Appl. Math. J. Chinese Univ.
2022, 37(1): 73-90

Learning a Discriminative Feature Attention Network for pancreas CT segmentation

HUANG Mei-xiang[1] WANG Yuan-jin[1] HUANG Chong-fei[2]
YUAN Jing[3,*] KONG De-xing[2,*]

Abstract. Accurate pancreas segmentation is critical for the diagnosis and management of diseases of the pancreas. It is challenging to precisely delineate pancreas due to the highly variations in volume, shape and location. In recent years, coarse-to-fine methods have been widely used to alleviate class imbalance issue and improve pancreas segmentation accuracy. However, cascaded methods could be computationally intensive and the refined results are significantly dependent on the performance of its coarse segmentation results. To balance the segmentation accuracy and computational efficiency, we propose a Discriminative Feature Attention Network for pancreas segmentation, to effectively highlight pancreas features and improve segmentation accuracy without explicit pancreas location. The final segmentation is obtained by applying a simple yet effective post-processing step. Two experiments on both public NIH pancreas CT dataset and abdominal BTCV multi-organ dataset are individually conducted to show the effectiveness of our method for 2D pancreas segmentation. We obtained average Dice Similarity Coefficient (DSC) of 82.82±6.09%, average Jaccard Index (JI) of 71.13± 8.30% and average Symmetric Average Surface Distance (ASD) of 1.69 ± 0.83 mm on the NIH dataset. Compared to the existing deep learning-based pancreas segmentation methods, our experimental results achieve the best average DSC and JI value.

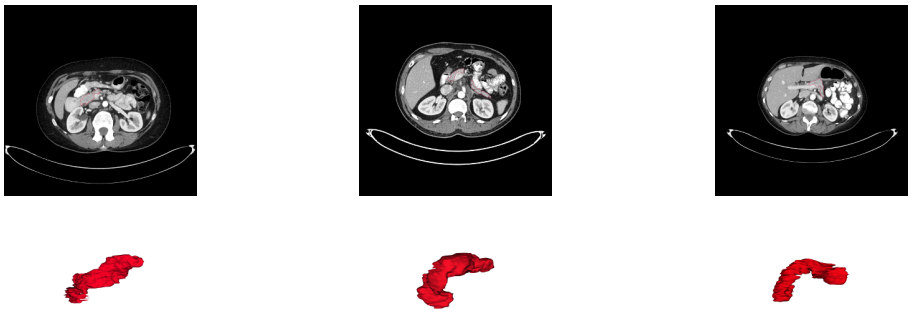
§1 Introduction

Organ segmentation usually refers to the process of extracting specific target organs from medical images. Accurate organ segmentation is a prerequisite for organ measurement, surgical guidance, and radiotherapy effect evaluation in computer-aided diagnosis technology [36]. The

Received: 2020-12-21. Revised: 2021-05-26.
MR Subject Classification: 97R40, 97R20, 65S05.
Keywords: attention mechanism, Discriminative Feature Attention Network, Improved Refinement Residual Block, pancreas CT segmentation.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-022-4346-4.
Supported by the Ph.D. Research Startup Project of Minnan Normal University(KJ2021020) and the National Natural Science Foundation of China(12090020 and 12090025) and Zhejiang Provincial Natural Science Foundation of China(LSD19H180005).
*Corresponding author.

pancreas is a soft organ located on the periphery of the abdomen, which lacks a fixed shape and is hidden behind the peritoneum [10]. Pancreas-related diseases are relatively hidden and difficult to detect and cure, especially for pancreatic cancers, which is still accompanied by higher mortality and lower postoperative survival rate [32]. In clinical practice, the pancreas volume is manually delineated by radiologists for the diagnose of pancreas disease and quantitative assessment. For example, the volume of pancreas enables the physicians to estimate endocrine and exocrine pancreatic functions [1]. However, manual annotation is a highly time-consuming and subject to operators. Hence, an accurate and robust automatic segmentation method of pancreas is highly demanded in the clinical management of pancreas diseases, which can allow to alleviate the workload of radiologists and improve the consistency of pancreas segmentation.

It is challenging to accurate segmentation of pancreas in CT images for the following reasons. First, the intensity distribution between the pancreas and its surrounding structural tissues is very close. As shown in Fig 1, the pancreas boundaries are difficult to distinguish even after contrast adjustments. Second, the pancreas is a small and soft abdominal organ with highly irregular shape, leading to severe class imbalance and difficulty in designing a method to adaptively cover all possible pancreas variabilities [10]. Third, it can be seen from Fig 1 that discontinuities exist in some pancreas slices, which is prone to over-segment and under-segment.



(a) Case #1 (b) Case #7 (c) Case #81

Figure 1. Examples of variations in appearance, shape and size of the pancreas. (a)-(c) denote example axial CT slice from three different patients and corresponding 3D ground truth, respectively.

To address the aforementioned challenges, many pancreas segmentation works have been proposed over the past few years, which can be categorized into two types: top-down and bottom-up methods [11]. In the top-down methods, segmentation is performed by multi-atlas registration and label fusion (MALF) [9, 16, 26, 31]. To reduce the misselection of similar atlas caused by CT intensity, Karasawa et al. proposed a new atlas selection strategy based on vessel structure around the pancreatic tissue for pancreas segmentation [9]. Experimental results show the atlas selection based on vessel structure is much more effective in selecting atlases with similar pancreatic shape and position. However, it is not trivial to select atlases that is general enough to cover all possible pancreas variabilities due to the highly irregular

shape and poor contrast with spatially adjacent abdominal tissues.

Recently, Dense prediction based on deep convolutional neural networks have achieved great success in computer vision and medical imaging, such as FCN [15] and Deeporgan [19], which also boost the pancreas segmentation. Since the pancreas often occupies a small proportion of the whole abdomen, most pancreas segmentation methods rely on multi-stage [2, 12], cascaded CNNs [6, 20, 21, 38], in order to improve the segmentation accuracy. Roth et al. firstly proposed a bottom-up, coarse-to-fine approach for pancreas CT segmentation, utilizing multi-level deep ConvNet model to learn robust pancreas features representation and effectively prune the coarse pancreas over-segmentation [19]. This framework is further improved by the holistically-nested segmentation networks [20, 21]. Zhou et al. proposed a 2D fixed-point models based on FCN-8s [38], in which coarse segmentation provides pancreas location for further fine-scaled models iteratively. Asaturyan et al. presented an approach for automatic pancreas segmentation based on a hierarchical pooling of information by classifying extracted image patches, superpixels and intensity distributions as pancreatic tissue or otherwise [2]. Li et al. proposed a new CAD model for pancreas cancer on PET/CT images based on a gray interval mapping (GIP) method and dual threshold principal component analysis [12]. Although, the multi-stage methods have demonstrated significant improvements over the traditional methods, it is complex to train and lack of generalization due to the presence of multiple learning stages [25].

