



Research article

Improved U-Net based on cross-layer connection for pituitary adenoma MRI image segmentation

Xiaoliang Jiang^{1,†}, Junjian Xiao^{1,*}, Qile Zhang^{2,†,*}, Lihui Wang³, Jinyun Jiang¹ and Kun Lan¹

¹ College of Mechanical Engineering, Quzhou University, Quzhou 324000, China

² Department of Rehabilitation, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou People's Hospital, Quzhou 324000, China

³ Department of Science and Education, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou People's Hospital, Quzhou 324000, China

† The authors contributed equally to this work.

* **Correspondence:** Email: xjj919@sohu.com, zjzhangqile@163.com.

Abstract: Pituitary adenoma is a common neuroendocrine neoplasm, and most of its MR images are characterized by blurred edges, high noise and similar to surrounding normal tissues. Therefore, it is extremely difficult to accurately locate and outline the lesion of pituitary adenoma. To solve these limitations, we design a novel deep learning framework for pituitary adenoma MRI image segmentation. Under the framework of U-Net, a newly cross-layer connection is introduced to capture richer multi-scale features and contextual information. At the same time, full-scale skip structure can reasonably utilize the above information obtained by different layers. In addition, an improved inception-dense block is designed to replace the classical convolution layer, which can enlarge the effectiveness of the receiving field and increase the depth of our network. Finally, a novel loss function based on binary cross-entropy and Jaccard losses is utilized to eliminate the problem of small samples and unbalanced data. The sample data were collected from 30 patients in Quzhou People's Hospital, with a total of 500 lesion images. Experimental results show that although the amount of patient sample is small, the proposed method has better performance in pituitary adenoma image compared with existing algorithms, and its Dice, Intersection over Union (IoU), Matthews correlation coefficient (Mcc) and precision reach 88.87, 80.67, 88.91 and 97.63%, respectively.

Keywords: pituitary adenoma; U-Net; cross-layer connection; image segmentation; inception-dense

1. Introduction

On the basis of the latest statistics from the World Health Organization, there were 19.29 million new cancers and 9.96 million deaths around the world in 2021. Among all tumor diseases, pituitary adenoma is one of the most common primary intracranial tumors, accounting for about 10–15% of all brain tumors [1]. Although most pituitary adenoma are benign and do not pose a threat to the life of patients, they can cause clinical problems such as visual impairment, infertility and metabolic syndrome. In recent years, countries around the world have put forward “precision medicine” plans, aiming to combine patients’ living environment and medical data, and apply modern genetics, molecular imaging, artificial intelligence and other technologies to develop personalized prevention plans and precision diagnosis and treatment plans. Doctors can use non-invasive methods to monitor the structure and function of organs and tissues to make faster and more accurate assessments and predictions of conditions. Therefore, the research on the assisted pituitary adenoma diagnosis based on deep learning theory has a strong frontier and clinical application value.

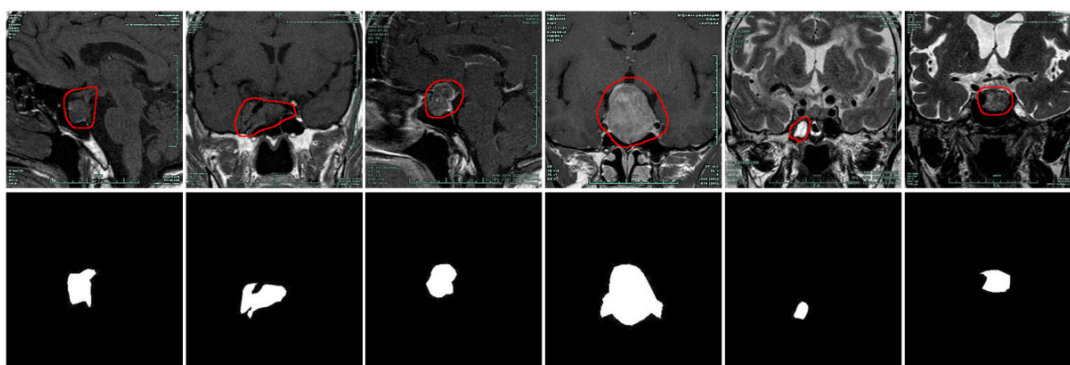


Figure 1. Examples of pituitary adenoma on MR images. The first row: six lesion samples. The second row: the corresponding masks.

In the medical Imaging diagnosis of pituitary adenoma, the conventional techniques mainly involve computed tomography (CT) and magnetic resonance imaging (MRI). Because MR images have good soft tissue resolution and can provide more detailed information on pituitary lesions, it is an ideal method for detecting pituitary adenomas, as shown in Figure 1. However, no matter which imaging technique is used, accurate localization and delineation of pituitary lesion are indispensable. In the clinical treatment implementation system, this work is still done manually, which heavily depends on the radiologist’s reading experience and work status. In addition, image diagnosis is highly subjective and the repetition of segmentation results is poor, which limits the promotion and application of “precision medicine” policy to a certain extent. At present, a large number of researchers and scholars have devoted themselves to the segmentation method of tumor MR images. By using the gray scale, shape and boundary information, various diseased tissues are detected, and then the corresponding diagnosis results are given. However, due to the different locations, shapes and sizes of tumors in MR images, and the presence of noise, field shift effects, and partial volume effects, fully automatic segmentation of medical images is still a complex and challenging research.

With the continuous improvement of medical imaging technology and mathematical theory, lots of MR image segmentation approaches based on deep learning [2–5] have emerged. To automatically segment and detect the location of brain tumors from MRI images, Rai et al. [6] presented a depth-based network framework. By choosing U-Net [7] as the basic skeleton and incorporating ResNext-50 [8] as the encoder, it not only solves the problem of gradient descent, but also saves computational overhead. Lu et al. [9] presented a new end-to-end medical image segmentation model inspired by multi-scale cross-fusion encoding network. In this study, dual context aggregation and attention guided modules are used to construct the network, and the segmentation performance is further improved by optimizing the corresponding objective function. Rehman et al. [10] introduced feature enhancement modules at each encoder stage to achieve the purpose of improving the receiving domain. Meanwhile, the loss function is redesigned to solve the problem of unbalanced class. Tang et al. [11] proposed a dense frame algorithm based on dual attention for nasopharyngeal carcinoma image segmentation, which helps to obtain better performance while alleviating the increase of parameters, thus overcoming the problem of gradient disappearance.

Recently, inception module [12–15] has achieved significant success in medical image segmentation tasks. Zhou et al. [16] added inception layer and extended residual module into the framework of U-Net to achieve accurate segmentation of superficial muscle. By redesigning the inception module, Chen et al. [17] can obtain more features on multi-scale channels. At the same time, BConvLSTM and self-attention layers were added to achieve more accurate segmentation results. Hoorali et al. [18] proposed a detection approach for anthrax microscopic image, which alleviated the problem of different semantic features by exploiting the advantages of inception block and residual block. Zhang et al. [19] presented a new CNN approach by adding inception-res and dense convolution operations into U-Net. The width of its network is increased without need for additional parameters, and the possible gradient disappearance is mitigated. Bala et al. [20] presented a generalized encoder-decoder network by modifying U-Net. Firstly, dense connection is used to replace simple jump connection, and then a new inception module is proposed to replace traditional convolution. Finally, image enhancement and normalization are carried out to increase the generalization ability.

