EA-UNet: Shape-aware UNet for Biomedical Image Segmentation with Spatio-channel Mechanism

Author First[*]
Department of Computer
Beijing University of Chemical Technology
`first@mail.buct.edu.cn`

Author Second
Department of Computer
Beijing University of Chemical Technology
`second@mail.buct.edu.cn`

Abstract

In recently years, deep learning-based models brought about performance breakthrough in medical image community. For image segmentation, registration and reconstruction tasks, deep models consistently outperforms traditional methods. Specifically, the U-Net model proposed by Olaf Ronneberger et al.[22] and some variants are becoming the de facto standard model and widely used in various medical image tasks. Despite great success, it is still a challenge task to accurately segment and identify the location and sizes of rectal tumors. Another challenge is the amorphous or irregularly shaped boundaries of object. In this study, we propose a novel U-Net architecture to address the aforementioned issues regarding segment performance and shape identification. our model mainly includes two improvements on U-Net. Firstly, we re-design the skip pathways with spatial and channel attentions which reduce the semantic gap between the feature maps of encoder and decoder sub-networks. Secondly, a novel loss function was proposed to emphasize the segmentation accuracy on tumor boundaries and encourages shape feature preservation. We have evaluated the ASU-Net in comparison with U-Net and some variants on multiple medical image segmentation tasks: liver segmentation in abdominal CT scans, nodule segmentation in the low-dose CT scans. Based on the experimental results, it was identified that the proposed method performed a significant role in improving accuracy and shape preservation.

1 Introduction

Medical images of different modality such as tomography (CT), positron emission tomography (PET), X-ray, Ultrasonic, and singlephoton emission computed tomography (SPECT) play an increasingly important role for clinician to analyse and diagnosis of diseases. For example, Diabetic Retinopathy (DR), which is a universal retinal disease caused by elevated blood sugar, accompanied by retinal vascular swelling. Early detection and treatment are the keys to determine the degree and severity of

---

[*]corresponding author

vitreoretinal traction. Ultrasound is a valuable diagnostic tool for this disease, which can evaluate the state of the retina in vitreous hemorrhage or vitreous opacity.

In general, medical image analysis including image segmentation, object detection and tracking tasks can be categorized into two classes. (1) manual and semi-automatic segmentation, (2) automatic segmentation. The traditional manual analysis requires physicians, radiologists, or experts with advanced technical knowledge to carry out the tasks. Furthermore, the segmentation quality relies heavily on the judgement of experts. The segmentation and diagnosis process is rather time-consuming and error-prone. In the past few decades, researchers have proposed a great number of methods for automatic segmentation, detection and tracking in Medical Image Community. Early methods were based on traditional pattern recognition techniques like statistical modeling and edge detection filters. For example, Bankhead et al[21] proposed a wavelet-based method to enhance the detection of vessel foreground and background. Later, machine learning approaches using hand-crafted features based on the modality and type of segmentation task were developed. For instance, Lupascu et al. [17] designed 41-D feature vectors for every pixel to train the AdaBoost classifier to classify each pixel in the retinal images.

With the advent of deep learning, convolutional neural networks (CNNs) have been showing promising results over the conventional methods in medical image tasks like image segmentation, registration and tracking and so on. Medical image segmentation is an important task of artificial intelligence assisted medical technology. The automatic segmentation of medical image not only help clinician identify the candidate regions, but also give helpful diagnosis suggestions. During the past decade, many effective models have been developed by researchers. The classic model is fully convolutional network (FCN) proposed by Jonathan Long et al[16]. Fully convolutional networks (FCN) take advantage of fully convolutional layers to train a segmentation model in an end-to-end way by replacing fully connected layers in CNN with convolutional layers. Despite outstanding performances, FCN suffers from conceptual limitations in complex segmentation tasks, e.g. when dealing with local visual ambiguities and low contrast between organs. Since then, several variants have been developed for medical image segmentation tasks like SegNet, U-Net and so on. What is worth noting is that most architectures developed for semantic segmentation in both computer vision and medical medical images are encoder-decoder type convolutional networks. SegNet was the first encoder-decoder type convolution networks. A major breakthrough in medical image segmentation was U-Net architecture proposed by Olaf Ronneberger[22] which has recently become backbone of almost all the leading models. Subsequently, a variety of network architectures based on U-Net are proposed to enhance the segmentation performance. Several main extensions includes UNet++, Attention Gate UNet, UNet-Transformer, Double U-Net, TVUNet and so on. In addition, extensions like 3D U-Net and V-Net are also proposed for volumetric segmentation in 3D medical scans. It is worth noting that all the above extensions of U-Net used the same encoder-decoder architecture and their contributions were either in encoder and decoder, using different convolutional layer or in skip connections, using different concatenation to combine the features in appropriate level.