Attention-based image classification [27] and semantic segmentation architectures [34] have recently witnessed increased focus. Attention mechanisms aim at emphasizing important information and filtering irrelevant information. Hu et al. proposed a compact module to explicitly explore the relationship between channels. In their squeeze-and-excitation module, they performed global average pooling to obtain channel-wise feature response vector [7]. Liu et al. proposed an adaptively spatial feature fusion (ASFF) [14], utilizing spatial attention to optimize the feature fusion process. Wang et al. presented non-local operations [28] to capture long-range dependencies, which perform well in modelling contextual information. Woo et al. proposed two simple and effective attention modules based on channel-wise and spatial-wise attention, named Convolutional Block Attention Module (CBAM) [30] and Bottleneck Attention Module (BAM) [17], which can learn to selectively focus on the salient features in channel and spatial dimensions, and then recalibrate the intermediate features expression effectively. Oktay et al. proposed a 3D Attention U-Net architecture for abdominal organs segmentation, by integrating additive gated attention module in the skip connections of the decoder part of U-Net, which could implicitly learn to focus on more discriminant regions of the image and suppress irrelevant information [24]. While 3D deep networks [22, 23, 33] can directly leverage the inherent spatial information between slices, they are more prone to overfit, especially for small datasets. In addition, large computational burden of 3D convolutional filters limit the depth and receptive field of networks, which are two key factors for the improvement of network performance.

Recently, the Discriminative Feature Network (DFN) [35] was proposed to tackle the intra-class inconsistency and inter-class indistinction issues in most semantic segmentation methods. Automatic pancreas segmentation is a semantic segmentation task. To address the challenges of fuzzy boundaries and large shape variations in the pancreas segmentation, we design a Modified

Discriminative Feature Attention Network (MDFAN) based on DFN to explore the strengths of attention mechanism for the pancreas segmentation.

In summary, this work has the following contributions:

- We design a Discriminative Feature Attention Network to simultaneously address the intra-class inconsistency and inter-class indistinction issues of the pancreas segmentation. Quantitative evaluation on two publicly available datasets validates the effectiveness of the proposed method.
- We apply attention mechanism in our network, which can enhance the discriminative information of the pancreas structures by concentrating attention close to the pancreas, which also contributes to remove the explicit pancreas location module or network.
- We propose a lightweight Improved Refinement Residual Block (IRRB), which can model the importance of the spatial positions within each feature map and aggregate contextual information over local features.
- We propose a simple but effective post-processing method to refine the segmentation results of the proposed network.

To the best of our knowledge, this is the first attempt to segment pancreas under the guidance of attention mechanism in a 2D single-step training network with a simple post-processing.

§2 Materials and Methods

In this section, we propose a Discriminative Feature Attention Network for the pancreas segmentation. Unlike cascaded methods–pancreas localization and pancreas segmentation, the proposed network aims to utilize the attention mechanism to adaptively locate the pancreas and improve the performance and efficiency of pancreas segmentation. Our proposed method is based on the DFN proposed in [35], we utilized the modified DFN as our baseline by replacing the pretrained residual block in the backbone ResNet-101 with the pretrained dense block in the DenseNet-121, aiming at enhancing feature propagation and encouraging feature reuse. We call the modified DFN as MDFN.

2.1 Network architecture

Fig 2 shows that the proposed network has three components: one shared attention-based feature extraction branch, Smooth sub-network and Border sub-network. To improve the capabilities of feature extraction, the four denseblocks and three transitions (denseblock1 $\sim$ denseblock4, transtion1 $\sim$ transition3) from the pre-trained DenseNet121 network [5], along with BAM [17] are utilized to enhance the learning of features and obtain discriminative hierarchical features by exploiting spatial-wise and channel-wise independence. BAM is designed to explicitly learn spatial (where) and channel-wise (what) attention separately. As shown in Fig 3, BAM composes of spatial attention branch and channel attention branch. For the given input feature

map $\mathcal{F} \in R^{C\times H\times W}$, BAM infers a spatial-and-channel attention map $\mathcal{M}(\mathcal{F}) \in R^{C\times H\times W}$. The refined feature map $\mathcal{F}^{'}$ is computed as:

$$\mathcal{F}^{'} = \mathcal{F} + \mathcal{F} \otimes \mathcal{M}(\mathcal{F}) \tag{1}$$

where $\otimes$ denotes element-wise multiplication. It can be observed that the BAM is placed at each bottleneck of the proposed model to highlight features from different layers and select relevant and useful features.

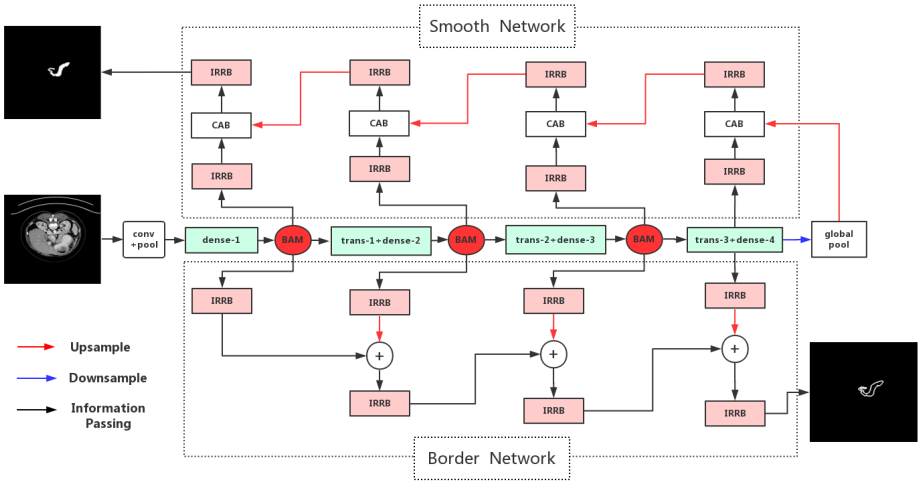


Figure 2. An overview of the proposed network (MDFAN_II). The middle part of MDFAN_II is the feature extraction stage based on the DenseNet121 and BAM. Here, ['dense-1', 'dense-2', 'dense-3', 'dense-4'], ['trans-1', 'trans-2', 'trans-3'] denote (denseblock1 $\sim$ denseblock4, transtion1 $\sim$ transition3) from the pre-trained DenseNet121 network. The top part of the model is the structure of smooth sub-network, dealing with the intra-class inconsistency issue, while the down part of the model is the border sub-network, resorting to make the bilateral features of boundary distinguishable.