Inspired by the above models, we propose a new automatic segmentation framework for pituitary adenoma which takes advantage of cross-layer connection and inception module. The creative points of this work are summarized: 1) A newly cross-layer is designed to capture richer multi-scale features and contextual information. 2) Full-scale skip structure is utilized, thereby enhancing the learning of pituitary tumor shape. 3) We utilize the inception-dense module for block unit convolution, so that the network can expand the effective receptive field and increase the depth. 4) The training process is optimized by employing the function based on binary cross-entropy and Jaccard losses to obtain better performance.

The remainder of our paper is structured as follows: Section 2 gives the proposed method. Section 3 demonstrates the superiority of our algorithm on real pituitary adenoma images. Finally, the conclusion is discussed in Section 4.

2. Materials and methods

Inspired by the original U-shaped encoder-decoder network in [7], we design a new end-to-end network and its overall framework is shown in Figure 2. The proposed network mainly includes

cross-layer connection, inception-dense module and full-scale skip structure. For down-sampling in coding stage, cross-layer connection is introduced and inception-dense block is used instead of the original convolution block to extract the input image features effectively. During the decoding stage, the feature map is up-sampled by convolution transpose operation and then fused with the multi-layer features information obtained by the down-sampling. Finally, a mixed function is developed to help achieve better training results. The detailed technical of this network will be described in the following subsection.

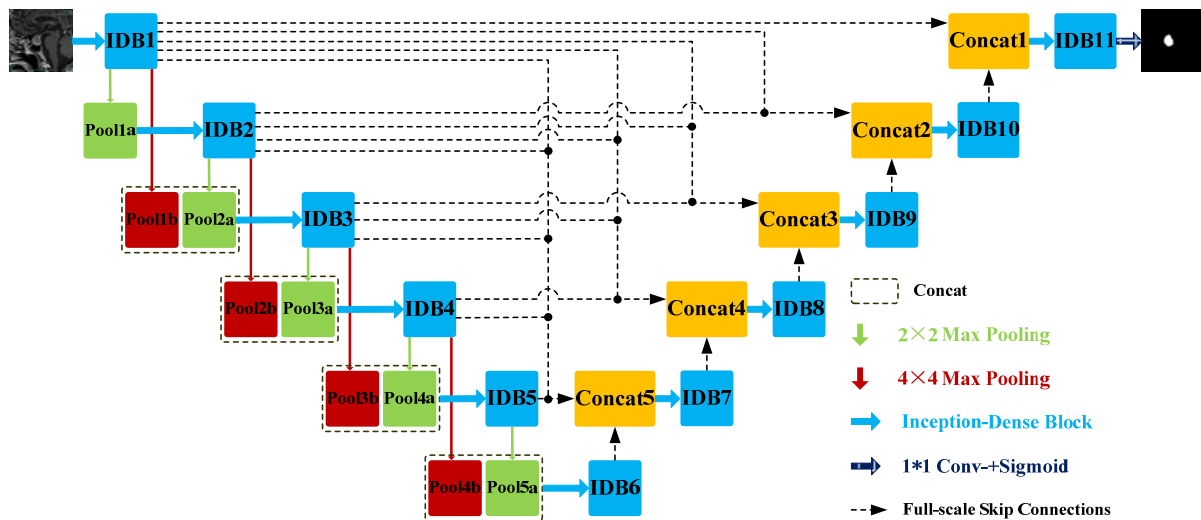


Figure 2. The overview of our proposed U-Net architecture for pituitary adenoma segmentation.

2.1. Cross-layer connection

For pituitary adenoma images, most of them have problems such as blurred edges, high noise and high similarity to the surrounding normal tissues. Therefore, it is difficult to retain target detail under a single scale. In general, multi-layer features can not only expand the receiving field, but also better extract semantic context information and spatial details of the image. From this point, we utilize cross-layer connection to effectively improve network performance [21,22]. Firstly, two multi-scale feature branches (Pool1a and Pool1b) are obtained by applying inception-dense block and Max-pooling layer with step size of 2×2 and 4×4 , as shown in Figure 2. Then, the above operations were repeated to extract multi-scale features Pool2a and Pool2b. Obviously, Pool1b and Pool2a have the same image dimension. The research shows that Pool2a can obtain higher level of context information by small stride, while Pool1b can maintain a larger receptive domain by bigger stride [23]. Therefore, the designed cross-layer connections can achieve the aggregation of multi-scale features and multi-level contextual information.

2.2. Inception-dense module

Inception module [24–26] is designed to improve performance by increasing network width. Its key point is to replace the classical convolution layer with pooling layer and convolution layer in different scales, and adopt 1×1 convolution layer for dimensionality reduction. Inception can make

self-adapt to convolutional kernels of different sizes, and can obtain more complex image features compared to convolutional kernels of fixed size. As shown in Figure 3(a), the original Inception module includes 1×1 , 3×3 , and 5×5 convolutional size filters and a 3×3 maximum pooling layer. The advantage of this structure is that it can learn spatial features at different scales and give different weights, so as to achieve better segmentation results. Based on the original Inception architecture, this paper proposes a modified version, as shown in Figure 3(b). Our improved Inception model adopts the idea of asymmetric convolution decomposition and proposes a lightweight feature information extraction module. The key improvement is to replace the 3×3 convolutional filter in the original Inception module with a set of 1×3 and 3×1 asymmetric convolutional operation. Similarly, we can decompose the 5×5 convolutional filter into 1×5 and 5×1 sequence groups. This asymmetric convolution operation can ensure the receiving field and reduce the module dimension, so as to improve the over-fitting problem.

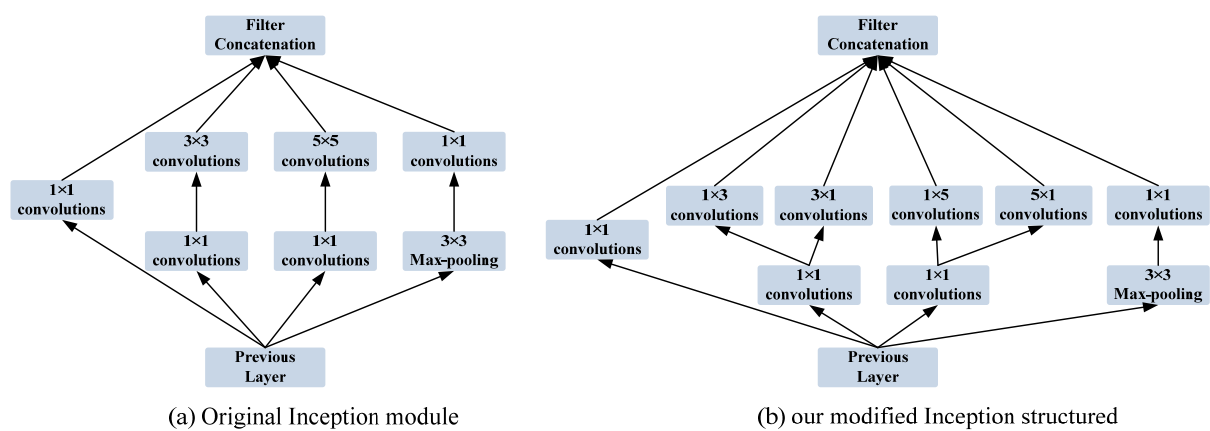


Figure 3. Original Inception module and our modified Inception structured.