Despite the success of U-Net and its variants, there are some disadvantages of above networks. In general, U-Net and its variants are good at segmenting large structure, they fail when the segment objects are small or have noisy boundaries which can be seen in Fig 1. From the figure above, the U-Net architecture lacks focus in extracting features for segmentation of small structures.
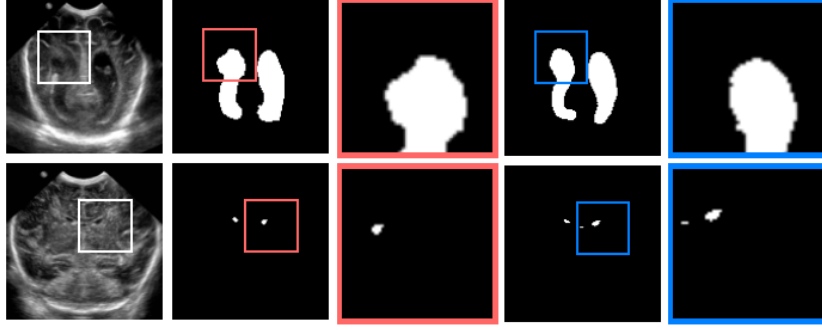
图 1: small structure segmentation

In addition, UNet network also faces the challenge to segment objects with enough good accuracy for retina segmentation due to shape variability and unclear boundaries of segmenting objects. In encoder part of network, feature extraction process is often performed by CNN networks which are are highly influenced by dense pixel values which are not robust features compared to shapes of objects. From the perspective of evaluation, most segmentation evaluation metrics to quantify segmentation performance, such as accuracy, spedificity, and jaccard similarity, do not directly reflect how well a object boundary is delineated which hinders to segment object with clear boundaries. Last but not least, UNet architecture family also shared the disadvantage by deep neural network is a lack of interpretability.

To address the issues illustrated above in medical image segmentation, in this paper, we proposed a new UNet variant named SA-UNet. The proposed model is inspired by growing interest on integrating attention mechanisms and shape-aware feature preservation in image segmentation networks for natural scenes. Attention mechanism has been shown to be a powerful technique in deep learning models and has been widely used in various tasks in natural language processing to capture keywords for context. Unlike the previous approaches, SA-UNet re-design the skip-connection to fuse several different attention mechanisms including spatial attention, channel-wise attention and self-attention. Spatial attention focus on some special region which contains the objects to be segmented; channel-wise attention vector are employed in the lower resolution to extract feature in some channel. self-attention effectively captures the long-range dependencies between visual features. Despite the growing interest on integrating attention mechanisms in medical image segmentation scenes, the adoption of several attention mechanisms in medical images remains scarce, being limited to simple attention models. Furthermore, we also consider shape-feature perservation in feature extraction procedure in imporved implemenation in the loss function. The shapes of objects should be learned to allow generalizability and robustness of the model which goes hand-in-hand with transparency which also increases the interpretability in a clinical setting. There are several advantages of our proposed architectures for segmentation task. Fristly, The new model augmented with multi-attention improves the performance in a variety of segmentation tasks. Secondly new loss function contributes to the shape preservation of objects to be segmented. The experimental results show superior performance on segmentation tasks compared to equivalent models including U-Net and serveral variants.

The contributions of this work can be summarized as follows:

1. We propose a novel multi-attention module that can be utilised in CNN based standard image analysis models for classification, segmentation and tracking application and so on.

2. We redesign the skip-connection to fuse different attention mechanisms and proposed the SAU-Net architecture for medical image segmentation task.

3. The architecture use new loss function for shape preservation for medical image analysis, in particular, in the context of image classification and segmentation.

The remainder of this paper is organized as follows.. In Section 2, we introduce background knowledge, including U-Net architecture family, existing attention mechanism and shape-aware related work in machine learning fields. Section 3 introduce the proposed method, including network architecture and attention modules and new loss function to preserve the shape feature. in Section **??**, we give some experiment and present the results in detail and section 6 conclude the paper.

