foundation of Jiangxi Province (grant no.20202BABL202029), and Basic study on public projects of Zhejiang Province (grant no. LGF18H030011).

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, et al., Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases[J], Radiology (2020), 200642.
[3] S.K. Zhou, H. Greenspan, C. Davatzikos, et al., A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises[J], Proc. IEEE 109 (5) (2021) 820–838.
[4] J. Zhang, Y. Xie, G. Pang, et al., Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection[J], IEEE Trans. Med. Imag. 40 (3) (2020) 879–890.
[5] Y. Chen, X.H. Yang, Z. Wei, et al., Generative adversarial networks in medical image augmentation: a review[J], Comput. Biol. Med. (2022), 105382.
[6] H. Su, D. Zhao, F. Yu, et al., Horizontal and vertical search artificial bee colony for image segmentation of COVID-19 X-ray images[J], Comput. Biol. Med. 142 (2022), 105181.
[7] L. Liu, D. Zhao, F. Yu, et al., Performance optimization of differential evolution with slime mould algorithm for multilevel breast cancer image segmentation[J], Comput. Biol. Med. 138 (2021), 104910.
[8] Q. Zhang, Z. Wang, A.A. Heidari, et al., Gaussian barebone salp swarm algorithm with stochastic fractal search for medical image segmentation: a COVID-19 case study[J], Comput. Biol. Med. 139 (2021), 104941.
[9] L. Liu, D. Zhao, F. Yu, et al., Ant colony optimization with Cauchy and greedy Levy mutations for multilevel COVID 19 X-ray image segmentation[J], Comput. Biol. Med. 136 (2021), 104609.
[10] D. Zhao, L. Liu, F. Yu, et al., Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy[J], Knowl. Base Syst. 216 (2021), 106510.
[11] H. Yu, J. Song, C. Chen, et al., Image segmentation of Leaf Spot Diseases on Maize using multi-stage Cauchy-enabled grey wolf algorithm[J], Eng. Appl. Artif. Intell. 109 (2022), 104653.
[12] J. Long, E. Shelhamer, T. Darrell, Fully Convolutional Networks for Semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[13] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015, pp. 234–241.
[14] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks[J], Adv. Neural Inf. Process. Syst. 25 (2012).
[15] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image recognition[J], 2014 arXiv preprint arXiv:1409.1556.
[16] K. He, X. Zhang, S. Ren, et al., Deep Residual Learning for Image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[17] G. Huang, Z. Liu, L. Van Der Maaten, et al., Densely Connected Convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial nets[J], Adv. Neural Inf. Process. Syst. 27 (2014).
[19] Q. Guan, Y. Chen, Z. Wei, et al., Medical image augmentation for lesion detection using a texture-constrained multichannel progressive GAN[J], Comput. Biol. Med. 145 (2022), 105444.
[20] H. Seo, C. Huang, M. Bassenne, et al., Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images[J], IEEE Trans. Med. Imag. 39 (5) (2019) 1316–1325.
[21] Q. Jin, Z. Meng, C. Sun, et al., RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans[J], Front. Bioeng. Biotechnol. (2020) 1471.
[22] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, et al., Unet++: redesigning skip connections to exploit multiscale features in image segmentation[J], IEEE Trans. Med. Imag. 39 (6) (2019) 1856–1867.
[23] P.F. Christ, F. Ettlinger, F. Grün, et al., Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural networks[J], 2017 arXiv preprint arXiv:1702.05970.
[24] L.I. Song, K.F. Geoffrey, H.E. Kaijian, Bottleneck feature supervised U-Net for pixel-wise liver and tumor segmentation[J], Expert Syst. Appl. 145 (2020), 113131.
[25] F. Milletari, N. Navab, S A. V-net Ahmadi, Fully Convolutional Neural Networks for Volumetric Medical Image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.
[26] Z. Liu, Y.Q. Song, V.S. Sheng, et al., Liver CT sequence segmentation based with improved U-Net and graph cut[J], Expert Syst. Appl. 126 (2019) 54–63.
[27] P.F. Christ, F. Ettlinger, F. Grün, et al., Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural networks[J], 2017 arXiv preprint arXiv:1702.05970.
[28] Y. Duan, F. Liu, L. Jiao, et al., SAR image segmentation based on convolutional-wavelet neural network and Markov random field[J], Pattern Recogn. 64 (2017) 255–267.
[29] H. Zhao, J. Shi, X. Qi, et al., Pyramid Scene Parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.
[30] W. Xie, C. Jacobs, J.P. Charbonnier, et al., Relational modeling for robust and efficient pulmonary lobe segmentation in CT scans[J], IEEE Trans. Med. Imag. 39 (8) (2020) 2664–2675.