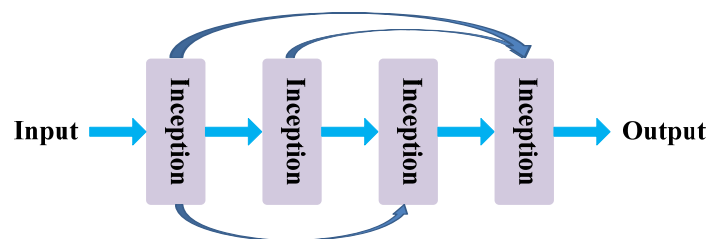


Figure 4. The proposed Inception-dense block.

To strengthen the information transfer between layers, our modified Inception module is inserted into the dense connection blocks, as shown in Figure 4. Therefore, feature-maps $[x_0, \dots, x_{l-1}]$ of all previous layers are used as input, and then the output can be expressed:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (1)$$

where x_l represents the input of layer l , H_l is a composite function that can affect information transfer. In this study, the advantages of dense connection and Inception module were used to obtain

different scale features with convolution kernels of different sizes. At the same time, without increasing the depth and width of network, the features are fully extracted to obtain better generalization and representation ability.

2.3. Full-scale skip structure

The classical U-Net [7] model fuses the feature map output by each layer of the encoder with the feature map generated by the corresponding decoder. It only combines the features of the same scale at the same height, so the global context information will be gradually weakened when it is gradually spread to the shallow layer. The U-Net++ [27] designs a nested and dense jump connection, and utilizes features from each layer of the network to automatically learn the importance level of different depth features. This model integrates the features of different receptive fields, but these features all come directly from the same layer or the next layer, so there are still phenomena such as insufficient fine granularity, loss of edge and location information, etc. To remedy the defects of the above models, a new full-scale skip structure is designed to improve segmentation accuracy. Modifications of this architecture can combine low level details with high level semantics from feature maps at different scales. The difference from [27] is that the modified full-scale skip structure can transmit the characteristics information of all previous layers, containing all layers in the encoder and deeper layers in the decoder, to a specific layer in the decoder. The advantage of this is that each decoder layer can introduce larger scale feature maps from different levels and scales, thus greatly mitigating the phenomenon of feature loss.

2.4. Loss function

To show more balanced segmentation, we design a combination of binary cross-entropy [28,29] and Jaccard [30,31] as loss function. The formula is as follows:

$$L_{total} = L_{BCE} + L_{JA} \quad (2)$$

where L_{BCE} represents binary cross-entropy loss, it is defined as:

$$L_{BCE} = -\sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3)$$

where N denotes the sum of pixels, y_i is the predicted mask, and \hat{y}_i is the ground truth mask. However, when dealing with an imbalanced number of classes, the optimizer can easily get stuck in local minima. For the stability of training, Jaccard loss is introduced:

$$L_{JA} = 1 - \frac{\sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N y_i \hat{y}_i} \quad (4)$$

where L_{JA} represents the Jaccard loss.

3. Experiments and results

In this section, we mainly discuss the experimental results on real patient data from Quzhou People's Hospital. We further prove the superiority of our method by comparing with some classical learning approaches, employing the Dice, IoU, Mcc and precision as the evaluation indicators.

3.1. Experimental settings

3.1.1. Dataset description

The pituitary adenoma MRI dataset used in this paper consists of 500 lesion images, obtained from the rehabilitation department of Quzhou People's Hospital. All images were taken by doctors for the subject's routine examination, with a size of 320×320 pixels. Although the images are acquired with the same equipment, they show high variability due to the influence of field of view size, camera view, appearance and other aspects, as shown in Figure 1. Based on the interactive segmentation results of Labelme, the ground truths are labelled by junior annotators and verified by an experienced radiologist. Considering the limiting memory usage, we adjusted all the images to 256×256 pixels. To eliminate the problem of small samples and unbalanced data, augmentation techniques (such as random rotating, scaling, translating, shearing, left-right flipping, and mosaic augmentation) are employed to extend the training dataset. In the current work, the original MR image dataset is 500 images, which 20% (100 images) are retained for testing and 80% (400 images) are used for enhancement, so the total training dataset is (400×7) 2800 images. The enhanced images and their corresponding label are shown in Figure 5.

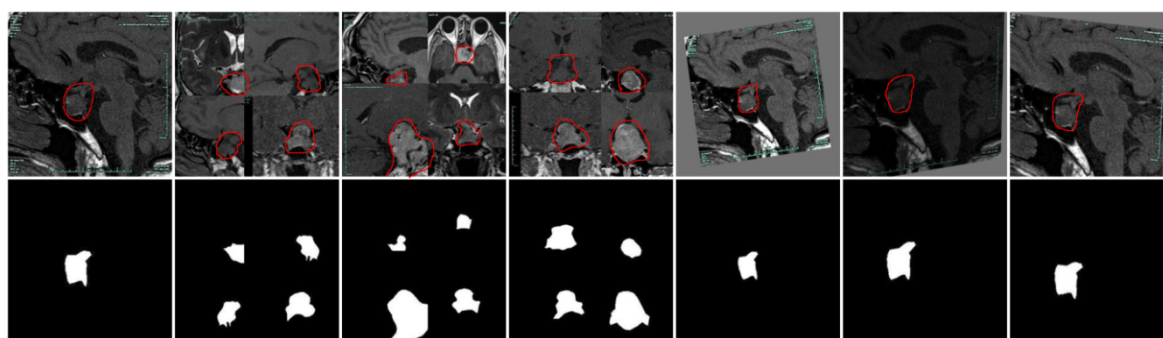


Figure 5. The enhanced images of pituitary adenoma (top) with their corresponding label (bottom).

3.1.2. Evaluation metrics

The Dice coefficient [32], Intersection over Union [33], Matthews correlation coefficient [34] and precision [35] were used to effectively evaluate the advantages of different methods. In addition, the larger these values, the better the segmentation performance. The evaluation matrices are defined as:

$$Dice = \frac{2TP}{2TP + FN + FP} \quad (5)$$

$$IoU = \frac{TP}{TP + FN + FP} \quad (6)$$

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \quad (7)$$

$$precision = \frac{TP}{TP + FP} \quad (8)$$

where TP, TN, FP, and FN denote numbers of true positives, true negatives, false positives, and false negatives on the pixels set, respectively.

3.1.3. Implementation details

Table 1. Results of our model with different batch size.

Batch size	Dice	IoU	Mcc	Precision
2	0.8592	0.7715	0.8632	0.9758
4	0.8687	0.7800	0.8707	0.9759
6	0.8706	0.7813	0.8715	0.9759
8	0.8766	0.7902	0.8778	0.9762
10	0.8771	0.7903	0.8776	0.9760
12	0.8771	0.7895	0.8763	0.9758
14	0.8760	0.7864	0.8767	0.9759
16	0.8887	0.8067	0.8891	0.9763
18	0.8837	0.7976	0.8838	0.9762
20	0.8820	0.7962	0.8821	0.9762

Table 2. Results of our model with different learning rate.