2 Related work

In this section, we review several lines of research related to our work in the following felds.

U-Net and variants   In medical image segmentation, Fully Convolution Networks (FCN) have been widely used for semantic segmentation and contextual aggregation. In 2015, Olaf Ronneberger et al. proposed a U-shape architecture that employs contracting and expanding paths together with skip connections, which combines both low-and high-level features for image segmentation[22]. With UNet as a backbone, researchers proproosed improved variants for volumetric segmentation in 3D medical scans, such as 3DUNet[4] and V-Net[18]. Zongwei Zhou et al. proposed UNet++, a new segmentation architecture based on nested and dense skip connections [28]. The architecture can more effectively capture fine-grained details of the foreground objects when high-resolution feature maps are gradually fusioned from encoder and decoder network. Wen Liu et al. presented an enhanced nested U-Net architecture named MDAN-UNet for segmentation of OCT images[14]. The approach designed a UNet++ alike network and redesigns the skip pathways, multi-scale input, multi-scale side output and accurately segment pixels of OCT images. Nabil Ibtehaz et al. propose some modifications to improve upon the already state-of-the-art U-Net model[9]. They replaces the convolutional layers with Inception-like blocks should facilitate the U-Net architecture to reconcile the features learnt from the image at different scales. They also introduced residual connections to them as they make the learning easier. Kevin Trebing et al propose SmaAt-UNet, an efficient convolutional neural networks based on the well known UNet architecture equipped with attention modules and depthwise-separable convolutions[25].

Attention   In recent years, attention has been a fairly popular concept and useful mechanism in deep learning community. Attention mechanism has been shown to be a powerful technique in deep learning models and has been widely used in various tasks in natural language processing to capture keywords for context. Some attention mechanisms and models, such as transformer and SNAIL are nearly becoming the de facto standard components of deep networks. In fact, the idea of attention is intuitive and simple, for an image, our visual system always focuses on some object of interest in the image, rather than focusing on the whole image. Inspired by the successful application of attention mechanism, Sahar Yousefi et al proposed Attention U-Net with Dilated Dense Attention for automatically segmenting the esophageal tumor[27]. The model introduced dilated dense spatial attention and channel attention blocks which are composed of dilated dense block (DDB) and a spatial attention gate and a channel attention gate. Jo Schlempera et al. introduce attention mechanism named Attention Gate (AG) which can be flexibly incorporated in existing CNN segmentation and

classification architectures[20]. Attention Gate mechanism also provides fine-scale attention maps that can be visualized, with minimal computational overhead. The experiments on 3D-CT abdominal image and 2D fetal ultrasound show increased sensitivity while not sacrificing specificity. Yancheng Lan et al. designed an novel encoder-decoder network using mixed-attention mechanism (MARU) to reconstruct the ultrasound image[12]. During the encoding phase, a lightweight mixed-attention block is designed to effectively enhance the image features and suppress some speckle noise. Haijun Gao et al proposed DPU-Net model further modified the U-Net structure which has only one path to two contracting paths[6]. DPU-Net has the function of covariance self-attention which is embedded into the up-sampling and bottom block sub-modules. The main advantage of DPU-Net not only integrates multi-scale semantic information, but the decoder can obtain sufficient fine-grained and coarse-grained semantic information at multiple scales. Mehrdad Noori et al. proposed an attention-guided extension of UNet network for Brain Tumor Segmentation in which two techniques are employed[19]. Firstly, an attention mechanism is adopted after concatenation of low-level and high-level features. Secondly, multi-views are fused from 3D contextual information of input images to enrich feature maps despite using a 2D model. Chen Li et al. incorporated the Attention Gate mechanism into nested UNet architecture and proposed the attention UNET++. The model redesigned dense skip connection and increased the weight of the target region while inhibiting the background region[13]. Changlu Guo et al. presented Channel Attention Residual U-Net(CAR-UNet) model which considers the relationship between the feature channels to strengthen the discriminative capability[8].

Shape The existing deep learning models for medical image segmentation are usually based on traditional loss functions that only regard objects at pixel level. Obviously, the medical images that suffer from low-quality and low signal-to-noise ratio without any shape constraint remains problematic even for deep models. Some works has been shown that incorporation of shape prior information significantly improves the performance of the segmentation algorithms, incorporation of such prior knowledge is a tricky practical problem. In general, incorporation of shape prior information can be devided into three main categories: 1) Conditional or Markov Field models that establish connections between different pixel regions; 2) Active/Statistical Shape Models that learn a special representation for valid shapes; 3) Active Contour Models or snakes that use deformable splines for shape detection;

These models are either applied as pre-processing steps to create initial segmentations, post-processing steps to refine the neural network segmentations, or used in multi-step models consisting of various models along a specific pipeline.