As shown in the top of Fig 2, the Smooth sub-network involves the Improved Refinement Residual Block (IRRB) and Channel Attention Block (CAB). CAB aims to enhance semantic consistency of the pancreas, it infuses high-level semantic information to low-level feature maps by learning the global semantic information relationship on different channel images, and generate discriminative feature representations (as shown in Fig 4). The goal of the Smooth sub-network is to exploit the high-level features with strong consistency to guide the low-level features prediction for intra-class consistency and retain boundary information. Specifically, the channel attention block and the proposed improved refinement residual block are utilized to recalibrate the feature maps separately along channel and space according to the response of feature maps. The combination of CAB and IRRB adaptively reassign large weights to high activation regions and useful channels to enhance the intra-class consistency. However, it is still not trivial to delineate pancreas boundary due to the fuzzy boundaries of pancreas. To differentiate the features beside pancreas boundary, we employed a bottom-up Border sub-network

[35], which utilized the pancreas semantic boundary of the existing target labels to supervise and recognize the shape of pancreas. Specifically, the feature maps obtained from lower stages contain spatial details information, while those generated by higher stages with larger reception fields contain more semantic context cues. The proposed IRRB can select more discriminative spatial features to gradually help the border sub-network restore boundaries and enlarge the edge discrimination, thus reduce the impact of inter-class indistinction.

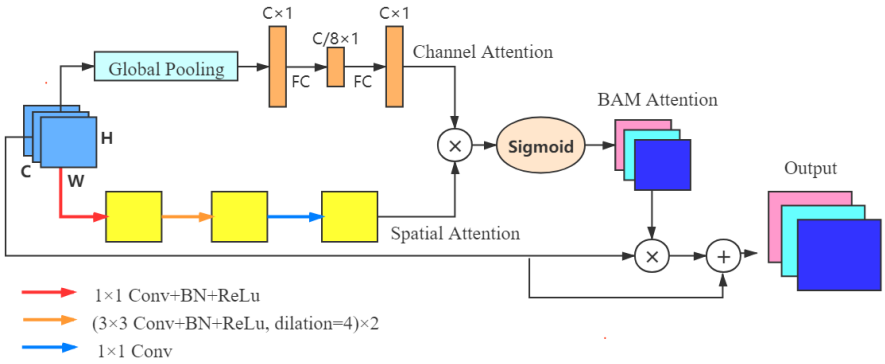


Figure 3. Detailed BAM architecture. Given the intermediate feature map, the module computes the BAM attention map through the channel attention branch and spatial attention branch.

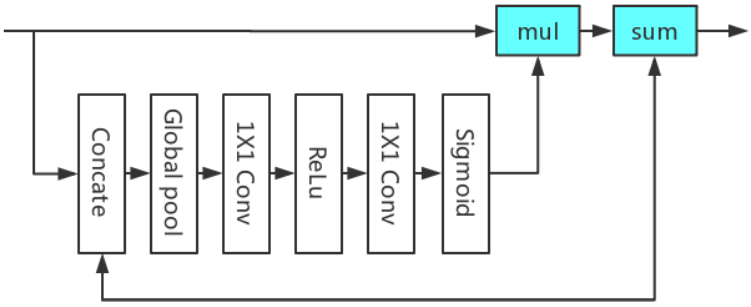


Figure 4. The structure of Channel Attention Block that utilizes channel attention to guide the selection of low-level features.

2.2 Improved Refinement Residual Block

Spatial attention mechanisms is widely used in classification and semantic segmentation [27], [34]. The goal of spatial attention is to assign large weight to target-related locations and

aggregate contextual information within each feature map. Yu et al. [35] proposed a Refinement Residual Block (RRB), which could enhance the recognition ability of each stage and refine the feature maps, as shown in Fig 5.

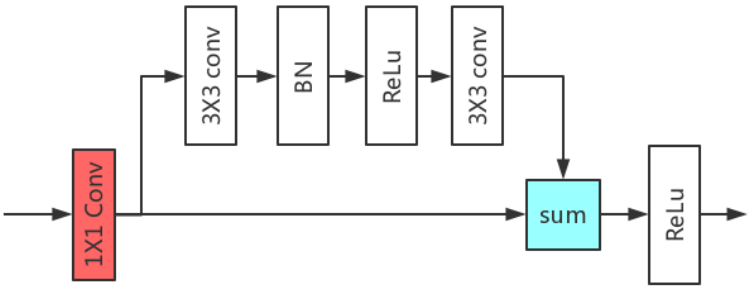


Figure 5. The structure of Refinement Residual Block.

However, we observed that the original smooth sub-network and border sub-network in [35] did not consider the spatial correlation within feature maps, which enlighten me to introduce spatial attention to the Refinement Residual Block, termed Improved Refinement Residual Block (IRRB). Fig 6 illustrates the architecture of the proposed IRRB. The IRRB consists of continuous convolution, batch normalization and ReLu layers. To exploit spatial-wise interdependencies, we first utilized two 1×1 convolution layers to gradually reduce the channels of input feature maps to 1 before sigmoid operation. Then, one 3×3 convolution, followed by BN and ReLu, as well as another 3×3 convolution are utilized to increase the receptive field and improve the awareness of contextual information within feature maps, which is helpful for the highly-varied pancreas size and position. To avoid information loss after spatial attention and speed up convergence, the residual connection is employed. In short, the output feature map of IRRB can be formulated as:

$$\mathcal{S}(\mathcal{F}) = ReLu(\mathcal{H}(\sigma(g^{1\times1}(f^{1\times1}(\mathcal{F}))) * \mathcal{F}) + (f^{1\times1}(\mathcal{F}))) \tag{2}$$

where σ denotes a sigmoid function, $f^{1\times1}$ and $g^{1\times1}$ are two convolution operation with the filter size of 1×1, H is an operation, consisting of two 3×3 convolution, BN and ReLu, $\mathcal{F}$ is an intermediate feature map. The IRRB learns a self-attention mask to enhance the targeted regions within feature maps, and then helps the network to emphasize the regions, which are more relevant to the semantic classes. For the smooth sub-network, the IRRB can attend to relevant spatial locations in the feature maps of low-level layers and gradually recover the spatial details in a top-down manner. For the bottom-up border sub-network, we gradually fusion spatial detailed features from low- to high-levels by explicitly modeling spatial-wise attention at various levels, which strengthens the semantic discrimination of high-level features with details, thus boost edge classification. Fig 7 qualitatively demonstrates that our proposed IRRB can effectively capture more detailed pancreatic features information during the decoding stage.