Learning rate	Dice	IoU	Mcc	Precision
0.1	0.8791	0.7913	0.8790	0.9759
0.01	0.8829	0.7979	0.8835	0.9762
0.001	0.8887	0.8067	0.8891	0.9763
0.0001	0.8837	0.7984	0.8842	0.9761

Table 3. Results of our model with different channel.

Channel	Dice	IoU	Mcc	Precision
16	0.8725	0.7854	0.8753	0.9756
24	0.8768	0.7888	0.8775	0.9758
32	0.8887	0.8067	0.8891	0.9763
48	0.8862	0.8027	0.8869	0.9761

All methods are implemented using Python 3.7 based on the open-source framework Keras package with Tensorflow 2.0 as backend. The experimental environment is based on a Windows10 workstation configured with an Intel Xeon Gold 6154 CPU@3.00 GHz processor, 64 GB of 2933 MHz DDR4 ECC RDIMM, NVIDIA Quadro RTX 6000 GPU memory of 24 GB. Since the convolutional neural network will continuously iterate during the training process, it is necessary to estimate the impact of these hyper-parameters on the model performance, mainly including batch size, learning rate and channel. Based on the speed and environment configuration, Adam optimizer [36] is utilized to optimize the segmentation network. As shown in Tables 1 and 2, we can clearly see that the network performance is best when the batch size is set to 16 and the learning rate is set to 0.001. From Table 3, the network achieves the best results when the channel is set to 32.

3.2. Comparison experiment

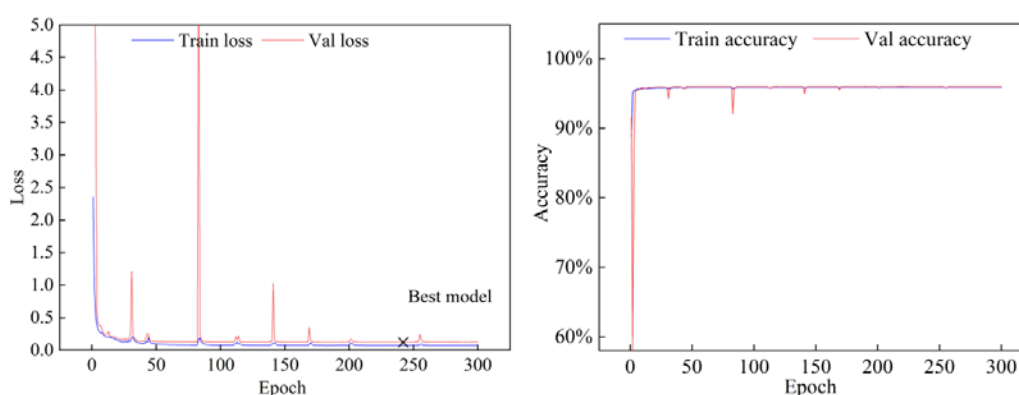


Figure 6. Loss and accuracy curve during training process.

We firstly observe the accuracy and loss values of 300 epochs on the training and validation sets, and the results are shown in Figure 6. It can be shown that for all cases, the proposed method can quickly converge. This is mainly due to the advantages of Inception-dense module in feature extraction. In addition, it should be noted that, except for some special points, the deviation of the loss and accuracy curves on the training set and test set is very small, which indicates that our model is reliable and robust.

Figure 7 qualitatively evaluates the segmentation results of different methods with ground truth. The first column is the original images, the second column is the ground truth hand-sketched by a radiologist, the third to last columns are segmentation results of AttUNet [37], DenseUnet [38], ENet [39], HRNet [40], ICNet [41], R2U-net [42], SegNet [43], U-Net [7], XNet [44], U-Net+++ [45], UNet++ [27] and our method, respectively. According to the visualization results, it is obvious that U-net and XNet are easily disturbed by noise in the complex background, so these two models perform the worst. Through the multi-scale feature extraction techniques, U-Net+++ and UNet++ can increase the completeness of complex lesion segmentation, but the edge details are obviously insufficient. Although AttUNet and DenseUnet show some advantages in dealing with complex shadows, they still suffer from insufficient segmentation. The other methods will also misclassify some shadows as lesions, resulting in over-segmentation or under-segmentation. However, our proposed segmentation model can well complete the task of

medical image segmentation, and has a good ability to extract details of lesions. This shows that our model is more stable than other methods during the evolution, and the network fluctuation is smaller.

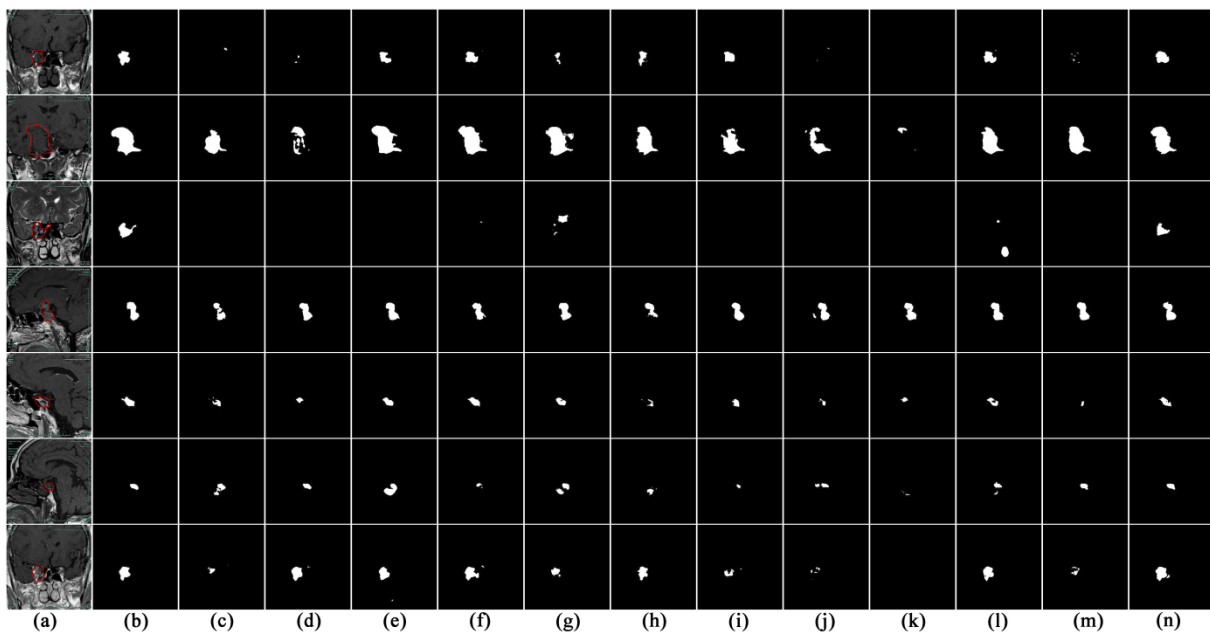


Figure 7. The segmentation result of pituitary adenoma images by different models. (a) original images; (b) their labels; (c) AttUNet; (d) DenseUNet; (e) ENet; (f) HRNet; (g) ICNet; (h) R2U-Net; (i) SegNet; (j) U-Net; (k) XNet; (l) U-Net+++; (m) UNet++; (n) our method.

