Computers & Graphics 90 (2020) 11–20

Contents lists available at ScienceDirect

Computers & Graphics

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

Special Section on SMI 2020

ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation

Chen Li[a], Yusong Tan[a,*], Wei Chen[a], Xin Luo[a], Yulin He[a], Yuanming Gao[a], Fei Li[b]

[a] College of Computer, National University of Defense Technology, Changsha 410073, China
[b] Computer Network Information Center, Chinese Academy of Sciences, China

ARTICLE INFO

Article history:
Received 4 May 2020
Accepted 5 May 2020
Available online 15 May 2020

Keywords:
Attention mechanism
Nested U-Net
Medical image segmentation
Model pruning

ABSTRACT

Organ cancer have a high mortality rate. In order to help doctors diagnose and treat organ lesion, an automatic medical image segmentation model is urgently needed as manually segmentation is time-consuming and error-prone. However, automatic segmentation of target organ from medical images is a challenging task because of organ's uneven and irregular shapes. In this paper, we propose an attention-based nested segmentation network, named ANU-Net. Our proposed network has a deep supervised encoder-decoder architecture and a redesigned dense skip connection. ANU-Net introduces attention mechanism between nested convolutional blocks so that the features extracted at different levels can be merged with a task-related selection. Besides, we redesign a hybrid loss function combining with three kinds of losses to make full use of full resolution feature information. We evaluated proposed model on MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge Dataset and ISBI 2019 Combined Healthy Abdominal Organ Segmentation (CHAOS) Challenge. ANU-Net achieved very competitive performance for four kinds of medical image segmentation tasks.

1. Introduction

For a long time, human beings have been plagued by malignant tumors and other organ diseases, and their lives and health have been seriously threatened. Cancer is the leading cause of death, which pose a huge threat to human health. Millions of people died from cancer every year in the world. However, early detection and treatment are still the main methods to increase cancer survival. Identifying the position of organ is a preparation step in diagnosing and plays an indispensable role in disease treatment. Generally speaking, locating the position of an organ from medical images is an medical image segmentation task.

The analysis of research on medical image segmentation tasks is as follows:

1. Clinically, medical images can not only detect a variety of lesions, but also assist doctors in selecting individual treatment options for different patients;
2. The patient's response to the treatment arrangement and the recovery condition are displayed more intuitively in medical images;
3. The treatment progress is customized according to personal needs to reduce the treatment cost, to achieve precise medical treatment and personalized treatment. Finally, it can reduce the psychological burden and economic burden of the patient.

Therefore, the researches and applications of medical image segmentation are of great significance.

Generally, researches on medical image segmentation can be categorized in to two classes: (1)manual and semi-automatic segmentation, (2)automatic segmentation. Manual segmentation largely relies on experts with advanced technical skills to perform such task. It is an extremely challenge to distinguish between human organs and tissues by observation of medical images. Besides, the subjective judgment of experts will seriously affect the result of manual segmentation. The quality of the segmentation relies heavily on the judgment of experts. As a consequence, these factors lead to the poor practicality of manual segmentation in the field of medical image segmentation. Meanwhile, semi-automatic segmentation still requires manual intervention, which leads to biases and errors. Therefore, automated medical image segmentation has become a preferred choice in this field and has been extensively studied.

In recent years, with the leaps in computing power brought by GPUs and the rapid development of deep learning technolo-

* Corresponding author.
E-mail address: ystan@nudt.edu.cn (Y. Tan).

https://doi.org/10.1016/j.cag.2020.05.003
0097-8493/

gies, the field of image analysis has also developed rapidly. Convolutional neural networks (CNNs) have achieved great success in image classification[2,1], semantic segmentation[3,4], object detection[5,6], and reconstruction[7,8]. Image semantic segmentation is one of the basic researches in the field of computer vision. It is also an important part of image understanding. The natural image segmentation technology based on CNNs has gradually matured. At the same time, the medical image segmentation technology is still a very challenging problem because medical images have characteristics such as gray unevenness, large contrast changes, and noise.

Since the fully convolutional neural network (FCN) [9] was proposed, image semantic segmentation has not only achieved significant breakthroughs in the field of natural image processing, but also has developed rapidly in medical image segmentation. In clinical researches, image semantic segmentation technology can accurately segment target organs and diseased tissues from medical images in a fully automatic manner. And it is also the most ideal way to assist physicians in personalized precise treatment, so as to be "the right medicine."

A large number of convolutional neural networks represented by variants of FCN and U-Net [10] have gradually been applied in the field of medical image segmentation. For instance, Li et al. [11] present a H-Dense UNet for liver and tumor segmentation through hybrid feature fusion layer. Christian et al. [12] proposed a pipeline of two fully convolutional networks for automatic multi-label heart segmentation from CT and MRI volume. Liao et al. [13] proposed a 3D deep neural convolutional network to find all malignant nodules from CT images for assessing lung cancer.

However, these methods usually divide the segmentation task into two steps: localization and segmentation. The extra localization step will increase the amount of model parameters and bring extra time consumption. In addition, the accuracy of model segmentation also depends heavily on the first step positioning accuracy. In clinical practice, even minor errors in medical image segmentation may lead doctors to diagnose patients mistakes.

Therefore, the current automatic segmentation technology for medical images is still not advanced enough. And the precise medical images segmentation is still challenging and demanding. In this field, there is still much room for improvement in deep learning segmentation models.

To address the need for more accurate segmentation result in medical images, we propose ANU-Net, a reliable segmentation model based on nested U-Net architecture and attention mechanism. The contributions of our work can be listed as follows:

1. ANU-Net is introduced for medical image segmentation;
2. The experiments conducted on public dataset (LiTS, CHAOS) show that Attention mechanism can focus on target organ of the whole image while suppressing the irrelevant tissue;
3. ANU-Net has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task;
4. ANU-Net redesigns nested U-Net architecture and then integrates features at different levels, which brings improved performance on various medical image segmentation tasks compared to other UNet-based models;
5. Due to the introduction of deep supervision, ANU-Net has a flexible network structure, which can perform pruning operations during testing. So a large amount of parameters in the pruned ANU-Net can be greatly reduced and model is accelerated at the cost of little performance.

Our paper is organized as follows: Section 2 will briefly review related works. Section 3 will detail our segmentation methodology including network architecture, deep supervision and model pruning. Then we illustrate the experimental setups in Section 4. After that, the experimental results will be displayed and analyzed. Conclusions and future works are given in Section 5.

2. Related works

2.1. Image semantic segmentation

Deep learning is dramatically driving the development of image analysis, leading to a series of state-of-the-art in image analysis tasks including classification[14–17], object detection[18–20], semantic segmentation[9,21,22] etc. Image semantic segmentation is an active research field. The goal of segmentation is fundamentally different from classification or other image analysis tasks. Image semantic segmentation is to classify each pixel in the image individually, which is fueled by different challenging and provided datasets in the field of image segmentation[23–25].