3 Proposed Method

Our novel architecture is inspired from the attention UNet, which uses attention to give importance to a certain region out of the entire image.

这里描述提出的方法，先描述模型整理，然后再描述模型的部分和细节内容。有些内容用数学语言来进行阐述。

Different from the original UNet architecture we have used attention gates which take the features from the encoder and the output of the pyramid pool as input and produced out-put is further concatenated with the up-sampled output of the previous pyramid-pool layer and mapped to the next subsequent layer. This network can map local features with their global counterparts with improved accuracy and emphasize on discriminative image regions by focusing on relevant local features only.
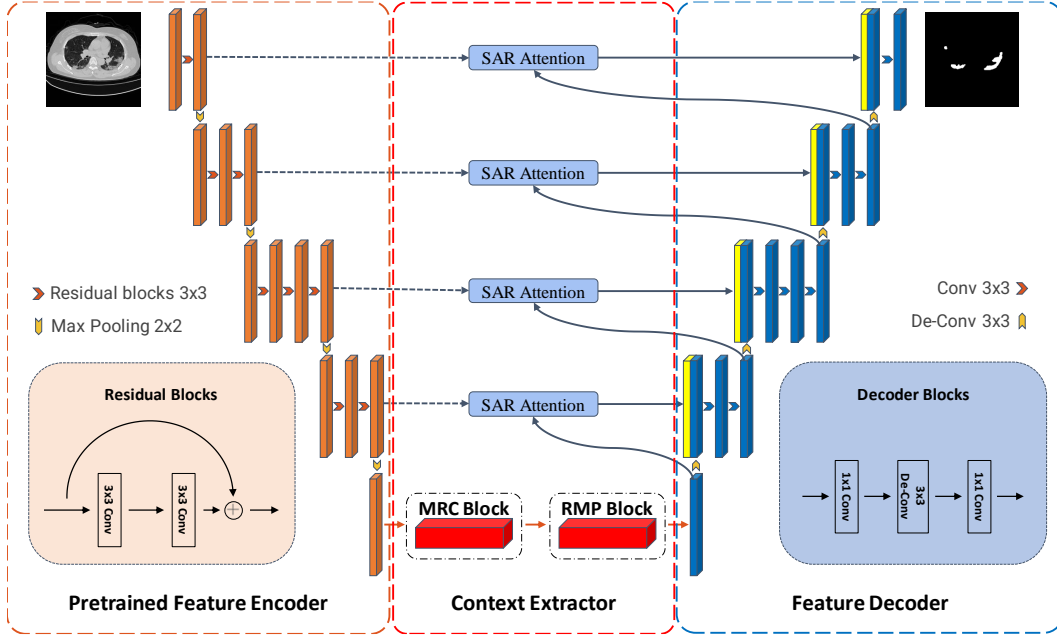
3.1 EA-UNet

图 2: EA-UNet Architecture

There are two commonly used attention types: multiplicative (Luong et al., 2015) and additive attention (Bahdanau et al., 2014). The former is faster to compute and more memory-efficient in practice since it can be implemented as a matrix multiplication. However, additive attention is experimentally shown to be performing better for large dimensional input features (Britz et al., 2017). For this reason, we use the latter to obtain the gating coefficient as can be seen in Figure 2, which is formulated as follows:

3.2 Attention Ensemble

在语义分割任务中，deep convolutional layers已经展现出其对提取图像特征的重要性。等人提出了CE-Net[7]，其在Bottleneck采用了DeepLab中的SPP-Block

3.3 Learning Shape

The active contour loss was recently introduced by Chen et al. [5] in an attempt to incorporate shape considerations in the learning process.

$$Loss_{AC} = Loss + \lambda Region \tag{1}$$

Backbone	Components	Params	Params size
CE-Net	SAR	135,571,137	517.16
ResU-Net	SAR+ASPP	92,573,377	353.14
ResU-Net	SAR+MRC+RMP	94,669,505	361.14