Table 4. Results of different models on pituitary adenoma images.

Method	Dice	IoU	Mcc	Precision
AttUNet [37]	0.8683	0.7802	0.8721	0.9759
DenseUNet [38]	0.8575	0.7620	0.8600	0.9755
ENet [39]	0.8722	0.7847	0.8745	0.9761
HRNet [40]	0.8804	0.7988	0.8837	0.9762
ICNet [41]	0.8596	0.7688	0.8640	0.9758
R2U-Net [42]	0.8745	0.7890	0.8777	0.9761
SegNet [43]	0.8569	0.7599	0.8587	0.9759
U-Net [7]	0.8636	0.7713	0.8664	0.9758
XNet [44]	0.8573	0.7630	0.8606	0.9753
U-Net+++ [45]	0.8794	0.7933	0.8806	0.9761
UNet++ [27]	0.8721	0.7846	0.8749	0.9759
Our model	0.8887	0.8067	0.8891	0.9763

To verify the proposed scheme, the above models are quantitatively evaluated and compared using the evaluation metrics of Dice, IoU, Mcc and precision, which is listed in Table 4. According to Table 4, the competitive performance of DenseUNet, ICNet, SegNet and XNet is generally lower than that of other networks. As the baseline, the Dice, IoU, Mcc and precision values of U-Net

reached 0.8636, 0.7713, 0.8664 and 0.8636. UNet++ and UNet+++ have nested dense jump paths, so there are obvious improvements in false detection and missed detection. On the basis of the U-Net network architecture, the segmentation effect of pituitary tumor lesions can be significantly improved whether the attention module or Recurrent Residual module is added. Compared with the U-Net, the Dice, IoU, Mcc and precision values of our method are improved to 2.51, 3.54, 2.27 and 0.05%, respectively. This result can show the effectiveness of the proposed network in the localization and extraction of pituitary tumor lesions.

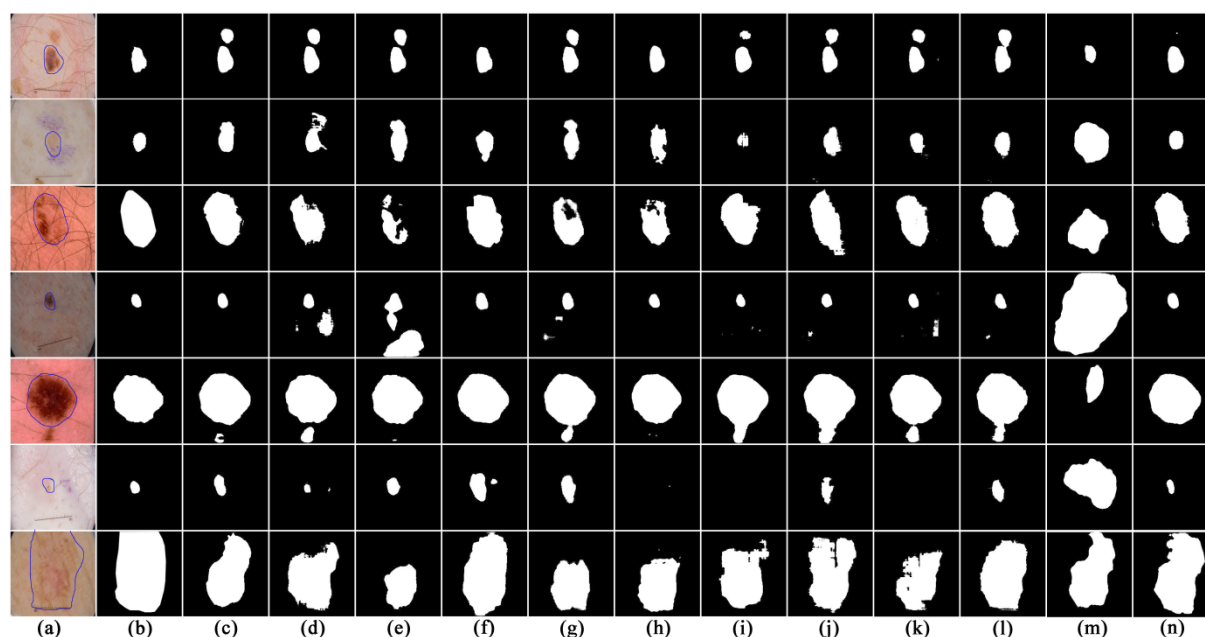


Figure 8. The segmentation result of ISIC-2018 dataset by different models. (a) original images; (b) their labels; (c) AttUNet; (d) DenseUnet; (e) ENet; (f) HRNet; (g) ICNet; (h) R2U-Net; (i) SegNet; (j) U-Net; (k) XNet; (l) U-Net+++; (m) UNet++; (n) our method.

Table 5. Results of different models on ISIC-2018 dataset.

Method	Dice	IoU	Mcc	Precision
AttUNet [37]	0.8460	0.7351	0.8062	0.9348
DenseUnet [38]	0.8387	0.7241	0.7969	0.9311
ENet [39]	0.8320	0.7141	0.7887	0.9333
HRNet [40]	0.8343	0.7177	0.7914	0.9307
ICNet [41]	0.8616	0.7589	0.8253	0.9334
R2U-Net [42]	0.8317	0.7140	0.7893	0.9328
SegNet [43]	0.8285	0.7092	0.7829	0.9259
U-Net [7]	0.8377	0.7225	0.7947	0.9288
XNet [44]	0.8332	0.7166	0.7927	0.9279
U-Net+++ [45]	0.8475	0.7372	0.8072	0.9315
UNet++ [27]	0.8415	0.7283	0.8086	0.9325
Our model	0.8683	0.7686	0.8344	0.9349

To verify the processing performance on complex data, the ISIC-2018 Challenge dataset [46] was used for further testing. The dataset is released by the international organization ISIC, and it has 2594 images of skin lesions with corresponding labels. Among them, 70% of these images are used for training, 10% for validation, and the rest for testing. The images in this database are characterized by large differences in the size of the lesion area, low contrast with the surrounding skin, hair interference, and irregular shape. Figure 8 shows seven examples of our proposed method and other models at ISIC-2018 dataset, where white represents the pixels of the lesion area, and black represents the normal skin pixels. It can be observed that our method can better cope with the influence of external interference, accurately distinguish lesions from surrounding tissues, and thus can completely segment skin lesions. Table 5 shows the experimental results of various methods. It can be seen that the proposed network can achieve Dice, IoU, Mcc and precision of 0.8683, 0.7686, 0.8344 and 0.9349, indicating that the performance of the modified model is very competitive.

3.3. Ablation experiment

Table 6. The ablation experiment on effectiveness of each structure in our model.