Since He et al [2] proposed a deep residual model ResNet, deep Convolutional Neural Networks (DCNNs) methods can use identity mapping to overcome the vanishing gradient problem during the training gradient, thereby simplifying and facilitating the training process. DCNNs methods provide superior performance for natural image segmentation [9] in relevant competitions and exploit the state-of-the-art detectors, such as Faster R-CNN [5], R-FCN [16], to get the region of each instance, and then predict the mask for each region. SegNet consists of encoder and decoder, encoding network is a 13-layer VGG16 network [1], and decoding network uses pixel-wise classification layers. Pinheiro et al [26] proposed DeepMask to segment and classify the center object in a sliding window fashion. He et al. [27] proposed Mask R-CNN that is built on the top of Faster R-CNN by adding an instance-level semantic segmentation branch. Based on Mask R-CNN, Chen et al [28] proposed MaskLab that used position-sensitive scores to obtain better results.Mask Scoring R-CNN [29] brings consistent and noticeable gain with different models, and outperforms the state-of-the-art Mask RCNN.

However, compared to natural image segmentation, there are fewer deep learning models that have been proposed specifically for the medical image segmentation, as they consider data insufficiency and class imbalance problems. Besides, duo to the formation of medical images is susceptible to noise and tissue movement, the images have characteristics such as blur and non-uniformity. In practice, there is a wide variety of medical imaging modalities used for the purpose of clinical diagnosis and in most cases the images look similar. These characteristics make medical images hard to segment precisely and increase the difficulty in diagnosis of clinical diseases. Moreover, compared to natural images, the collection of medical image data is more difficult, and the training of deep neural networks requires a large amount of data, so training deep neural networks with strong generalization ability is still challenging.

To tackle these difficulties, many segmentation methods have been proposed for medical image segmentation, including intensity thresholding, region growing, and deformable models. These traditional methods, however, rely on hand-crafted features, and have limited feature representation capability. So more automated segmentation model is demanded. Recently, fully convolutional neural networks (FCNs) have achieved great success on a broad array of recognition problems[30–33]. Many researchers advance this stream using FCN based methods in the medical image segmentation problem and these researches can be classified into two categories broadly:

1. 2D FCNs, such as U-Net architecture [10] and its variants, the multi-channel FCN [34], and the FCN based on backbone such as VGG-16 [35].
2. 3D FCNs, where 2D convolution layers are replaced by 3D convolution layers with volumetric data input[36,37].

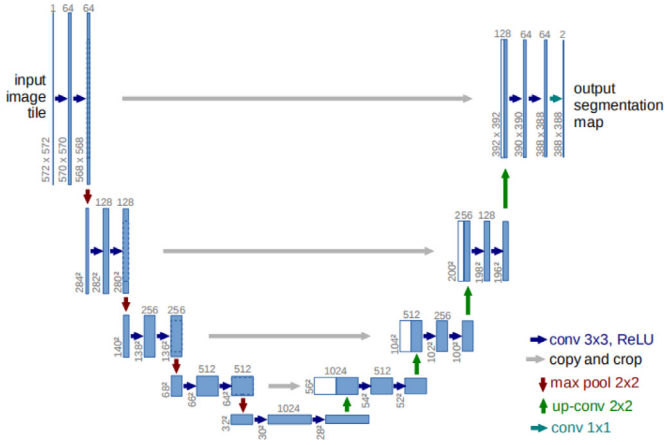

Fig. 1. Diagram of U-Net structure proposed by Ronneberger et al. [10].

2.2. U-Net architecture and variants

One of the most popular approaches for semantic medical image segmentation is U-Net. A diagram of the basic U-Net architecture is shown in Fig. 1. U-Net is proposed for automatic medical image segmentation where the network consists of symmetrical encoder and decoder. This architecture synthesizes the significant information by minimizing a cost function in the encoder and construct an image mask at the decoder.

According to the structure, the network consists of two parts: the convolutional encoder and decoder. In the The basic convolution operations are performed followed by ReLU activation in both parts of the network. For downsampling in the encoder, 2x2 max-pooling operations are performed. Downsampling is necessary for the segmentation network because it can increase the robustness to the disturbance of the input image, such as image translation, rotation, etc., thereby it can reduce the risk of overfit and the amount of computation, and increase the size of the receptive field. In the decoder, the convolution transpose (deconvolution) operations are performed to upsample the feature maps. Upsampling can restore the abstract features and decode them to the size of the original image, and finally obtain the segmentation result. In addition, the skip connection between encoder and decoder ensure that U-Net preserves the full context of the input images, which is a major advantage when compared to patch-based segmentation approaches[38,10]. So the U-Net performs well medical image segmentation task.

Due to its outperformance, U-Net is not only limited to the applications in the domain of biomedical image, nowadays this architecture is massively applied for computer vision tasks as well[39,40]. Meanwhile, different variants of U-Net architecture have been proposed. ResUNet-a [4] consists of a U-Net backbone and a novel loss function based on the Dice loss function, combining with residual connections, pyramid scene parsing pooling and multi-tasking inference. H-DenseUNet [11] is a hybrid densely connected U-Net, which consists of 2D and 3D dense-parts for efficiently extracting features for segmentation. R2U-Net [41] utilize U-Net architecture, residual unit and RCNN, which structure is shown in Fig. 2. Under same number of network parameters, R2U-Net performs better for medical image segmentation when compared to basic U-Net model.

2.3. Attention mechanism

Attention mechanism originates from human visual cognitive science. Due to the bottleneck of information processing, different parts of the human retina have different information processing capabilities, the central recessed part has the strongest ability while the other parts weaken in turn. In order to make reasonable use of limited visual information processing resources, the human cognitive system will choose to focus on a specific part of all information collected, and then focus on it, while ignoring other useless information.

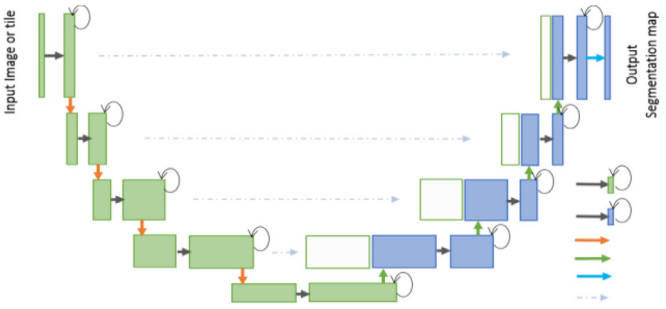

Fig. 2. Diagram of R2U-Net structure proposed by Alom et al. [41].

Attention mechanism firstly emerged in the natural language processing(NLP) and quickly gained dominance. It was Non-local [42] proposed by He's team that first introduced the attention mechanism to computer vision. Since then, [3] combined the shared network to the field of semantic segmentation. RA-UNet [43] combined with residual and attention mechanism to obtain a deep network. Attention Mechanism is currently very popular and widely used in many fields such as machine translation [44], speech recognition [45], and computer vision[43,46]. Attention Mechanism is usually used as a unit to improve the performance of the Encoder-Decoder architecture.