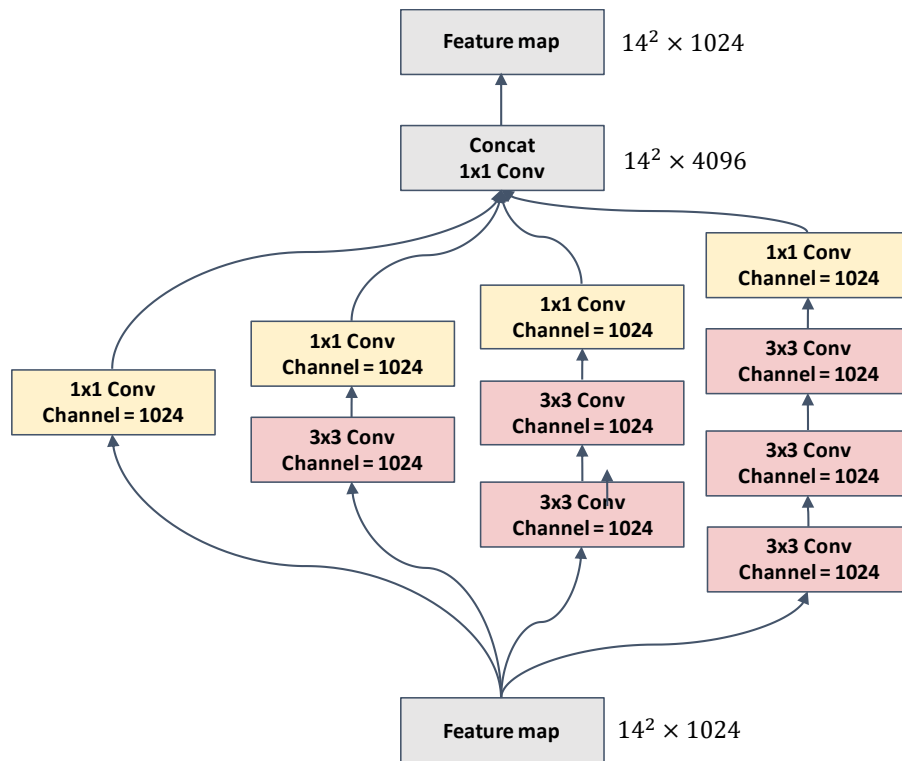表 1: Memorize Cost Comparison With Different Methods In Bottleneck
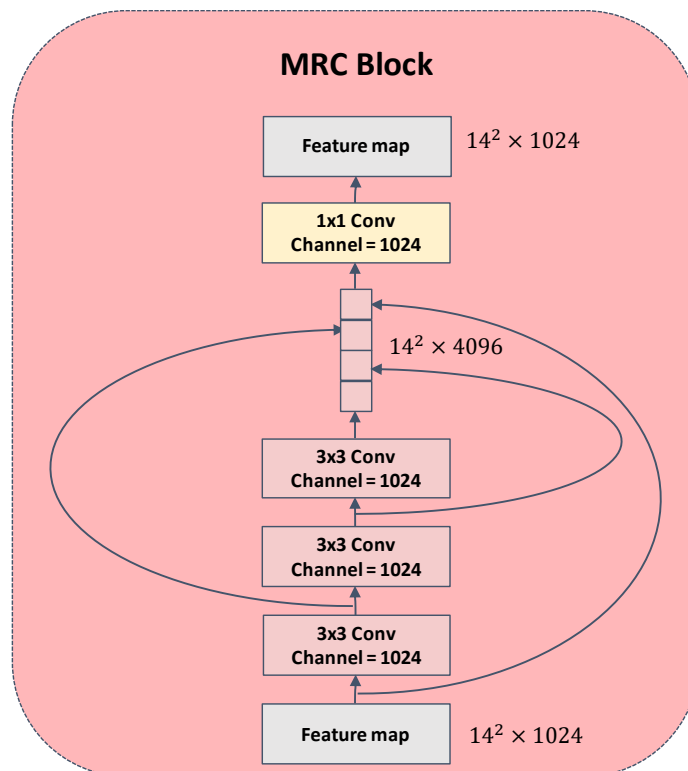
图 3: DAC BLock