Method	Dice	IoU	Mcc	Precision
Baseline (U-Net)	0.8636	0.7713	0.8664	0.9758
Baseline + Cross-layer	0.8683	0.7764	0.8698	0.9759
Baseline + Cross-layer + Inception-dense	0.8692	0.7785	0.8709	0.9761
Baseline + Cross-layer + Full-scale-skip	0.8736	0.7872	0.8767	0.9761
Baseline + Cross-layer + Inception-dense+ Full-scale-skip	0.8887	0.8067	0.8891	0.9763

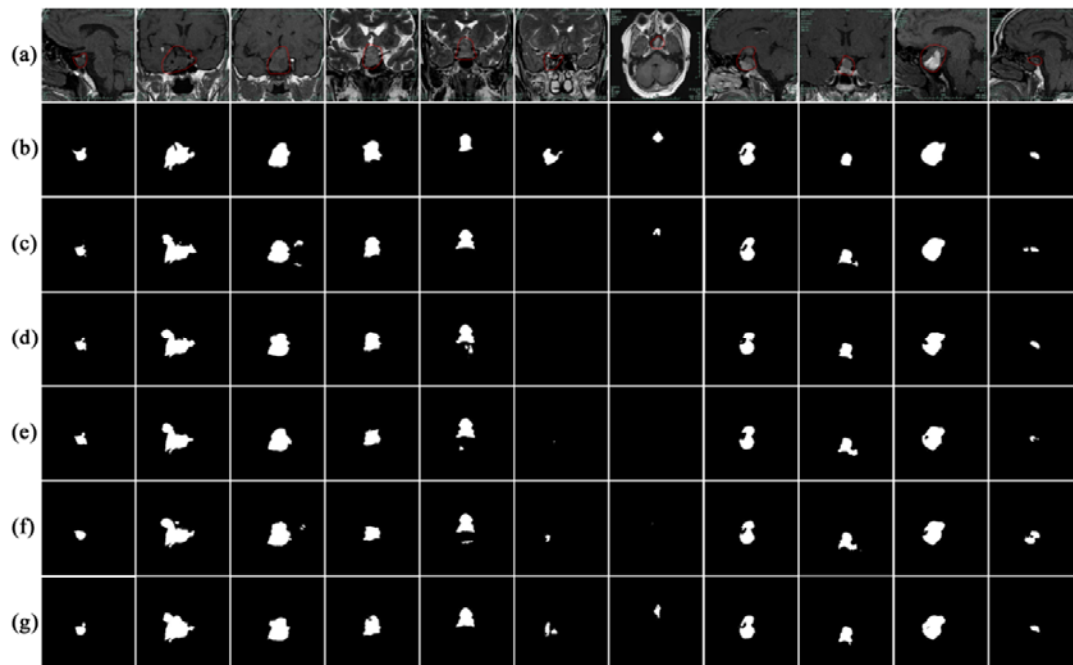


Figure 9. The segmentation result of different structures on pituitary adenoma images. (a) original images; (b) their labels; (c) Baseline (U-Net); (d) Baseline + Cross-layer; (e) Baseline + Cross-layer + Inception-dense; (f) Baseline + Cross-layer + Full-scale skip, (g) Baseline + Cross-layer + Inception-dense + Full-scale skip.

Next, the contribution of each module in the proposed model was performed for ablation study to verify their performance. For a fair comparison, the U-Net model is selected as the benchmark, and the results are shown in Table 6. It can be observed that under the same segmentation framework, the overall segmentation performance is slightly improved after the introduction of the cross-layer module, and the Dice, IoU, Mcc and prcision values are increased by 0.47, 0.51, 0.34 and 0.01% respectively. Under the framework of Baseline+Cross-layer, inception-dense module and full-scale skip structure are added respectively, and the segmentation performance is further improved. When all modules are fused into the U-Net framework, the values of Dice, IoU, Mcc and prcision all reach the optimal value. This indicates that the cross-layer connection, inception-dense module and full-scale skip structure proposed in our model can well strengthen the feature dependencies between channels and capture the lesion features of pituitary tumors, which can help the network to make better decisions.

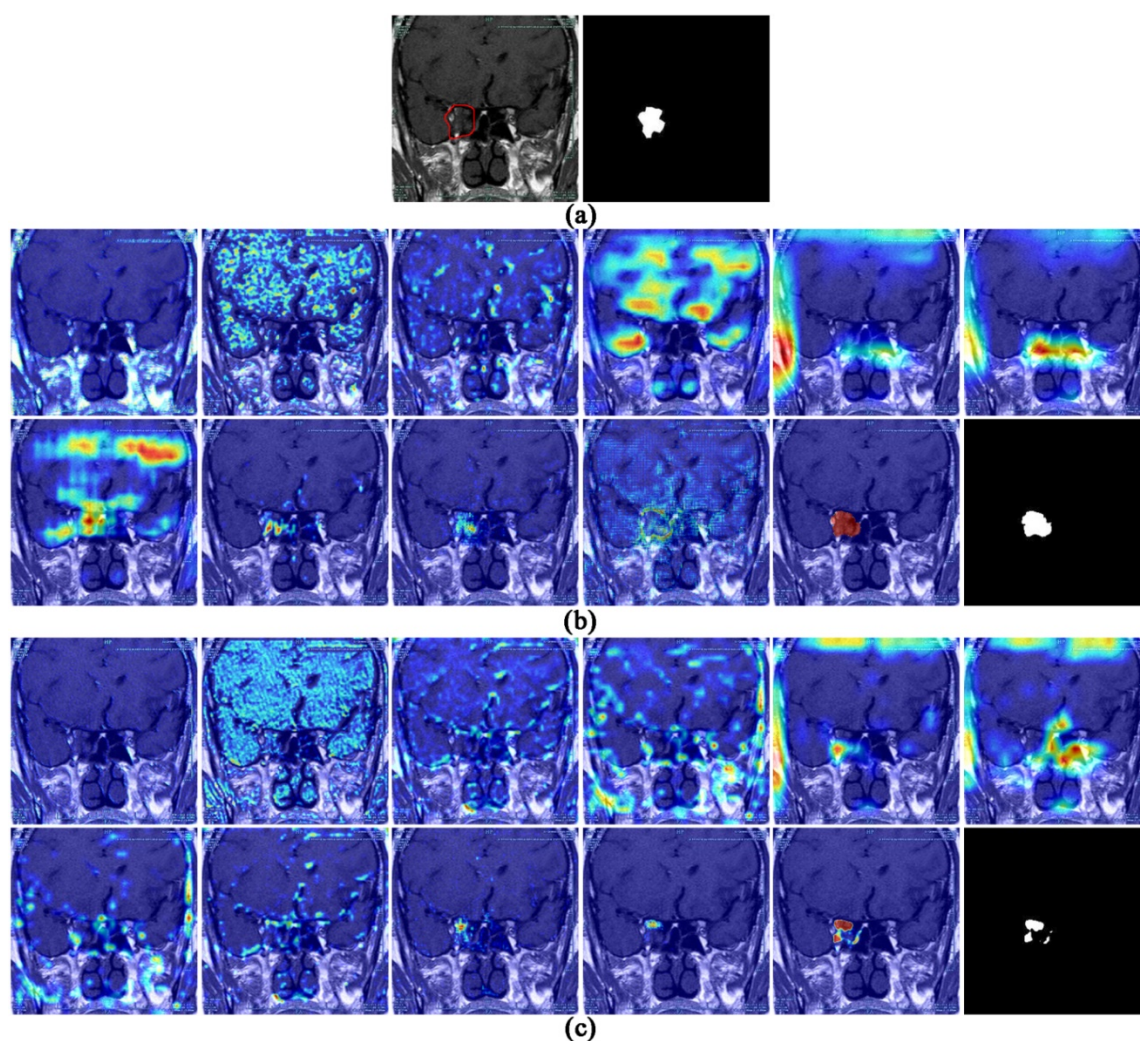


Figure 10. Predicted heatmaps obtained by our model with and without the Inception-dense module. (a) Pituitary adenoma image (left) with its corresponding label (right). (b) Our model with Inception-dense module (from left to right, the heatmaps of images passing through the modules of IDB1, IDB2, IDB3, IDB4, IDB5, IDB6, IDB7, IDB8, IDB9, IDB10 and IDB11, and the last one is the predicted mask). (c) Our model without Inception-dense module.