The reason why Attention Gate is so popular is that attention mechanism gives network the ability to distinguish and focus. For example, in speech recognition applications, each word in a sentence is given a different weight, making the learning of network models become more soft and targeted. Meanwhile, Attention itself can be used to explain the alignment relationship between input and output data and explain what the model has learned, thus providing a window for us to open the black box of deep learning.

3. Methodology

3.1. Nested U-Net model

Most researches used U-Net as the backbone and made some changes for different segmentation tasks in the past. UNet++ proposed by Zhou et al. [47] is one of the most representative UNet-based segmentation architectures. Our work is based on its structure, improving the loss function (Section. 3.4) and changing the dense connections between nested convolutional blocks (Section. 3.2).

Nested U-Net architecture is inspired by DenseNet [48] and then integrates a series of U-Nets with different depths. What distinguishes our nested architecture from U-Net is that the former redesigned dense skip connections between encoder and decoder at different depth and used nested convolutional blocks. Each nested convolutional block in nested U-Net extracts semantic information by several convolution layers. And every convolution layer in the block is connected through dense skip connections so that concatenation layer can fuse different levels semantic information. Our improved nested architecture brings following benefits:

1. Our nested architecture can learn the importance of features at different depths by itself, thereby avoiding complex selection of deep and shallow features.

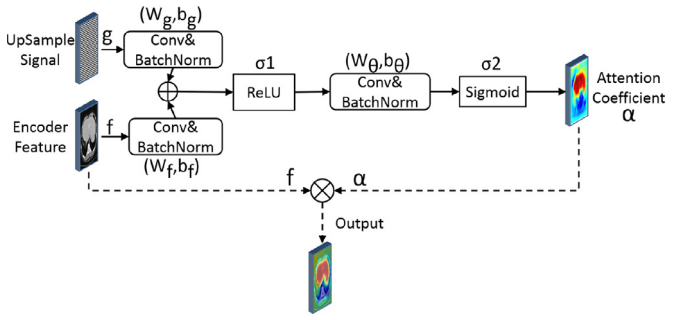

Fig. 3. Diagram of Attention Gate.

2. Our nested architecture shares a feature extractor, so there is no need to train the series of U-Nets, but only an encoder.
3. In our nested architecture, different levels of features are restored by different decoder paths separately, so we can obtain hierarchical decoded masks from different levels.

3.2. Attention Gate(AG)

In order to focus on locations that are relevant for target organ, we refer to the method proposed by PASSRnet [49] and add a simple but effective Attention Gate to the our nested architecture (Section 3.1). The architecture of Attention Gate is shown in Fig. 3.

The detailed analysis of the Attention Gate is as follows:

1. Attention Gate has two inputs: (1) the upsampling feature (g) in the decoder and (2) the corresponding depth feature (f) in the encoder. The first input (g) is used as the gating signal to enhance the learning of the second input (f). In a word, this gating signal (g) can select more useful feature from encoded feature (f) and send it to the upper decoder.
2. After convolutional operation (W_g, W_f) and BatchNorm (b_g, b_f), these two inputs are merged pixel by pixel.
3. Then the added result will be sent to Rectified Linear Unit (ReLU, $\sigma_1(x) = max(0, x)$) for the activation.
4. After activation, the features will get convolutional operation (W_θ) and BatchNorm (b_θ) again.
5. The S-shaped activation function sigmoid ($\sigma_2(x) = \frac{1}{1+e^{(-x)}}$) is selected to train the convergence of the parameters in the Gate and to get the attention coefficient (α).
6. The output can be obtained by multiplying the encoder feature by coefficient α pixel by pixel.

The process of feature selection in Attention Gate can be formulated as follows:

$$F = \sigma_1\left[\left(W_f^T \times f + b_f\right) + \left(W_g^T \times g + b_g\right)\right] \tag{1}$$

$$\alpha = \sigma_2\left(W_\theta^T \times F + b_\theta\right) \tag{2}$$

$$output = f \times \alpha \tag{3}$$

The Attention Gate has a good function selection function and can enhance the learning of the target area related to the segmentation task while suppressing the area irrelevant in the task. Therefore, our work integrates Attention Gate into the novel proposed network to improve the efficiency of propagating semantic information through skip connections.

3.3. Attention-based nested U-Net

We design an integrated network based on **A**ttention mechanism and **N**ested **U-Net** architecture, called ANU-Net for medical image segmentation. The high-level overview of the ANU-Net is shown in Fig. 4.

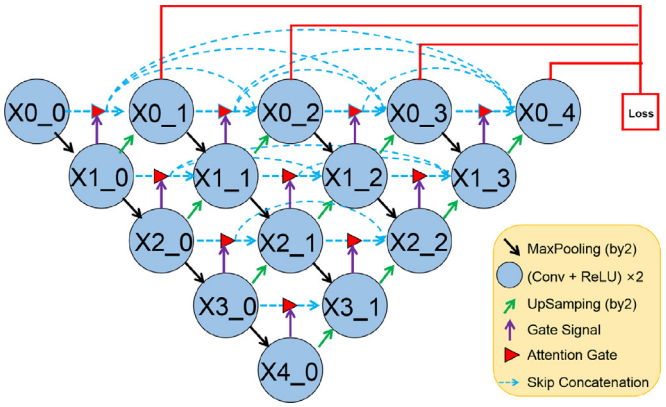

Fig. 4. Structure of ANU-Net.

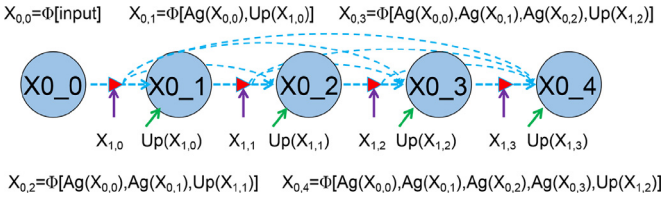

Fig. 5. Detailed analysis of dense skip connections between the nested convolution blocks in first layer.

As we can see, ANU-Net uses nested U-Net as the basic network framework, where encoder and decoder are symmetrically arranged on both sides of the network. The context information extracted by the encoder is propagated to the decoder of the corresponding layers through the dense skip connections, so that more efficient hierarchical features can be extracted.

With dense skip connections, the inputs of each convolution block in the decoder consists of two equal-scale feature maps: (1) The intermediate feature maps come from the outputs of prior Attention Gates along skip connection at the same depth; (2) The final feature map comes from the output of deeper block deconvolution operation. After all features maps received and concatenated, the decoder restores features in a bottom-up manner. The reason why all prior feature maps accumulate and arrive at the current block is that dense skip connections can make full use of these feature maps from prior nested convolution blocks in this layer.