[31] X. Chen, L. Yao, Y. Zhang, Residual Attention U-Net for Automated Multi-Class Segmentation of Covid-19 Chest Ct images[J], 2020 arXiv preprint arXiv: 2004.05645.
[32] V. Mnih, N. Heess, A. Graves, Recurrent models of visual attention[J], Adv. Neural Inf. Process. Syst. 27 (2014).
[33] M. Yu, M. Han, X. Li, et al., Adaptive soft erasure with edge self-attention for weakly supervised semantic segmentation: thyroid ultrasound image case study[J], Comput. Biol. Med. 144 (2022), 105347.
[34] X. Fu, L. Bi, A. Kumar, et al., Multimodal spatial attention module for targeting multimodal PET-CT lung tumor segmentation[J], IEEE J. Biomed. Health Inf. 25 (9) (2021) 3507–3516.
[35] P. Zhao, J. Zhang, W. Fang, et al., SCAU-net: spatial-channel attention U-net for gland segmentation[J], Front. Bioeng. Biotechnol. 8 (2020) 670.
[36] J. Fu, J. Liu, H. Tian, et al., Dual Attention Network for Scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154.
[37] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[38] S. Woo, J. Park, J.Y. Lee, et al., Cbam: Convolutional Block Attention module[C]//Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.
[39] F. Yu, V. Koltun, Multi-scale Context Aggregation by Dilated convolutions[J], 2015 arXiv preprint arXiv:1511.07122.
[40] P. Wang, P. Chen, Y. Yuan, et al., Understanding Convolution for Semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Ieee, 2018, pp. 1451–1460.
[41] D.P. Fan, T. Zhou, G.P. Ji, et al., Inf-net: automatic covid-19 lung infection segmentation from ct images[J], IEEE Trans. Med. Imag. 39 (8) (2020) 2626–2637.
[42] Covid-19 ct segmentation dataset. https://medicalsegmentation.com/covid19/. (Accessed 28 August 2020).
[43] K. Zhang, X. Liu, J. Shen, et al., Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography[J], Cell 181 (6) (2020) 1423–1433, e11.
[44] J. Ma, Y. Wang, X. An, et al., Toward data-efficient learning: a benchmark for COVID-19 CT lung and infection segmentation[J], Med. Phys. 48 (3) (2021) 1197–1210.
[45] F. Shan, Y. Gao, J. Wang, et al., Lung Infection Quantification of COVID-19 in CT Images with Deep learning[J], 2020 arXiv preprint arXiv:2003.04655.
[46] Y. Qiu, Y. Liu, S. Li, et al., Miniseg: an extremely minimum network for efficient covid-19 segmentation[C], Proc. AAAI Conf. Artif. Intell. 35 (6) (2021) 4846–4854.
[47] I. Laradji, P. Rodriguez, O. Manas, et al., A weakly supervised consistency-based learning method for covid-19 segmentation in ct images[C], in: /Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 2453–2462.
[48] F. Yu, Y. Zhu, X. Qin, et al., A multi-class COVID-19 segmentation network with pyramid attention and edge loss in CT images[J], IET Image Process. 15 (11) (2021) 2604–2613.
[49] X. Zhao, P. Zhang, F. Song, et al., D2A U-net: automatic segmentation of COVID-19 CT slices based on dual attention and hybrid dilated convolution[J], Comput. Biol. Med. 135 (2021), 104526.
[50] Y. Zhang, Q. Liao, L. Yuan, et al., Exploiting shared knowledge from non-covid lesions for annotation-efficient covid-19 ct lung infection segmentation[J], IEEE J. Biomed. Health Inf. 25 (11) (2021) 4152–4162.
[51] Z. Gu, J. Cheng, H. Fu, et al., Ce-net: context encoder network for 2d medical image segmentation[J], IEEE Trans. Med. Imag. 38 (10) (2019) 2281–2292.
[52] R. Wang, C. Ji, Y. Zhang, et al., Focus, fusion, and rectify: context-aware learning for COVID-19 lung infection segmentation[J], IEEE Transact. Neural Networks Learn. Syst. 33 (1) (2021) 12–24.
[53] J. Schlemper, O. Oktay, M. Schaap, et al., Attention gated networks: learning to leverage salient regions in medical images[J], Med. Image Anal. 53 (2019) 197–207.
[54] S. Yin, H. Deng, Z. Xu, et al., SD-UNet: A novel segmentation framework for CT images of lung infections[J], Electronics 11 (1) (2022) 130.
[55] N. Ibtehaz, M.S. Rahman, MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation[J], Neural Network. 121 (2020) 74–87.
[56] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, et al., 3D U-Net: Learning Dense Volumetric Segmentation from Sparse annotation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2016, pp. 424–432.