Figure 9 shows the visualization results of the ablation experiment, the first two rows show pituitary tumor lesions and their corresponding label images, respectively. It can be found that due to the influence of other tissues and blurred boundary, the segmentation effect of Baseline for the lesion is poor, and the edge part is not smooth. After adding the cross-layer module, the performance of the model is improved, but there are still some classification errors. When the inception-dense module and full-scale skip structure is incorporated, our model can accurately segment the pituitary tumor lesion area, and the segmentation edge is smoother and has higher similarity with the gold standard.

To further demonstrate that the Inception-density module outperforms traditional convolutional modules for segmentation problems, the experiments are carried out with and without the Inception-dense module. Figure 10 shows the heatmaps of our model in the process of pituitary tumor image training. The heat map shows that by integrating multi-scale features and multi-level context information, the Inception-dense module can depict more and clearer lesion features, which is helpful to achieve better segmentation performance.

3.4. Complexity analysis

The computational complexity and the number of parameters can evaluate the complexity of a model and limit its application in real scenarios to some extent. As shown in Table 7, due to the introduction of cross-layer connection, inception-Dense Module and full-scale skip structure, our model has a large number of parameters, but the processing time is in the middle among all methods. Therefore, our model achieves a good balance between detection accuracy and efficiency.

Table 7. Comparison of Parameters and computational efficiency on pituitary adenoma dataset.

Method	Parameter (M)	Time (ms/step)
AttUNet [37]	8.49	35
DenseUnet [38]	0.58	32
ENet [39]	0.35	23
HRNet [40]	27.28	57
ICNet [41]	6.44	17
R2U-Net [42]	16.83	45
SegNet [43]	2.80	17
U-Net [7]	1.12	11
XNet [44]	11.19	39
U-Net+++ [45]	21.57	30
UNet++ [27]	8.62	34
Our model	24.41	29

4. Conclusions

In this paper, by modifying the original U-Net architecture, we propose a novel automatic segmentation framework for pituitary adenoma MRI image segmentation. The core idea is to design a newly cross-layer connection in the coding phase to capture richer multi-scale features and contextual information. Meanwhile, full-scale skip structure is employed to combine the output of each path, which can make full use of the above information obtained by different layers.

Furthermore, in the whole framework, an improved inception-dense module is used to replace the standard convolution block of U-Net, which can expand the effective receptive field and increase the depth. Experimental results on pituitary adenoma images indicate that the proposed model can effectively and accurately achieve high-quality lesion segmentation when compared with other advanced deep learning-based method.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62102227, 51805124, 62101206), Zhejiang Basic Public Welfare Research Project (No. LZY22E050001, LZY22D010001, LGG19E050013, LZY21E060001), Science and Technology Major Projects of Quzhou (2021K29).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. Feng, H. Gao, Q. Zhang, Y. Zhou, C. Li, S. Zhao, et al., Metabolic profiling reveals distinct metabolic alterations in different subtypes of pituitary adenomas and confers therapeutic targets, *J. Transl. Med.*, **17** (2019), 1–13. <https://doi.org/10.1186/s12967-019-2042-9>
2. X. M. Liu, Q. Yuan, Y. Z. Gao, K. L. He, S. Wang, X. Tang, et al., Weakly supervised segmentation of COVID-19 infection with scribble annotation on CT images, *Pattern Recognit.*, **122** (2022), 108341. <https://doi.org/10.1016/j.patcog.2021.108341>
3. B. J. Kar, M. V. Cohen, S. P. McQuiston, C. M. Malozzi, A deep-learning semantic segmentation approach to fully automated MRI-based left-ventricular deformation analysis in cardiotoxicity, *Magn. Reson. Imaging*, **78** (2021), 127–139. <https://doi.org/10.1016/j.mri.2021.01.005>
4. N. Mu, H. Y. Wang, Y. Zhang, J. F. Jiang, J. S. Tang, Progressive global perception and local polishing network for lung infection segmentation of COVID-19 CT images, *Pattern Recognit.*, **120** (2021), 108168. <https://doi.org/10.1016/j.patcog.2021.108168>
5. X. M. Liu, Z. S. Guo, J. Cao, J. S. Tang, MDC-net: A new convolutional neural network for nucleus segmentation in histopathology images with distance maps and contour information, *Comput. Biol. Med.*, **135** (2021), 104543. <https://doi.org/10.1016/j.combiomed.2021.104543>
6. H. M. Rai, K. Chatterjee, S. Dashkevich, Automatic and accurate abnormality detection from brain MR images using a novel hybrid UnetResNext-50 deep CNN model, *Biomed. Signal Process. Control*, **66** (2021), 102477. <https://doi.org/10.1016/j.bspc.2021.102477>
7. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, Cham, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