We define the feature map as follows, let $X_{i,j}$ represents the output of convolution block where i denotes the feature depth in network and j denotes the sequence of the convolution block in the ith layer along the skip connections, so extracted feature map of in ANU-Net can be formulated as follows:

$$X_{i,j} = \begin{cases} \Phi\left[X_{i-1,j}\right] & , j = 0 \\ \Phi\left[\sum_{k=0}^{j-1} Ag(X_{i,k}), Up\left(X_{i+1,j-1}\right)\right] & , j > 0 \end{cases} \tag{4}$$

where $\Phi[\,]$ denotes the convolution block followed by concatenation merger, $Up()$ and $Ag()$ mean upsampling and the attention gate selection respectively, $\sum_{k=0}^{j-1} Ag(X_{i,k})$ denotes that concatenate the output results of the Attention Gates from node $X_{i,k=0}$ to $X_{i,k=j-1}$ in the ith layer.

After the concatenation operation, convolution blocks in the decoder will only learn to use the selected same-scale feature maps from encoder instead of all collected feature maps along dense skip connections. Fig. 5 further explains Eq. (4) by detailed analysing how the feature maps travel through the skip connection in the first layer. For example, block $X_{0,4}$ owns five inputs, of which four

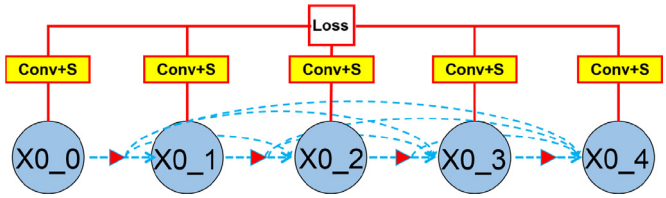

Fig. 6. Diagram of deep supervision in ANU-Net.

inputs are the outputs of the previous j blocks in this layer and another input is the unsampling feature from block $X_{1,3}$ in second layer.

Two of the main innovations in ANU-Net is: The network transfer extracted feature from the encoder to the decoder through dense skip connections for integration of hierarchical representations. Besides, Attention Gate is added between nested convolution blocks, so that the features extracted at different layers can be merged with a focused selection in the decoder path. As a consequence, the accuracy of ANU-Net should be improved.

3.4. Deep supervision

It was Lee et al. that firstly proposed deep supervision in DSN [50], where proved that deep supervision can alleviate the vanishing gradient problem and accelerate convergence speed. In addition, deep supervision can also assist loss function to play the role of regularization. In order to introduce deep supervision, ANU-Net adds 1x1 convolutional layer and sigmoid activation function after every output block($X_{0\_1}, X_{0\_2}, X_{0\_3}, X_{0\_4}$) in the first layer. Besides, ANU-Net connects these layers directly to the final output for calculation of the loss and back propagation of the error. Fig. 6 shows the structure of deep supervision designed in ANU-Net.

Because of dense skip connections designed between nested convolution blocks, ANU-Net obtains full resolution feature maps at different semantic levels from blocks. In order to make full use of these semantic information, we redesign a hybrid loss function combining with soft dice coefficient loss(DICE), focal loss(FOCAL) and pixel-wise binary cross entropy loss(BCE).

Dice coefficient is derived from binary classification and is essentially a measure of the overlap between two samples. The indicator range is [0,1], where 1 means complete overlap and 0 means no overlap. The calculation formula is:

$$Dice\ Coefficient(\bar{Y}, Y) = \frac{2 \times \left|\bar{Y} \cap Y\right|}{\left|\bar{Y}\right| + |Y|} = \frac{2 \times \bar{Y} \cdot Y}{\bar{Y}^2 + Y^2} \tag{5}$$

where $\left|\bar{Y} \cap Y\right|$ represents the overlap area contrasted pixel by pixel between ground true and predict result, which is usually approximated as the result of pixel-by-pixel accumulation after point multiplication between the prediction result and ground true. $Diceloss = 1 - Dice\ Coefficient$ is proposed to solve the problem that the proportion of foreground in the image is too small.

Cross entropy is another common loss function, which is defined as:

$$Cross\ Entropy = -\sum_{c=1}^{C} \bar{Y} log(Y) \tag{6}$$

where C is number of classes. When C is equal to 2, this loss is the binary cross entropy. Cross entropy loss can be used in most semantic segmentation scenarios, but it has a significant disadvantage: When it is only used to segment the foreground and background, if the number of foreground $\bar{Y} = 0$ pixels is much smaller than the number of background $\bar{Y} = 1$ pixels, and the $\bar{Y} = 0$ component in the loss function will dominate, causing the model to be heavily biased to the background, resulting in poor results.

In order to solve the problem of serious imbalance between the positive and negative samples in the field of target detection, Focal loss was first introduced in RetinaNet [6]. Considering that the class imbalance in the samples will adversely affect the final training loss, it can be corrected by setting a coefficient(α) in the loss function that is inversely proportional to the target existence probability. Focal loss is defined as follows:

$$Focal\ loss(p) = -\alpha \times (1-p)^{\gamma} \times \log(p) \tag{7}$$

where p ∈ [0,1] is the model's predicted probability for the positive class (label $\bar{Y} = 1$). Focal loss is a modification based on the cross entropy loss function. RetinaNet has proved that the coefficient α added by focal loss can automatically adjust the contribution of positive and negative samples to loss. When one sample class is easy to distinguish, its contribution to the overall loss is relatively small. On the contrary, when one sample class is hard to distinguish, its contribution to the overall loss is relatively large. Such a setting will induce the model to focus on those difficult samples, which can effectively improve the overall target segmentation accuracy.

In ANU-Net, we refer to advantages of the above three loss functions and re-integrate them to get a novel hybrid loss function, which is defined as follows:

$$\mathrm{Loss} = \sum_{i=1}^{4} \left(1 - \left[\frac{\alpha \times \overline{Y} \times \log Y_i}{|Y_i - 0.5|^{\gamma}} + \frac{2 \times Y_i \times \overline{Y} + s}{Y_i^2 + \overline{Y}^2 + s}\right]\right) \tag{8}$$

where $\overline{Y}$ is the real result and Y_i is the segmentation output from node X_{0_i}, | | is the absolute value function, s is the smooth in soft dice coefficient part. $\frac{\alpha \times \overline{Y} \times \log Y_i}{|Y_i - 0.5|^{\gamma}}$ combines with focal loss and binary cross entropy loss, where balance factor α is added to solve the imbalance problem between positive and negative samples. In order to solve this problem, different weights are given to different kind of samples in the loss function. In medical image processing tasks, for example, due to the specificity of its data, there are more positive samples (healthy), so the weight of the positive sample loss is reduced (α decrease when $\overline{Y}$ is equal to 1), and the weight of the negative sample loss is correspondingly increased (α increase when $\overline{Y}$ is equal to 0).