图 4: MRC BLock
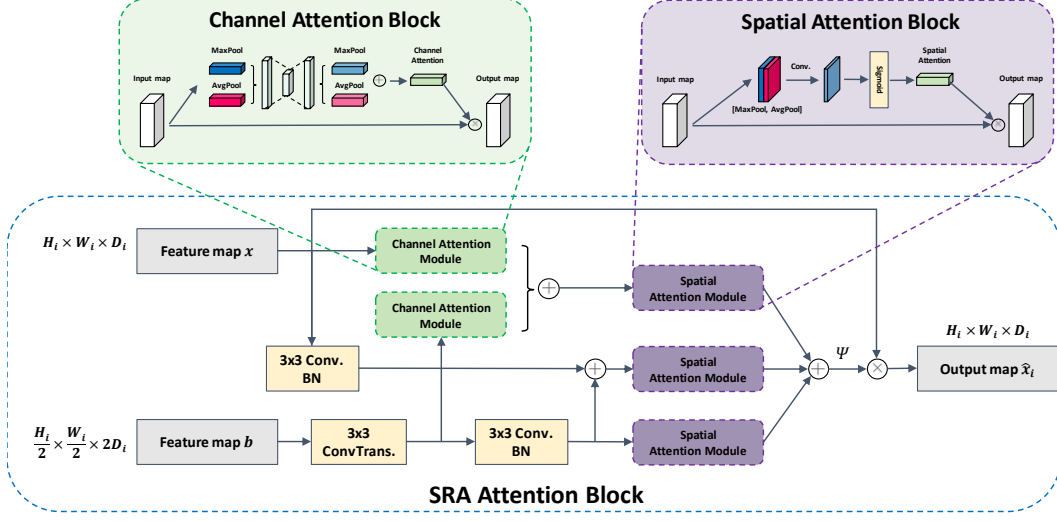
7

3.4 Attention Modules

图 5: attention

4 Experiments

4.1 Baseline and implementation

We used a server equipped with an Intel Core i9-9980XE CPU @ 3.00GHz with 64GB RAM and 12GB of RTX2080Ti GPU for our proposed networks training. The operating system of the sever is 64-bits Ubuntu 18.04. The structure of the network is implemented under the open source deep learning library Pytorch with VSCode implementation.

4.2 Dataset

For this study, we conduct our experiments on four differents segmentation tasks. Covering lesions/organs from most commonly used medical imaging modalities including microscopy, computed tomography (CT), and magnetic resonance imaging (MRI). Table 4.4.1 summarize those datasets in our study.

Dataset	Image	Input Size	Modality	Provider
Cell	30	512×512	EM	ISBI 2012[1]
Liver	4,000	512×512	CT	LiTS 2017[3]
DSB2018	670	256×256	EM	Kaggle[2]
COVID19	1,800	630×630	CT	Web[26, 10]

表 2: Summaey Of Biomedical Image Segmentation Datasets Used In Our Experiments

Cell The datset is the segmentation of neuronal structures in electron microscopic recordings. The dataset is provided by the EM segmentation challenge[1] that is started at ISBI 2012. The data is a set of 30 images (512×512 pixels) from serial section transmission electron microscopy of the Drosophila first instar larva ventral nerve cord (VNC). Each image comes with a corresponding fully annotated ground truth segmentation map for cells (white) and membranes (black).

Liver Liver tumor Segmentation Challenge (LiTS)[3] contain 131 contrast-enhanced CT images provided by hospital around the world with 512×512 resolution. The ground truth segmentation provides two different labels: liver and lesion. For our experiments, we only consider liver as positive class and others as negative class.

COVID19 Dataset[26] includes whole volumes and includes, therefore, both positive and negative slices (373 out of the total of 829 slices have been evaluated by a radiologist as positive and segmented). Dataset[10] contains 20 CT scans of patients diagnosed with COVID-19 as well as segmentations of lungs and infections made by experts. These volumes are converted and normalized in a similar way as above, meanwhile we resize the data to 512×512.

4.3 Evaluation metrics

The experiments are implemented using the Pytorch framework. We use Adam optimizer[11] as our models' optimizer with a learning rate of 0.00001, batch size of 2. All of datasets are splitted into training set, validation set and test set with the ratio of 8:1:1 using sklearn library. To numerically evaluate, we use five widely adopted metrics, $i.e.$, the Dice similarity coefficient(Dice.), F1 score., Sensitivity(Sen.), Iou. and hausdorff distance(Hd)., the expressions of them are defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{DSC}(G, S) = \frac{2|G \cap S|}{|G| + |S|} \tag{3}$$

$$\text{IOU}(G, S) = \frac{|G \cap S|}{|G| \cup |S|} \tag{4}$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

$$h(G, S) = \max_{g \in G} \left\{ \min_{c \in C} \|g - c\| \right\} \tag{6}$$

4.4 Medical image Segmentation Results

For comparsion, we use five origianl network FCN with 32s[15], U-Net[23], U-Net++[29] , CE-Net[7] and U-Net with Attention Gate[24] to evaluate our proposaed method.