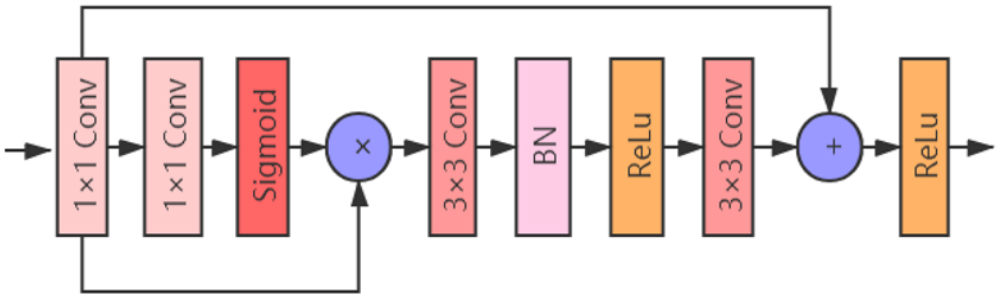


Figure 6. The structure of Improved Refinement Residual Block.

2.3 Loss function

We employed a hybrid loss based on the Dice loss and Focal loss [13] for pancreas segmentation. The aim of Dice loss is to learn the imbalanced class distribution of Smooth sub-network, which is defined as:

$$L_{Dice} = 1 - 2 * \frac{\sum_{k=1}^{N} p_k g_k + \epsilon}{\sum_{k=1}^{N} p_k + g_k + \epsilon} \tag{3}$$

Because pancreas boundaries occupy a very small region of the whole CT scan and pixels on the boundary are easy to misclassify. We adopt the Focal loss, a dynamically scaled cross entropy loss, which can adaptively reduce the contribution of easy examples during training and focus the Border sub-network on hard examples, it is defined as:

$$L_{Focal} = -\frac{1}{N} \sum_{k=1}^{N} g_k (1 - p_k)^{\gamma} log p_k \tag{4}$$

In all experiments, we use the Dice loss in conjunction with the Focal loss:

$$L = L_{Dice} + \lambda L_{Focal} \tag{5}$$

where $g_k \in \{0, 1\}$ and $p_k \in [0, 1]$ denote the manual annotations and automatic segmentations, respectively. N denotes the total number of pixels in an image and ϵ provides numerical stability to prevent division by zero. In our experiments, we trained all models with $\lambda = 0.025$ to balance the boundary Focal loss and the regional Dice loss and set $\gamma = 2.0$.

2.4 Post-processing

Many prior studies [8, 18, 37] have demonstrated that post-processing is an efficient way to improve the segmentation performance by refining the results of CNNs. Conditional random field (CRF) algorithm is widely used as a post-processing step in [8, 37]. In this work, we present a simple yet effective post-processing method to refine the predictions of the proposed network. Our post-processing is based on connected component operation. Table 1 shows that the MDFN_II can produce relatively good pancreas predictions. However, it is difficult to avoid over-segmentation due to the low contrast between pancreas and the complex surrounding tis-

sues. Moreover, the pancreas only occupies a small part of the whole abdomen and has irregular shape, which further increases the possibility of false segmentation. In order to separate the over-segmented regions, which are weakly connected with pancreas, connected component algorithm is utilized to to keep the largest connected component and reduce the false positives in the predictions. Specifically, the pancreas segmentations from the MDFN_II were post-processed by eliminating connected component comprising <20% of the total label volume. As shown in Table 1, the proposed post-processing significantly improves the average DSC and ASD of the pancreas. Here, we termed MDFAN_II with post-processing as MDFAN_III. The average inference time for post-processing per volume is 1.68 seconds.

2.5 Experimental setup

2.5.1 Data pre-processing

To quantitatively evaluate the effectiveness and generalization of the proposed model, two different abdominal CT datasets are used:

(1) A public pancreas dataset, which contains 82 contrast-enhanced abdominal CT volumes, is acquired at the National Institutes of Health Clinical Center from pre-nephrectomy healthy kidney donors or patients with neither major abdominal pathologies nor pancreatic cancer lesions [16]. The resolution of each CT volume is $512 \times 512 \times L$, where $L \in [181, 466]$ is the number of sampling slices along the long axis of the body. The slice thickness varies from 0.5 mm to 1.0 mm.

(2) The 'Beyond the Cranial Vault' (BTCV) segmentation challenge dataset (`https://www.synapse.org/#!Synapse:syn3193805/wiki/89480`) consists of 30 training data, which have annotations of all abdominal organs except duodenum, and 20 unseen testing data. The in-plane resolution of BTCV dataset varies from 0.54 mm to 0.98 mm and the slice thickness ranges from 2.5 mm to 5.0 mm. 17 patients from the 20 unseen testing data have manual annotations of eight abdominal organs, which is provided by Gibson et al.[4]. To quantitatively assess the generalization of the proposed model, we utilized 30 training data to train our proposed model, and then test the segmentation performance on the 17 testing data with annotations.

The image intensity values in a CT slice of both datasets were clipped to $[-100, 240]$ HU to filter out irrelevant information, and further normalized with zero mean and unit variance. It is important to note only axial slices are used to train our models.

2.5.2 Evaluation metrics

Five metrics including the Dice Similarity Coefficient (DSC), Jaccard index (JI), Precision, Recall and Symmetric Average Surface Distance (ASD) are used to quantitatively evaluate the segmentation performance of different methods.