In clinical diagnosis, there are a large number of simple samples (the patient's symptoms are clear), and the physician spend lots of time and energy on these simple samples. Therefore, in order to effectively improve the segmentation efficiency of the proposed model and make the model pay more attention to the difficult samples, we use hyper parameter γ. It can be seen from the Eq. (8) that when the sample predicted value Y_i is close to 0 (or 1), the sample is more likely to be judged as negative (or positive). This type of sample is a simple sample. At this time $\frac{1}{|Y_i - 0.5|^{\gamma}}$ will be so small that the loss of the simple samples will decrease significantly. Similarly, when the sample prediction value Y_i is close to 0.5, it is more difficult for the sample to judge a certain category, and this type of sample is a difficult sample. At this time, $\frac{1}{|Y_i - 0.5|^{\gamma}}$ will be so large that the loss of difficult samples will increase significantly, so the model will pay more attention to the optimization of difficult samples. In [6], the author has tried to adjust the hyper parameter γ. After the experimental search, it is concluded that the model gets the best detection performance when the recommended value is two.

As a consequence, our proposed novel hybrid loss function takes full advantages of all three loss functions: provides smooth gradient and solves the imbalance between positive and negative samples.

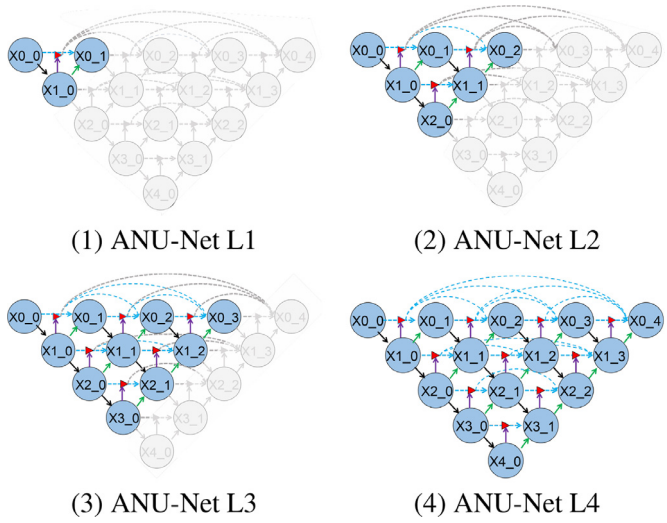

Fig. 7. Trained ANU-Net with deep supervision makes segmentation results available at different level. Region in gray means these nodes and attention gates are removed during predicting.

3.5. Model pruning

In Fig. 7, we use ANU-Net L1,L2,L3,L4 to denote network pruned at four different depth, where region in gray means these blocks and attention gates are removed during predicting.

It is easy to find that during the testing phase, removing these gray parts has no influence on the prior output because the input image will only propagate forward. However, during the training phase, because there is both forward and back propagation, the gray part will help other parts to update the weight when the error is back propagated. In short, removing does not affect the remaining structure when testing, removing has an impact on the remaining part when training. So we can conclude that the decoder path at depth d is independent of the deeper decoder path when the trained network predicts the test data.

Under the deeply supervised design structure (in Section 3.4), the output of each pruned ANU-Net (L1,L2,L3,L4) is the segmentation mask with the same-scale. So if the output of one pruned ANU-Net is satisfactory, we can decisively remove those irrelevant decoder paths and use smaller trained ANU-Net to segment when testing. We use pruned ANU-Net LN to represent taking final result from X_{0_N}. Selecting the extent of model pruning is weighed by evaluating the performance and inference time of the four subnetworks during validation. For example, if final result comes from X_{0_1}, ANU-Net L1 is a maximally pruned architecture and get qualified segmentation performance during validation. Similarly, there is no pruning in ANU-Net L4 when final result comes from X_{0_4}.

As a consequence, the introduction of deep supervision enables model pruning during the predicting time, leading to a significant speedup and a significant decrease in network parameters with only modest drop in performance.

4. Experiments and results analysis

4.1. Dataset setup and preprocessing

As shown in Table 1, four medical image segmentation datasets from famous challenges are used in this work, covering organs from common medical imaging modalities including CT and MRI.

LiTS (Liver Tumor Segmentation Challenge) dataset has 130 training CT scans and 70 testing CT scans. The data and annotations are provided by various clinical sites around the world. This dataset is supported by International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI, 2017).

Table 1
Summary of medical image segmentation datasets used in the experiments.

Target	Input Size	Modality	Source
Liver	256x256	CT	MICCAI 2017 LiTS
Spleen	256x256	MRI	ISBI 2019 CHAOS
Kidney	256x256	MRI	ISBI 2019 CHAOS
Liver	256x256	MRI	ISBI 2019 CHAOS

CHAOS [51] (Combined Healthy Abdominal Organ Segmentation) dataset consists of abdominal CT and MRI images. Only MRI images are used in this work, including 120 DICOM data sets from two different MRI sequences. This dataset is supported by The IEEE International Symposium on Biomedical Imaging (ISBI, 2019).

We split the all annotated cases at 5:1 ratio into training dataset and test dataset respectively. And then, we truncate the Hounsfield units(HU) value range of all images in the database to [−200,200] to remove irrelevant useless details.

In LiTS, the ground truth segmentation provides three different labels: liver, tumor(lesion) and background. For image preprocessing, we only consider liver as positive class and others as negative class. In CHAOS, there are four labels in the ground truth representing four abdomen organs, including liver, right kidney, left kidney and spleen. And then, we merge the left and right kidneys into kidney labels, obtained three abdominal organ labels, and trained them separately.

4.2. Evaluation metrics

We use dice similarity coefficient, intersection over union (IoU), precision and recall as performance indicators to evaluate the performance of medical image segmentation. These four indicators can be formulated as follows:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{9}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{10}$$

$$Precesion = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

where TP, FP, FN denote True Positive, False Positive, False Negative respectively. The larger the values of these four indicators, the larger the overlapping area between the segmentation result and the ground true, the higher the similarity, and the greater the accuracy of the segmentation.

4.3. Segmentation results

The proposed ANU-Net is benchmarked against the other five popular model (U-Net, R2U-Net, UNet++, Attention U-Net and Attention R2U-Net) on medical image segmentation task. Tables 2–5 summarize the segmentation performance of the above models on four medical image segmentation tasks.

In the liver CT image segmentation task shown in Table 2, the performance of ANU-Net is the best. Compared with Attention U-Net, ANU-Net achieves the IoU ratio increased 7.99%, the Dice coefficient increased by 3.7%, and the precision increased 5%, recall rate increased 4% over U-Net. The comparison between the segmentation results and the manually annotated result (ground true) is shown in Fig. 8.

Table 2
Comparison of liver segmentation performance of the models in CT images of LiTS test dataset.