我们利用U-Net，ResU-Net作为baseline，与我们提出的模型分别在Cell，Liver，COVID19三个数据集上进行对比，每次实验所使用的参数，训练集，验证集，测试集均相同。

4.4.1 Live segmentation results

Segmentation results of cell segmentation are shown in tables 4.4.1，在Dice. and F1 score指数上相比于 CE-Net, U-Net++, U-Net with Attention Gate, U-Net and FCN32s分别提升了1.30%, 1.49%, 4.35%, 6.11% and 16.95%. 在Iou.指数上，分别提升了2.05, 2.27%, 6.08%, 7.30% and 19.70%. 在Sens.指数上分别提升了9.80%, 6.53%, 7.39%, 11.81% and 15.36%. 在Hd.指数上分别降低了0.4074, 0.9083, 0.7471, 0.8825 and 2.4968.
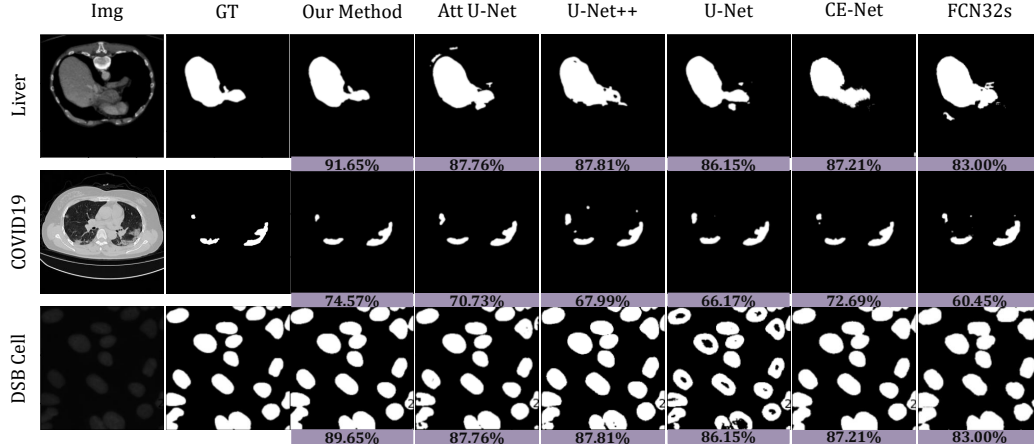
图 6: Medical image segmentation examples.

Datasets	Methods	Shape Loss	Dice.	F1 score.	Iou.	Sens.	Hd.
Cell	Our proposal	√	**0.8588**	**0.8588**	**0.7623**	**0.9296**	**4.6224**
	CENet	×	0.8458	0.8458	0.7418	0.8316	5.2098
	UNet++	×	0.8439	0.8439	0.7396	0.8643	5.5307
	Attention UNet	×	0.8153	0.8153	0.7015	0.8557	5.3695
	UNet	×	0.7977	0.7977	0.6893	0.8115	5.5049
	FCN32s	×	0.6895	0.6895	0.5653	0.7760	7.1192
Liver	Our proposal	√	**0.9551**	**0.9551**	**0.9165**	**0.9389**	**3.8854**
	U-Net++	×	0.9351	0.9351	0.8781	0.9156	5.8218
	Attention UNet	×	0.9346	0.9346	0.8776	0.9056	4.836
	CENet	×	0.9315	0.9315	0.8721	0.9045	4.904
	U-Net	×	0.9253	0.9253	0.8615	0.9106	6.6785
	FCN32s	×	0.9065	0.9065	0.8300	0.9381	7.97
COVID19	Our proposal	√	**0.8489**	**0.8489**	**0.7457**	**0.8570**	**4.313**
	CENet	×	0.8348	0.8348	0.7290	0.9359	4.7711
	UNet++	×	0.8014	0.8014	0.6799	0.9426	5.0301
	Attention UNet	×	0.8229	0.8229	0.7073	0.9435	4.8845
	UNet	×	0.7874	0.7874	0.6617	0.9528	5.2231
	FCN32s	×	0.7409	0.7409	0.6045	0.9935	5.8644