8. S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, (2017), 5987–5995. <https://doi.org/10.1016/j.cmpb.2021.106566>
9. H. C. Lu, S. W. Tian, L. Yu, L. Liu, J. L. Cheng, W. D. Wu, et al., DCACNet: Dual context aggregation and attention-guided cross deconvolution network for medical image segmentation, *Comput. Methods Programs Biomed.*, **214** (2022), 106566. <https://doi.org/10.1016/j.cmpb.2021.106566>
10. M. U. Rehman, S. Cho, J. Kim, K. T. Chong, BrainSeg-Net: Brain tumor MR image segmentation via enhanced encoder-decoder network, *Diagnostics*, **11** (2021), 169. <https://doi.org/10.3390/diagnostics11020169>
11. P. Tang, C. Zu, M. Hong, R. Yan, X. C. Peng, J. H. Xiao, et al., DA-DSUNet: Dual attention-based dense SU-Net for automatic head-and-neck tumor segmentation in MRI images, *Neurocomputing*, **435** (2021), 103–113. <https://doi.org/10.1016/j.neucom.2020.12.085>
12. U. Latif, A. R. Shahid, B. Raza, S. Ziauddin, M. A. Khan, An end-to-end brain tumor segmentation system using multi-inception-UNet, *Int. J. Imaging Syst. Technol.*, **31** (2021), 1803–1816. <https://doi.org/10.1002/ima.22585>
13. X. F. Du, J. S. Wang, W. Z. Sun, Densely connected U-Net retinal vessel segmentation algorithm based on multi-scale feature convolution extraction, *Med. Phys.*, **48** (2021), 3827–3841. <https://doi.org/10.1002/mp.14944>
14. Z. Y. Wang, Y. J. Peng, D. P. Li, Y. F. Guo, B. Zhang, MMNet: A multi-scale deep learning network for the left ventricular segmentation of cardiac MRI images, *Appl. Intell.*, **52** (2022), 5225–5240. <https://doi.org/10.1007/s10489-021-02720-9>
15. M. Yang, H. W. Wang, K. Hu, G. Yin, Z. Q. Wei, IA-Net: An inception-attention-module-based network for classifying underwater images from others, *IEEE J. Oceanic Eng.*, **47** (2022), 704–717. <https://doi.org/10.1109/JOE.2021.3126090>
16. J. S. Zhou, Y. W. Lu, S. Y. Tao, X. Cheng, C. X. Huang, E-Res U-Net: An improved U-Net model for segmentation of muscle images, *Expert Syst. Appl.*, **185** (2021), 115625. <https://doi.org/10.1016/j.eswa.2021.115625>
17. S. Y. Chen, Y. N. Zou, P. X. Liu, IBA-U-Net: Attentive BConvLSTM U-Net with redesigned inception for medical image segmentation, *Comput. Biol. Med.*, **135** (2021), 104551. <https://doi.org/10.1016/j.compbiomed.2021.104551>
18. F. Hoorali, H. Khosravi, B. Moradi, IRUNet for medical image segmentation, *Expert Syst. Appl.*, **191** (2022), 116399. <https://doi.org/10.1016/j.eswa.2021.116399>
19. Z. Zhang, C. D. Wu, S. Coleman, D. Kerr, Dense-inception U-Net for medical image segmentation, *Comput. Biol. Med.*, **192** (2020), 105395. <https://doi.org/10.1016/j.cmpb.2020.105395>
20. S. A. Bala, S. Kant, Dense dilated inception network for medical image segmentation, *Int. J. Adv. Comput. Sci. Appl.*, **11** (2020), 785–793. <https://doi.org/10.14569/IJACSA.2020.0111195>
21. L. Wang, J. Gu, Y. Chen, Y. Liang, W. Zhang, J. Pu, et al., Automated segmentation of the optic disc from fundus images using an asymmetric deep learning network, *Pattern Recognit.*, **112** (2021), 107810. <https://doi.org/10.1016/j.patcog.2020.107810>
22. Z. Zheng, Y. Wan, Y. Zhang, S. Xiang, D. Peng, B. Zhang, CLNet: Cross-layer convolutional neural network for change detection in optical remote sensing imagery, *ISPRS J. Photogramm. Remote Sens.*, **175** (2021), 247–267. <https://doi.org/10.1016/j.isprsjprs.2021.03.005>

23. H. S. Zhao, J. P. Shi, X. J. Qi, X. G. Wang, J. Y. Jia, Pyramid scene parsing network, in *IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 6230–6239.
24. S. Ran, J. Ding, B. Liu, X. Ge, G. Ma, Multi-U-Net: Residual module under multisensory field and attention mechanism based optimized U-Net for VHR image semantic segmentation, *Sensors*, **21** (2021), 1794. <https://doi.org/10.3390/s21051794>
25. R. M. Rad, P. Saeedi, J. Au, J. Havelock, Trophoctoderm segmentation in human embryo images via inceptioned U-Net, *Med. Image Anal.*, **62** (2020), 101612. <https://doi.org/10.1016/j.media.2019.101612>
26. N. S. Punn, S. Agarwal, Multi-modality encoded fusion with 3d inception u-net and decoder model for brain tumor segmentation, *Multimed. Tools Appl.*, **80** (2020), 30305–30320. <https://doi.org/10.1007/s11042-020-09271-0>
27. Z. W. Zhou, M. M. R. Siddiquee, N. Tajbakhsh. J. M. Liang, UNet++: A nested U-Net architecture for medical image segmentation, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (2018), 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
28. B. Zuo, F. F. Lee, Q. Chen, An efficient U-shaped network combined with edge attention module and context pyramid fusion for skin lesion segmentation, *Med. Biol. Eng. Comput.*, **60** (2022), 1987–2000. <https://doi.org/10.1007/s11517-022-02581-5>
29. D. P. Li, Y. J. Peng, Y. F. Guo, J. D. Sun, MFAUNet: Multiscale feature attentive U-Net for cardiac MRI structural segmentation, *IET Image Proc.*, **16** (2022), 1227–1242. <https://doi.org/10.1049/ipr2.12406>
30. V. S. Bochkov, L. Y. Kataeva, wUUNet: Advanced fully convolutional neural network for multiclass fire segmentation, *Symmetry*, **13** (2021), 98. <https://doi.org/10.3390/sym13010098>
31. D. John, C. Zhang, An attention-based U-Net for detecting deforestation within satellite sensor imagery, *Int. J. Appl. Earth Obs. Geoinf.*, **107** (2022), 102685. <https://doi.org/10.1016/j.jag.2022.102685>
32. Y. Y. Yang, C. Feng, R. F. Wang, Automatic segmentation model combining U-Net and level set method for medical images, *Expert Syst. Appl.*, **153** (2020), 113419. <https://doi.org/10.1016/j.eswa.2020.113419>
33. I. Ahmed, M. Ahmad, G. Jeon, A real-time efficient object segmentation system based on u-net using aerial drone images, *J. Real-Time Image Process.*, **18** (2021), 1745–1758. <https://doi.org/10.1007/s11554-021-01166-z>
34. M. Jiang, F. Zhai, J. Kong, A novel deep learning model DDU-net using edge features to enhance brain tumor segmentation on MR images, *Artif. Intell. Med.*, **121** (2021), 102180. <https://doi.org/10.1016/j.artmed.2021.102180>
35. D. Li, A. Cong, S. Guo, Sewer damage detection from imbalanced CCTV inspection data using deep convolutional neural networks with hierarchical classification, *Autom. Constr.*, **101** (2019), 199–208. <https://doi.org/10.1016/j.autcon.2019.01.017>
36. M. M. Ji, Z. B. Wu, Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic, *Comput. Electron. Agric.*, **193** (2022), 106718. <https://doi.org/10.1016/j.compag.2022.106718>
37. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., Attention U-Net: Learning where to look for the pancreas, preprint, arXiv:1804.03999.

38. G. Huang, Z. Liu, V. Laurens, K. Q. Weinberger, Densely connected convolutional networks, in *IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
39. A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, ENet: A deep neural network architecture for real-time semantic segmentation, preprint, arXiv:1606.02147.
40. K. Sun, B. Xiao, D. Liu, J. D. Wang, Deep high-resolution representation learning for human pose estimation, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2019), 5693–5703. <https://doi.org/10.1109/CVPR.2019.00584>
41. H. H. Zhao, X. J. Qi, X. Y. Shen, J. P. Shi, J. Y. Jia, ICNet for real-time semantic segmentation on high-resolution images, in *Proceedings of the European Conference on Computer Vision*, (2018), 405–420.
42. M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, V. K. Asari, Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation, preprint, arXiv:1802.06955.
43. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
44. J. Bullock, C. Cuesta-Lázaro, A. Quera-Bofarull, XNet: A convolutional neural network (CNN) implementation for medical X-Ray image segmentation suitable for small datasets, in *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, (2019), 453–463. <https://doi.org/10.1117/12.2512451>
45. H. Huang, L. Lin, R. Tong, H. Hu, J. Wu, UNet 3+: A full-scale connected UNet for medical image segmentation, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (2020), 1055–1059. <https://doi.org/10.1109/ICASSP40776.2020.9053405>
46. P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Sci. Data*, **5** (2018), 180161. <https://doi.org/10.1038/sdata.2018.161>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)