Method	IoU(%)	Dice (%)	Precision	Recall
		(higher is better)		
U-Net [10]	89.49	94.45	0.9324	0.9570
R2U-Net [41]	90.69	95.11	0.9380	0.9648
UNet++ [47]	94.46	97.15	0.9816	0.9617
Attention UNet [46]	93.39	96.58	0.9679	0.9637
Attention R2U-Net	92.38	96.03	0.9712	0.9498
ANU-Net	**97.48**	**98.15**	**0.9815**	**0.9931**

Table 3
Comparison of spleen segmentation performance of the models in MRI images of CHAOS test dataset.

Method	IoU(%)	Dice (%)	Precision	Recall
		(higher is better)		
U-Net [10]	76.18	86.48	0.8234	0.9106
R2U-Net [41]	81.50	89.77	0.9360	0.8624
UNet++ [47]	81.05	89.53	0.8637	0.9293
Attention UNet [46]	84.13	91.38	0.9154	0.9122
Attention R2U-Net	81.62	89.79	0.9088	0.8873
ANU-Net	**89.23**	**94.31**	**0.9519**	**0.9344**

Table 4
Comparison of kidney segmentation performance of the models in MRI images of CHAOS test dataset.

Method	IoU(%)	Dice (%)	Precision	Recall
		(higher is better)		
U-Net [10]	84.46	91.58	0.8920	0.9408
R2U-Net [41]	85.54	92.21	0.9192	0.9250
UNet++ [47]	86.58	92.81	0.9087	0.9482
Attention UNet [46]	85.77	92.34	0.9097	0.9376
Attention R2U-Net	87.34	93.24	0.9141	0.9515
ANU-Net	**90.10**	**94.79**	**0.9400**	**0.9560**

Table 5
Comparison of liver segmentation performance of the models in MRI images of CHAOS test dataset.

Method	IoU(%)	Dice (%)	Precision	Recall
		(higher is better)		
U-Net [10]	75.37	85.96	0.8731	0.8465
R2U-Net [41]	77.80	87.50	0.9211	0.8339
UNet++ [47]	84.23	91.39	0.9306	0.8979
Attention UNet [46]	76.00	86.37	0.9111	0.8209
Attention R2U-Net	82.44	90.35	**0.9474**	0.8640
ANU-Net	**87.89**	**93.55**	0.9423	**0.9288**

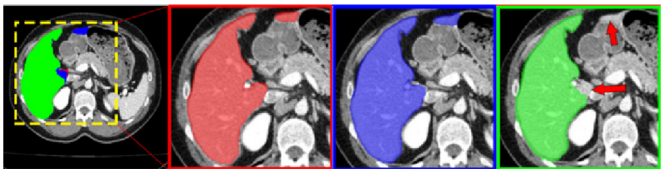

Fig. 8. Comparison between liver segmentation results and manually annotated result. The ground true is shown in a red region, the outputs of ANU-Net and U-Net are displayed in a blue region and a green region. U-Net's prediction missed area is highlighted with red arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

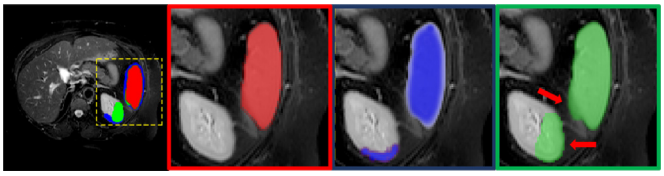

Fig. 9. Comparison between spleen segmentation results and manually annotated result. The ground true is shown in a red region, the outputs of the ANU-Net and Attention UNet are displayed in a blue region and a green region. Attention UNet's prediction missed area is highlighted with red arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

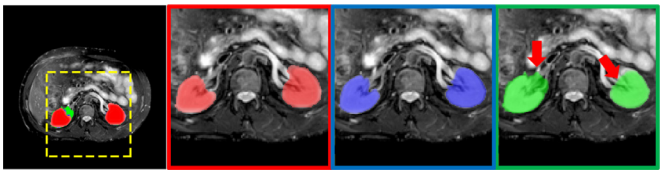

Fig. 10. Comparison between kidney segmentation results and manually annotated result. The ground true is shown in a red region, the outputs of ANU-Net and UNet++ are displayed in a blue region and a green region. UNet++'s prediction missed area is highlighted with red arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

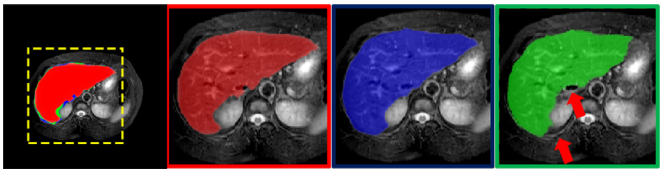

Fig. 11. Comparison between liver segmentation results and manually annotated result. The ground true is shown in a red region, the outputs of ANU-Net and Attention R2U-Net are displayed in a blue region and a green region. Attention R2U-Net's prediction missed area is highlighted with red arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In the spleen MRI image segmentation task shown in Table 3, the performance of ANU-Net is the best. Compared with Attention UNet, ANU-Net achieves the IoU ratio increased 6.06%, the Dice coefficient increased by 3.21%, and the precision increased 3.99%, recall rate increased 2.43% over Attention UNet. The comparison between the segmentation results and the manually annotated result (ground true) is shown in Fig. 9.

In the kidney MRI image segmentation task shown in Table 4, the performance of ANU-Net is the best. Compared with UNet++, ANU-Net achieves the IoU ratio increased 4.06%, the Dice coefficient increased by 2.13%, and the precision increased 3.44%, recall rate increased 0.82% over UNet++. The comparison between the segmentation results and the manually annotated result (ground true) is shown in Fig. 10.

In the liver MRI image segmentation task shown in Table 5, the performance of ANU-Net is the best. Compared with Attention R2U-Net, ANU-Net achieves the IoU ratio increased 6.61%, the Dice coefficient increased by 3.54%, and the precision decreased only 0.53%, recall rate increased 7.50% over Attention R2U-Net. The comparison between the segmentation results and the manually annotated result (ground true) is shown in Fig. 11.

It can be concluded that in the four medical image segmentation tasks, the proposed Attention Nested U-Net(ANU-Net) outperforms other methods. As seen, due to the use of the nested convolution structure, UNet++ always performs better than U-Net on this four indicators, and our improved nested structure in ANU-Net has achieved better results than UNet++ in the X medical image segmentation task. These comparisons strongly prove that the nested U-Net architecture can essentially improve the

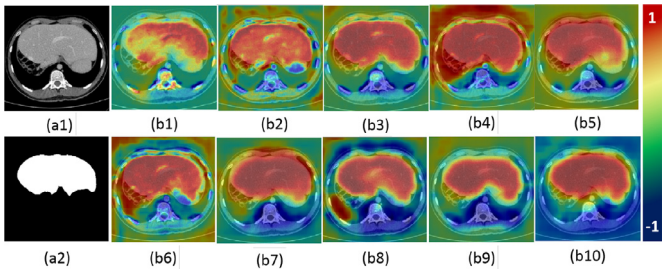

Fig. 12. (a1) and (a2) are the CT images and ground true in LiTS. (b1) to (b10) are the attention coefficients in the first layer of attention gates in different training periods. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

segmentation performance, which meets the expectations for the design in Section 3.1.