- Dice Similarity Coefficient (DSC) and Jaccard index (JI) measure the volumetric overlap degree between manually labeled ground truths and network predictions. They are defined

as [3]:

$$DSC = \frac{2 \parallel V_{gt} \cap V_{seg} \parallel}{\parallel V_{gt} \parallel + \parallel V_{seg} \parallel} \tag{6}$$

$$JI = \frac{\parallel V_{gt} \cap V_{seg} \parallel}{\parallel V_{gt} \cup V_{seg} \parallel} \tag{7}$$

- Precision measures the proportion of truly positive voxels in the predictions. It is defined as:

$$Precision = \frac{\parallel V_{gt} \cap V_{seg} \parallel}{\parallel V_{seg} \parallel} \tag{8}$$

- Recall measures the proportion of truly positive voxels in the manually labeled ground truths. It is defined as:

$$Recall = \frac{\parallel V_{gt} \cap V_{seg} \parallel}{\parallel V_{gt} \parallel} \tag{9}$$

- Average Surface Distance (ASD) measures the average distance between the surface of manual and automatic segmentations [29]. It is defined as:

$$ASD = \frac{(\sum_{z \in S_{seg}} d(z, S_{gt}) + \sum_{u \in S_{gt}} d(u, S_{seg}))}{\parallel S_{gt} \parallel + \parallel S_{seg} \parallel} \tag{10}$$

where V_{gt}, V_{seg} represent the voxel sets of manual annotations and automatic segmentations, respectively, S_{gt} and S_{seg} are the corresponding surface voxel sets of V_{gt} and V_{seg}. $d(z, S_{gt})$ denotes the minimum Euclidean distance of voxel $z \in S_{seg}$ to all voxels in S_{gt}. For DSC, JI and ASD metrics, the experimental results are all reported as the mean with standard deviation over all testing samples. For precision and recall metrics, we reported the mean score over all testing samples.

2.5.3 Implementation

We implement our method based on the PyTorch platform. An Adam optimizer with initial learning rate of 0.0001 is used to train all models. For the NIH dataset, we trained our proposed method for 16 epochs under the standard 4-fold cross-validation. For the 47 patients from BTCV segmentation dataset, we utilized 30 training data to train the proposed method for 50 epochs, and tested the model performance on the remaining 17 testing data. For both datasets, the batch size is set to 4 and the learning rate is reduced by a factor of 10 every 10 epochs. All models are trained with a NVIDIA Tesla P40 GPU of 24G memory for acceleration. During training, each input image is randomly rotated ($r \in [-45°, 45°]$) and scaled ($s \in [0.9, 1.1]$) (with probability 0.5) in order to improve the generalization performance on the validation-set. The reason why we set 50 epochs for the BTCV subset is that the number of training data is smaller and the images resolution is lower, which requires more epochs to converge.

§3 Experimental results

To evaluate the proposed method, we conducted two experiments on the NIH dataset [20] and the 'Beyond the Cranial Vault' (BTCV) segmentation challenge dataset. Experimental results demonstrate that the proposed method shows consistent performance on the two datasets.

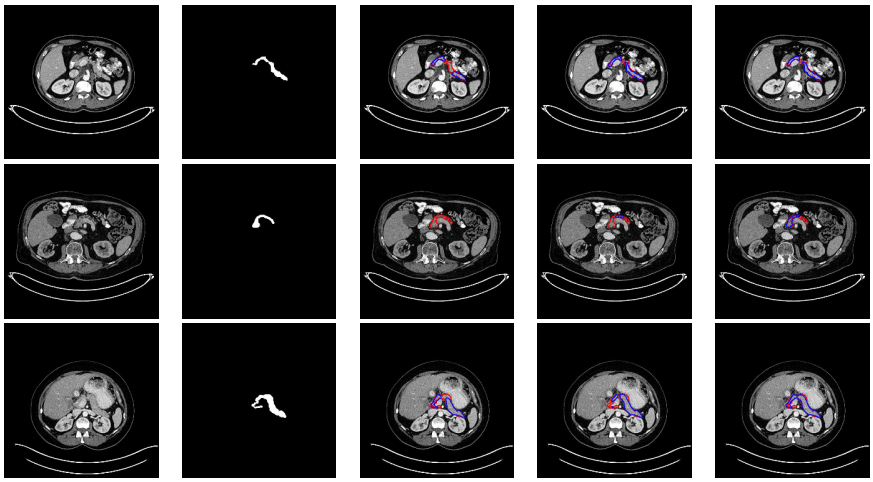
3.1 Segmentation results on the NIH dataset

To assess the effectiveness of the Bottleneck Attention Block (BAM) and the proposed Improved Refinement Residual Block (IRRB) in our method, we compared three models-MDFN, MDFAN_I, and MDFAN_II. For fair comparisons, we kept model structure and settings unchanged with only blocks being replaced or added. Fig 7 qualitatively shows the improvements brought by the Bottleneck Attention Block (BAM) and the proposed Improved Refinement Residual Block (IRRB). It is easy to note that our MDFAN_II does a better job in the pancreas localization and classification. Specifically, the comparison between third column and fourth column in Fig 7 demonstrates the Bottleneck Attention Module (BAM) can force network to pay more attention on the pancreas regions and extract more pancreas information. Similarly, the comparison between the fourth column and the fifth column validates our proposed Improved Refinement Residual Block(IRRB) can encode a wider range of contextual information into local features, which enhances pancreas features recognition capability.

The quantitative comparisons of the Precision, Recall, DSC, JI and ASD of different models are reported in Table 1. The MDFAN_II outperforms the MDFN and MDFAN_I with improvements of average DSC up to 2.03% and 1.5%. It is worth noting our proposed MDFAN_II reports the highest average Recall with 83.54%, which demonstrates the proposed IRRB can effectively filter the features spatially to get accurate saliency maps and aggregate spatial information within feature maps. Although MDFAN_II can well recognize pancreas and extract more detailed pancreas information, there inevitably exist over-segmentation. To handle the over-segmentation problem, we utilized a simple connected component detection algorithm as post-processing to refine the pancreas segmentations from MDFAN_II. The experimental results in Table 1 show the post-processing greatly improves the mean Precision, DSC, JI and ASD of MDFAN_II by 3.37%,1.6%, 2.3% and 1.9 mm, respectively. This is a result of balanced precision and recall scores, which denotes a good quality segmentation. Compared to the baseline MDFN, our final model improves the average DSC by 3.63%. Additionally, our final model takes about 1.864 seconds for each 3D scan, which consists of 0.184 seconds on the end-to-end prediction by MDFAN_II and 1.68 seconds on post-processing. Fig 8 visualizes the 3D overlap of segmentations from different models with respect to the manually labelled ground truths. Visual inspection shows MDFAN_II can capture more pancreas details and enhance the pancreas features response, and MDFAN_II with post-processing can effectively prune the over-segmentation regions to increase the average DSC and ASD measurements.