4.4. Attention learning results

As shown in the Tables 2–5, Attention U-Net consistently outperforms U-Net under four indicators due to the introduction of attention mechanism, and our proposed Attention Gate has the best performance in all attention-aware methods. These comparisons prove that Attention Gate can improve the segmentation performance, which meets the expectations for the design in Section 3.2.

Fig. 12 is a group of graphs in the liver CT image segmentation task, showing the attention learning results brought by Attention Gates. The following conclusions can be drawn:

- The attention coefficient α obtained is visualised during different trainingperiods.
- Attention Gates initially have a uniform distribution at all locations. This is gradually updated and localised towards the targeted organ boundaries.
- Attention coefficient gradually changes. When training, weights in the target organ (red region) where related to the segmentation task is enhancing.
- Weights in the confusing tissue (blue region) where unrelated to the segmentation task is suppressing.

As a consequence, it is clearing that introducing attention mechanism can enhance the learning of target regions and suppress irrelevant regions. Finally, Attention Gates increase the effectiveness of dense skip connections. Besides, there is no need to trim the ROI and locate the target object in our model. As a consequence, ANU-Net is one-step method because there is no need for extra localization step, according to Section 1.

4.5. Model pruning results

As we can see in Table 6: ANU-Net L1, L2, L3, L4 have huge differences in network parameters, where L1 has only 0.1M parameter amount, but L4 has 8.9M.

Numerically, if the result of ANU-Net L1 is satisfactory, the parameters that the network can save after pruning reaches 98.8%. Meanwhile, ANU-Net L1 achieves on average 17.128% reduction in prediction time while decreasing IoU by 13.354% and decreasing Dice by 27.18%. Similarly, ANU-Net L3 achieves on average 7.734% reduction in prediction time and 75.512% reduction in parameters while decreasing IoU by only 0.615% and decreasing Dice by only 2.558%.

However, according to the results of the liver segmentation experiment on the LiTS dataset, the segmentation performance of ANU-Net L1 is not so ideal because it is too shallow. Besides, the

Table 6
Liver CT image segmentation performance of pruned models.

Pruned ANU-Nets	Predict Time(s)* *(lower is better)*	Parameters Amounts(M)	Dice (%) *(higher is better)*	IoU (%)
L1	5.3795	0.1038	71.3670	83.2553
L2	5.7283	0.5134	90.2052	94.8468
L3	5.9893	2.1918	95.4969	95.4960
L4	6.4914	8.9506	98.0042	96.0871

* Predict time is the time that every model takes to diagnose 2K CT images in LiTS.

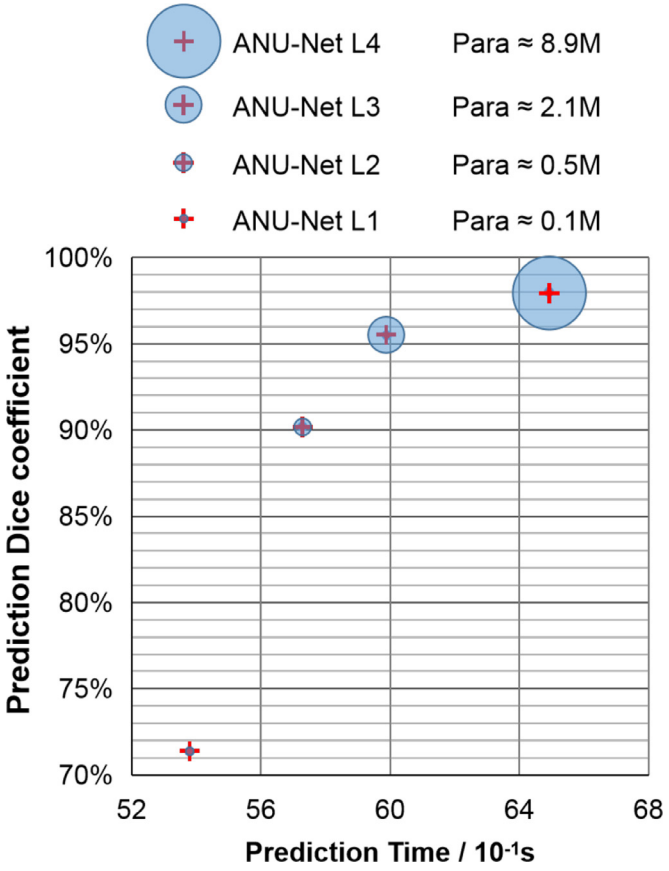

Fig. 13. Parameters amount, inference time, and Dice coefficient of pruned ANU-Net under different extent of pruning. The area of the blue circle represents the parameter amount of the network.

results also show that the segmentation performance of ANU-Net L2 and L3 are very close to L4. That is to say, for liver CT image segmentation task, there need to use ANU-Net L4 (9M) during the testing phase. A 0.5M or 2.1M network is sufficient, so that we can save at least 75.512% of the parameter amount (if ANU-Net L3 used). Fig. 13 shows the results of Table 6. This figure can more intuitively confirm the benefits of model pruning in ANU-Net from the other side.

So we conclude that model pruning can significantly reduce the model parameters and prediction time, but segmentation performance decline at same time. Therefore, we need to make a reasonable judgment based on the actual condition before pruning. In addition, since most deep CNN segmentation models take a long time to calculate and require a large amount of memory, it is more

practical to apply the pruned ANU-Net to computer-aided diagnosis in small computers, especially in mobile devices.

5. Conclusions

In this paper, an attention-based nested segmentation network (ANU-Net) is proposed and applied to medical image segmentation, covering various organs from common medical imageing modalities. Our experiments on the LiTS and CHAOS dataset demonstrated that the competitive performance of proposed ANU-Net in medical image segmentation. The improvement is attributed to the combination of dense skip connection and attention mechanism. Dense skip connections designed in ANU-Net between nested convolution blocks obtains full resolution feature maps at different semantic levels from blocks. In order to make full use of these full resolution feature information, we redesign a hybrid loss function combining with soft dice coefficient loss, focal loss and pixel-wise binary cross entropy loss. Experiments also proved that ANU-Net has the ability to increase the weight of the target region while inhibiting the background region that is unrelated to the segmentation task. Besides, due to the introduction of deep supervision, the prediction speed of the pruned ANU-Net is accelerated at the cost of modest performance degradation.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Chen Li: Conceptualization, Methodology, Software, Writing - original draft. **Yusong Tan:** Supervision. **Wei Chen:** Project administration, Writing - review & editing. **Xin Luo:** Resources, Visualization. **Yulin He:** Data curation, Formal analysis. **Yuanming Gao:** Investigation. **Fei Li:** Funding acquisition.