3.2 Segmentation results on the BTCV dataset

Since the NIH dataset is a widely used public dataset in previous pancreas segmentation works, to enable fair comparisons with the existing pancreas segmentation methods, the same 4-fold cross-validation was employed for evaluating the performance of the proposed method.



(a) Image (b) Label (c) MDFN (d) MDFAN_I (e) MDFAN_II

Figure 7. Examples of segmentation results of different models. Every row denotes a sample CT axial slice. (a) Image. (b) Label. (c)-(e) denote the overlap between the labels and segmentations of different models (Blue denotes segmentation results while red denotes label).

Table 1. Quantitative comparison of segmentation results of different models on the NIH dataset based on Precision, Recall, DSC, JI (%) and ASD (mm) (The best results are marked in bold).

Method	Precision	Recall	DSC	JI	ASD
MDFN	78.69	80.82	79.19 ± 7.0	66.1 ± 9.05	3.38 ± 1.67
MDFAN_I	79.49	80.9	79.72 ± 6.45	66.78 ± 8.6	3.33 ± 1.56
MDFAN_II	79.79	**83.54**	81.22 ± 6.12	68.83 ± 8.21	3.59 ± 1.59
MDFAN_III	**83.16**	83.30	**82.82 ± 6.09**	**71.13 ± 8.30**	**1.69± 0.83**

However, 4-fold cross-validation may generate relatively ideal results. To further verify the effectiveness and generalization of the proposed model, we conducted another experiment on the 47 patients from the 'Beyond the Cranial Vault' (BTCV) segmentation challenge. As shown in Table 2, compared with the baseline MDFN, the MDFN_II improve the segmentation accuracy by 1.59%, 1.96% and 2.31 mm in terms of average DSC, JI and ASD, which demonstrates the attention mechanism can enhance the feature representations and improve segmentation accuracy. Furthermore, we adopted the proposed post-processing to refine the segmentation results from MDFAN_II, in contrast to the baseline MDFN, the results significantly improved to 79.34% and 1.15 mm in terms of average DSC and average ASD, yielding increase of 5.87% and 5.22 mm respectively, which demonstrates the proposed post-processing can effectively prune the false positive regions and then achieve more robust performance. Above all, although the BTCV challenge dataset is smaller and has much lower image resolution than the NIH dataset, we still achieves comparable performance. Specifically, the experimental results on the BTCV dataset outperform the multi-stage models [2, 12] and cascaded models [19, 20, 21],

which utilized the explicit location modules or networks. In addition, our pancreas segmentation results on the BTCV challenge dataset achieve rank three in the Abdomen Leaderboard (`https://www.synapse.org/#!Synapse:syn3193805/wiki/217785`), which further demonstrates the effectiveness of the proposed method.

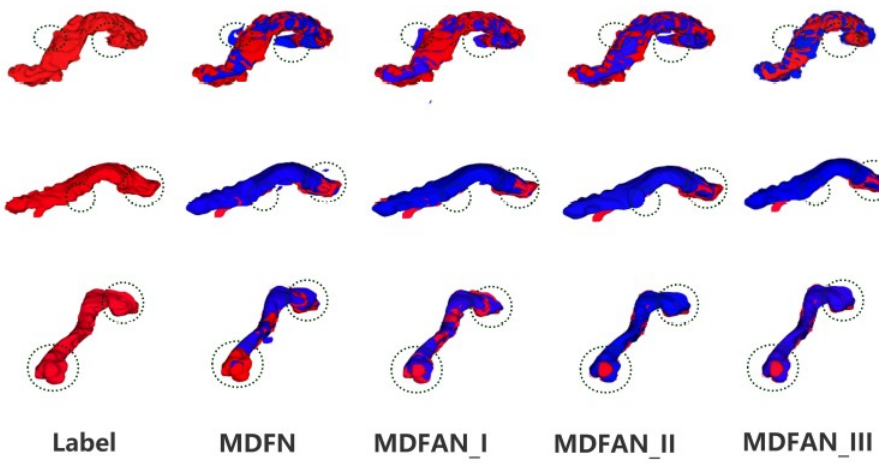


Figure 8. Examples of 3D fusion maps between predictions from different models and the ground truths, showing MDFAN_II can capture more detailed information of pancreas, and post-processing helps to prune the over-segmentation regions. The red denotes the ground truths, the blue denotes network predictions.

Table 2. Quantitative comparison of different models on the BTCV subset using DSC, JI (%) and ASD (mm) (The best results are marked in bold).

Method	DSC	JI	ASD
MDFN	73.47 ± 6.46	58.48 ± 8.33	6.37± 3.37
MDFAN_I	74.81 ± 6.76	60.24 ± 8.85	4.50 ± 2.91
MDFAN_II	75.06 ± 5.84	60.44 ± 7.64	4.06± 1.53
MDFAN_III	**79.34 ± 4.80**	**66.02 ± 6.66**	**1.15± 0.48**

3.3 Comparison with other state-of-the-art methods

We compared the final MDFAN_III (i.e. MDFAN_II with post-processing) with seven state-of-the-art pancreas segmentation methods [2, 12, 19, 20, 21, 24, 38]. To ensure fair comparisons, all methods were implemented on the NIH dataset. Note that the experimental results of other seven methods were obtained directly from their corresponding literatures. As shown in Table 3, our method achieves the average DSC of 82.82% and average JI of 71.13%, which outperforms all comparison methods. Despite the segmentation performance, the proposed method is more efficient than most comparison methods. Specifically, [2, 12, 19, 20, 21, 38] are multi-stage,

cascaded methods, which perform pancreas localization and pixel-wise classification separately, leading to low computation efficiency and generalizability. In addition, different from our simple post-processing, [20, 21] both rely on post-processing with random forest to further refine CNN's outputs. Overall, the experimental results show that our method has advantages over the coarse-to-fine methods [19, 20, 21, 38], multi-level method [2, 12]. In particular, compared to 3D method [24], our proposed method achieved slightly better segmentation performance in average DSC, which is a good proof of the effectiveness of our proposed MDFAN_III.

Table 3. Comparison of the DSC and JI results (%) with the state-of-the-art pancreas segmentation methods on the NIH dataset (The best results are marked in bold).

Method	Min DSC	Max DSC	Mean DSC	Mean JI	Protocol
Roth et al. [19]	23.99	86.29	71.42±10.11	N/A	CV-4
Roth et al. [20]	34.11	88.65	78.01±8.20	N/A	CV-4
Roth et al. [21]	50.69	88.96	81.27±6.27	68.87±8.12	CV-4
Li et al. [12]	N/A	N/A	78.9	65.4	CV-10
Asaturyan et al. [2]	**72.8**	86.0	79.3±4.4	65.7	CV-4
Zhou et al. [38]	62.43	**90.85**	82.37±5.68	N/A	CV-4
Oktay et al. [24]	N/A	N/A	82.1±5.7	N/A	CV-4
MDFAN_III	51.88	89.44	**82.82±6.09**	**71.13± 8.30**	CV-4

§4 Discussion

The pancreas is an important digestive organ in the abdomen, which plays a significant role in the decomposition and absorption of blood sugar and nutrients. Accurate pancreas segmentation can provide useful information for clinicians. To address the inefficiency of coarse-to-fine methods and unclear boundaries in the pancreas segmentation, we introduce attention mechanism to realize implicit localization for the pancreas, and propose a composite loss to force network pay more attention on boundary pixels. To the best of our knowledge, the proposed algorithm outperformed all 2D pancreas segmentation approaches on the NIH dataset under 4-fold cross-validation without the help of explicit pancreas localization, which demonstrates channel-wise and spatial attention can implicitly localize and highlight the pancreas regions, and thus enhance the representation of pancreas features. What's more, Table 3 shows the proposed algorithm outperformed the 3D attention model [24] in term of average DSC, which indicates the attention mechanism can automatically aggregate the contextual information over local features, and then utilize spatial context to capture pancreas features, and thus improve the performance of network. Overall, the proposed algorithm not only keeps a high segmentation accuracy on the pancreas, but also improve the efficiency of pancreas segmentation.

In order to gain a better understanding of the Bottleneck Attention Block (BAM) and the proposed Improved Refinement Residual Block (IRRB), we conducted the same post-processing on the baseline MDFN, the experimental results are reported in Table 4. As shown in Table 4, under the same post-processing, the MDFAN_II improves the average DSC, JI and ASD

Table 4. Quantitative comparison of the post-process on the baseline MDFN and the proposed MDFAN_II based on DSC, JI (%), average ASD (mm) and run time (s) (The best results are marked in bold).

Method	DSC	JI	ASD	Run time
MDFN	80.36 ± 7.78	67.87 ± 10.48	1.97 ± 1.18	**1.61** s
MDFAN_II	**82.82 ± 6.09**	**71.13 ± 8.30**	**1.69 ± 0.83**	1.68 s

Table 5. Quantitative comparison of the MDFAN_II with cross entropy loss (abbreviate as CE) and focal loss (abbreviate as FL) on DSC, JI (%) and average ASD (mm) (The best results are marked in bold).

Method	DSC	JI	ASD
MDFAN_II+CE	81.20 ± 6.20	68.82 ± 8.27	3.77 ± 1.82
MDFAN_II + FL	**81.22 ± 6.12**	**68.83 ± 8.21**	**3.59 ± 1.59**

by 2.46%, 3.26% and 0.28 mm over the baseline MDFN, which demonstrates the combination of the Bottleneck Attention Block (BAM) and the proposed Improved Refinement Residual Block (IRRB) can effectively improve the segmentation accuracy. In addition, to validate the effectiveness of focal loss used to guide Border sub-network training, we conducted another comparison experiment between the focal loss and cross entropy loss under the same architecture MDFAN_II, as well as the regional Dice loss. As shown in Table 5, the MDFAN_II with focal loss improve the overall performance in terms of the average DSC, JI and ASD, especially for ASD, which demonstrates the modulation factor in focal loss [13] can force network focus on hard samples, such as boundary pixels, to better delineate pancreas boundary.

There are several limitations in this study. First, over-segmentation exists in the predictions from the MDFAN_II, this is mainly because attention mechanism may suffer from semantic confusion due to the highly similarity in intensity between target pancreas and surrounding organs and tissues. Next, we will consider how to design more discriminative attention modules to effectively locate the pancreas and reduce the interference of background. Second, as shown in Table 2, there still have space to improve the generalization of the proposed algorithm on different dataset, such as the BTCV dataset. Since the number of training set in the BTCV dataset is small and the resolution of images is low, the model with large numbers of parameters is prone to overfit, and then degrade network performance, which pushes us to consider an adaptive regularization technique in our future works.

§5 Conclusion

Accurate delineation of pancreas can assist doctors in the diagnosis of pancreas diseases. In this paper, we propose a single-stage Discriminative Feature Attention Network for the pancreas segmentation. Our method has two advantages: 1) we integrate channel-wise and spatial-wise attention into the baseline MDFN to enhance feature extraction and eliminate the necessity of

using explicit pancreas localization modules. 2) we adopt a simple yet effective post-processing to refine the segmentation results. The experimental results show our network can effectively handle the issues of intra-class inconsistency and inter-class indistinction in the pancreas segmentation. Because the proposed method is a single-step end-to-end training framework with simple post-processing, it is simple to implement. Above all, the proposed method achieves consistently experimental results on the two pancreas datasets, which demonstrates the effectiveness and generalization of our proposed method.

References

[1] R R Almeida, G C Lo, M Patino, B Bizzo, D V Sahani. *Advances in Pancreatic CT Imaging*, AJR Am J Roentgenol, 2018, 211(1): 1-15.

[2] H Asaturyan, A Gligorievski, B Villarini. *Morphological and multi-level geometrical descriptor analysis in CT and MRI volumes for automatic pancreas segmentation*, Comput Med Imaging Graph, 2019, 75: 1-13.

[3] L R Dice. *Measures of the amount of ecologic association between species*, Ecology, 1945, 26(3): 297-302.

[4] E Gibson, F Giganti, Y Hu. *Automatic multi-organ segmentation on abdominal CT with dense V-networks*, IEEE Trans Med Imaging, 2018, 37(8): 1822-1834.

[5] G Huang, Z Liu, L Van Der Maaten, K Q Weinberger. *Densely connected convolutional networks*, In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 2261-2269.

[6] M X Huang, C F Huang, J Yuan, D X Kong. *Fixed-Point Deformable U-Net for Pancreas CT Segmentation*, In Proc. ISICDM, 2019, 283-287.

[7] J Hu, L Shen, G Sun. *Squeeze-and-excitation networks*, In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 7132-7141.

[8] K Kamnitsas, et al. *Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation*, Med Image Anal, 2016, 36: 61-78.

[9] K Karasawa, M Oda, T Kitasaka, K Misawa, M Fujiwara, C Chu, G Zheng, D Rueckert, K Mori. *Multi-atlas pancreas segmentation: Atlas selection based on vessel structure*, Med Image Anal, 2017, 39: 18-28.