Acknowledgments

This work is supported by National Key Research and Development Program of China (No. 2018YFB0204301), General Program of National Natural Science Foundation of China (81973244) and Science and Technology Program Projects of Shenzhen (JCYJ20170818110101726).

References

[1] Simonyan K., Zisserman A.. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556 2014;.
[2] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: European conference on computer vision. Springer; 2016. p. 630–45.
[3] Chen L-C, Yang Y, Wang J, Xu W, Yuille AL. Attention to scale: scale-aware semantic image segmentation. In: The IEEE conference on computer vision and pattern recognition (CVPR); 2016.
[4] Diakogiannis FI, Waldner F, Caccetta P, Wu C. Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens 2020;162:94–114.
[5] Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems; 2015. p. 91–9.
[6] Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 2980–8.
[7] Eilertsen G, Kronander J, Denes G, Mantiuk RK, Unger J. Hdr image reconstruction from a single exposure using deep CNNs. ACM Trans Graph 2017;36(6):1–15.
[8] Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 9446–54.
[9] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal MachIntelli 2014;39(4):640–51.
[10] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 234–41.
[11] Li X, Chen H, Qi X, Dou Q, Fu C-W, Heng P-A. H-Denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. IEEE Trans Med Imaging 2018;37(12):2663–74.
[12] Payer C, Štern D, Bischof H, Urschler M. Multi-label whole heart segmentation using CNNs and anatomical label configurations. In: International workshop on statistical atlases and computational models of the heart. Springer; 2017. p. 190–8.
[13] Liao F, Liang M, Li Z, Hu X, Song S. Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network. IEEE Trans Neural Netw Learn Syst 2019.
[14] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.
[15] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097–105.
[16] Dai J, Li Y, He K, Sun J. R-FCN: object detection via region-based fully convolutional networks. In: Advances in neural information processing systems; 2016. p. 379–87.
[17] Tang P, Wang X, Huang Z, Bai X, Liu W. Deep patch learning for weakly supervised object classification and discovery. Pattern Recognit 2017;71:446–59.
[18] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. SSD: Single shot multibox detector. In: European conference on computer vision. Springer; 2016. p. 21–37.
[19] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2014. p. 580–7.
[20] Tang P, Wang C, Wang X, Liu W, Zeng W, Wang J. Object detection in videos by high quality object linking. IEEE Trans Pattern Anal Mach Intell 2019.
[21] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 2881–90.
[22] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell 2017;40(4):834–48.
[23] Brostow GJ, Fauqueur J, Cipolla R. Semantic object classes in video: a high-definition ground truth database. Pattern Recognit Lett 2009;30(2):88–97.
[24] Song S, Lichtenberg SP, Xiao J. Sun RGB-D: a RGB-D scene understanding benchmark suite. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 567–76.
[25] Kistler M, Bonaretti S, Pfahrer M, Niklaus R, Büchler P. The virtual skeleton database: an open access repository for biomedical research and collaboration. J Med Internet Res 2013;15(11):e245.
[26] Pinheiro PO, Collobert R, Dollár P. Learning to segment object candidates. In: Advances in neural information processing systems; 2015. p. 1990–8.
[27] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 2961–9.
[28] Chen L-C, Hermans A, Papandreou G, Schroff F, Wang P, Adam H. Masklab: instance segmentation by refining object detection with semantic and direction features. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 4013–22.
[29] Huang Z, Huang L, Gong Y, Huang C, Wang X. Mask scoring R-CNN. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2019. p. 6409–18.
[30] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2016. p. 424–32.
[31] Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: International conference on medical image computing and computer-assisted intervention. Springer; 2013. p. 246–53.
[32] Roth HR, Lu L, Farag A, Shin H-C, Liu J, Turkbey EB, et al. Deeporgan: multi--level deep convolutional networks for automated pancreas segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 556–64.
[33] Wang J, MacKenzie JD, Ramachandran R, Chen DZ. Detection of glands and villi by collaboration of domain knowledge and deep learning. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 20–7.
[34] Sun C, Guo S, Zhang H, Li J, Chen M, Ma S, et al. Automatic segmentation of liver tumors from multiphase contrast-enhanced ct images based on FCNs. Artif Intell Med 2017;83:58–66.
[35] Ben-Cohen A, Diamant I, Klang E, Amitai M, Greenspan H. Fully convolutional network for liver segmentation and lesions detection. In: Deep learning and data labeling for medical applications. Springer; 2016. p. 77–85.
[36] Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng P-A. 3d deeply supervised network for automatic liver segmentation from CT volumes. In: International conference on medical image computing and computer-assisted intervention. Springer; 2016. p. 149–57.

[37] Lu F, Wu F, Hu P, Peng Z, Kong D. Automatic 3d liver location and segmentation via convolutional neural network and graph cut. Int J Comput Assist Radiol Surg 2017;12(2):171–82.
[38] Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). IEEE; 2016. p. 565–71.
[39] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net. IEEE Geosci Remote Sens Lett 2018;15(5):749–53.
[40] Li R, Liu W, Yang L, Sun S, Hu W, Zhang F, et al. Deepunet: a deep fully convolutional network for pixel-level sea-land segmentation. IEEE J Sel Top Appl Earth Obs Remote Sens 2018;11(11):3954–62.
[41] Alom M.Z., Hasan M., Yakopcic C., Taha T.M., Asari V.K.. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:180206955 2018;.
[42] Wang X, Girshick R, Gupta A, He K. Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 7794–803.
[43] Jin Q., Meng Z., Sun C., Wei L., Su R.. RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in ct scans. arXiv preprint arXiv:181101328 2018;.
[44] Bahdanau D., Cho K., Bengio Y.. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:14090473 2014;.
[45] Bahdanau D, Chorowski J, Serdyuk D, Brakel P, Bengio Y. End-to-end attention-based large vocabulary speech recognition. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2016. p. 4945–9.
[46] Oktay O., Schlemper J., Le Folgoc L., Lee M., Heinrich M., Misawa K., et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999 2018;.
[47] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: a nested U-Net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2018. p. 3–11.
[48] Iandola F., Moskewicz M., Karayev S., Girshick R., Darrell T., Keutzer K.. Densenet: implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:14041869 2014;.
[49] Wang L, Wang Y, Liang Z, Lin Z, Yang J, An W, et al. Learning parallax attention for stereo image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2019. p. 12250–9.
[50] Lee C-Y, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply-supervised nets. In: Artificial intelligence and statistics; 2015. p. 562–70.
[51] Kavur A.E., Selver M.A., Dicle O., Bar M., Gezer N.S.. CHAOS – Combined (CT-MR) healthy abdominal organ segmentation challenge data2019; 10.5281/zenodo.3362844