[10] H Kumar, S V DeSouza, M S Petrov. *Automated pancreas segmentation from computed tomography and magnetic resonance images: A systematic review*, Comput Methods Programs Biomed, 2019, 178: 319-328.

[11] F Y Li, W S Li, Y C Shu, S Qin, B Xiao, Z W Zhan. *Multiscale receptive field based on residual network for pancreas segmentation in CT images*, Biomed Signal Process Control, 2020, 57: 101828-101840.

[12] S Li, H Jiang, Z Wang, G Zhang, Y Yao. *An effective computer aided diagnosis model for pancreas cancer on PET/CT images*, Comput Methods Programs Biomed, 2018, 165: 205-214.

[13] T Y Lin, P Goyal, R Girshick, K He, P Dollar. *Focal loss for dense object detection*, In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 2999-3007.

[14] S Liu, D Huang, Y Wang. *Learning spatial fusion for single-shot object detection*, arXiv preprint arXiv:1911.09516, 2019.

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

[16] M Oda, N Shimizu, K Karasawa. *Regression forest-based atlas localization and direction specific atlas generation for pancreas segmentation*, In: International conference on medical image computing and computer-assisted intervention, Springer, Berlin, 2016, 556-563.

[17] J Park, S Woo, J Lee, I Kweon. *A Simple and Light-Weight Attention Module for Convolutional Neural Networks*, International journal of computer vision, 2020, 128(4): 783-798.

[18] S Pereira, A Pinto, V Alves, C A Silva. *Brain tumor segmentation using convolutional neural networks in MRI images*, IEEE Trans Med Imaging, 2016, 35(5): 1240-1251.

[19] H R Roth, L Lu, A Farag, H C Shin, J Liu, E B Turkbey, R M Summers. *Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation*, In: International conference on medical image computing and computer-assisted intervention, Springer, Berlin, 2015, 556-564.

[20] H Roth, L Lu, A Farag, A Sohn, R Summers. *Spatial aggregation of holistically-nested networks for automated pancreas segmentation*, In: International conference on medical image computing and computer-assisted intervention, Springer, Berlin, 2016, 451- 459.

[21] H Roth, L Lu, N Lay, A P Harrison, A Farag, A Sohn, R M Summers. *Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation*, Med Image Anal, 2018, 45: 94-107.

[22] H R Roth, M Oda, N Shimizu, H Oda, Y Hayashi, T Kitasaka, M Fujiwara, K Misawa, K Mori. *Towards dense volumetric pancreas segmentation in CT using 3D fully convolutional networks*, Progress in Biomedical Optics and Imaging-Proceedings of SPIE, 2018, 105740B-105740B-6.

[23] H R Roth, H Oda, X Zhou, N Shimizu, Y Yang, Y Hayashi, M Oda, K Mori. *An application of cascaded 3D fully convolutional networks for medical image segmentation*, Comput Med Imag Graph, 2018, 66: 90-99.

[24] J Schlempera, O Oktaya, M Schaapb, M Heinrichc, B Kainza, B Glockera, D Rueckerta. *Attention gated networks: learning to leverage salient regions in medical images*, Med Image Anal, 2019, 53: 197-207.

[25] V A Sindagi, V M Patel. *HA-CCN: Hierarchical Attention-Based Crowd Counting Network*, IEEE Trans Image Process, 2020, 29: 323-335.

[26] T Tong, R Wolz, Z Wang, Q Gao, K Misawa, M Fujiwara, K Mori, J Hajnal, D Rueckert. *Discriminative dictionary learning for abdominal multi-organ segmentation*, Med Image Anal, 2015, 23: 92-104.

[27] F Wang, M Jiang, C Qian, S Yang, C Li, H Zhang, X Wang, X Tang. *Residual Attention Network for Image Classification*, In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 6450-6458.

[28] X Wang, R Girshick, A Gupta, K He. *Non-local neural networks*, In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 7794-7803.

[29] S Wang, K He, D Nie, S Zhou, Y Gao, D Shen. *CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation*, Med Image Anal, 2019, 54: 168-178.

[30] S Woo, J Park, J Lee, I Kweon. *CBAM: Convolutional Block Attention Module*, In Proc Eur Conf on Computer Vision, 2018, 3-19.

[31] R Wolz, C Chu, K Misawa, M Fujiwara, K Mori, D Rueckert. *Automated abdominal multi-organ segmentation with subject-specific atlas generation*, IEEE Trans Med Imaging, 2013, 32(9): 1723-1730.

[32] J, Wu. *2D MRI Pancreas Segmentation based on Transfer Learning*, Dissertation, Xidian University, 2014.

[33] J Xue, K He, D Nie, E Adeli, Z Shi, S Lee, Y Zheng, X Liu, D Li, D Sheng. *Cascaded MultiTask 3-D Fully Convolutional Networks for Pancreas Segmentation*, IEEE Trans Cybern, 2019, 1-13.

[34] X Yang, X Wang, Y Wang, H Doub, S Li, H Wen, Y Lin, P Heng, D Ni. *Hybrid attention for automatic segmentation of whole fetal head in prenatal ultrasound volumes*, Comput Methods Programs Biomed, 2020, 194: 105519-105528.

[35] C Yu, Y Wang, C Peng, C Gao, G Yu, N Sang. *Learning a discriminative feature network for semantic segmentation*, In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 1857-1866.

[36] S Zhou, D Nie, E Adeli, J Yin, J Lian, D Shen. *High-Resolution Encoder-Decoder Networks for Low-Contrast Medical Image Segmentation*, IEEE Trans Image Process, 2020, 29: 461-475.

[37] X Zhao, Y Wu, G Song, Z Li, Y Zhang, Y Fan. *A deep learning model integrating FCNNs and CRFs for brain tumor segmentation*, Med Image Anal, 2018, 43: 98-111.

[38] Y Zhou, L Xie, W Shen, Y Wang, E K Fishman, A L Yuille. *A fixed-point model for pancreas segmentation in abdominal CT scans*, In: International conference on medical image computing and computer-assisted intervention, Springer, Berlin, 2017, 693-701.

[1]The School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China.
Email: 1228561480@qq.com

[2]The Department of Mathematics, Zhejiang University, Hangzhou 310027, China.
Email: 11735032@zju.edu.cn, dxkong@zju.edu.cn

[3]The School of Mathematics and Statistics, Xidian University, Xi'an 710069, China.
Email: jyuan@xidian.edu.cn