表 3: Comparsion With Other Methods In Cell[2], Liver[3] and COVID19[10] Dataset

4.4.2 Cell segmentation results

Segmentation results of cell segmentation are shown in table4.4.1, 在Dice. and F1 score指数上相比于 CE-Net, U-Net++, U-Net with Attention Gate, U-Net and FCN32s分别提升了2.01%, 2.05%, 2.36%, 2.98% and 4.86%. 在Iou.指数上，分别提升了3.84%, 3.89%, 4.44%, 5.50% and 8.65%. 在Sens.指数上分别提升了2.33%, 3.33%, 3.44%, 2.83% and 0.08%. 在Hd.指数上分别降低了1.9362, 0.9506, 1.0186, 2.7931 and 4.0846.

4.4.3 COVID19 segmentation results

Segmentation results of covid19 segmentation are shown in table4.4.1, 在Dice. and F1 score指数上相比于 CE-Net, U-Net++, U-Net with Attention Gate, U-Net and FCN32s分别提升了1.41%, 4.75%, 2.60%, 6.15% and 10.80%. 在Iou.指数上，分别提升了1.67%, 6.58%, 3.84%, 8.40% and

14.12%. 在Hd.指数上分别降低了0.4581, 0.7171, 0.5715, 0.9101 and 1.5514. 敏感度方面，通过观察Figure 6我们可以发现，其在分割目标图像时更加保守（改）。

综上所述，我们的模型在三个数据集上均consistently outperforms CE-Net and U-Net with Attention-Gate. Figure 6展示了我们在三个数据集上与其余5个模型的对比示例。COVID-19 ……，Liver ……，Cell ……。

5 Ablation study

To justify the effectiveness of the pretrained U-Net[23], Res-UNet[5], MRC(multi residual convolution) block, RMP block and SAR(spatial channel and gateway) Attention block in our proposed method, we conduct the following ablation study using the COVID19 and Cell dataset as examples.

5.1 Loss function

Methods	Loss	Dice.	F1 score.	Iou.	Sens.	Hd.
Our proposal	BCE	0.8076	0.8076	0.6874	0.8772	5.0112
Our proposal	BCE+DiceLoss	0.8375	0.8375	0.7282	0.8827	4.7241
Our proposal	Ours	0.8489	0.8489	0.7457	0.8570	4.4313

表 4: Comparsion With Other loss functions In COVID19[26] Dataset

为了验证我们使用的损失函数组合模型$loss_{oveall}$的有效性，我们以自己设计的模型作为baseline，在其他参数一致的情况下将其与$loss_{BCE}$和 $loss_{BCE+Dice}$在COVID19数据集上进行对比，$loss_{oveall}$ 和$loss_{BCE+Dice}$的定义如下所示：

$$loss_{oveall} = \alpha loss_{BCE} + \beta loss_{Dice} + \theta loss_{ACE} \tag{7}$$

$$loss_{BCE+Dice} = \alpha loss_{BCE} + \beta loss_{Dice} \tag{8}$$

在本次实验中，我们预先设置$\alpha = 0.5, \beta = 0.5$。对于$\theta$，因为ACEloss与图像梯度有关，其loss value的范围通常为$[10^1, 10^5]$，所以θ值应该显著小于α, β，在本次试验中，我们预先设置$\theta = 0.0005$。实验结果如表5.1所示，对比于$loss_{oveall}$和$loss_{BCE+Dice}$，在Dice.指标上我们分别提升了4.13%, 1.14%; 在F1 score.指标上我们分别提升了4.13%, 1.14%; 在Iou.指标上我们分别提升了5.83%, 1.75%; 在Sens.指标上我们分别降低了2.02%, 2.57%; 在Hd.指标上我们分别降低了0.5799, 0.2928，其推理结果如图7所示。

5.2 Modules

Our proposed method is based on ResU-Net and our high level feature extractor is based on CE-Net, therefore ResU-Net as well as CE-Net are the most fundamental baseline model. To further verify that the modules designed by ourselves are making contribute to segmentation task，we combined MRC Block, RMP Block, SRA Attention Block and shape loss function with ResU-Net individually：

- ResU-Net + MRC + RMP: to verify wheather these blocks can ehance the capacity to extract deep image features or not.

- ResU-Net + SAR Attention Module: to verify if our Attention module can enhance pixel-wised segmentation from spatial and channels dimensions than skip connection as well as attention-gate.
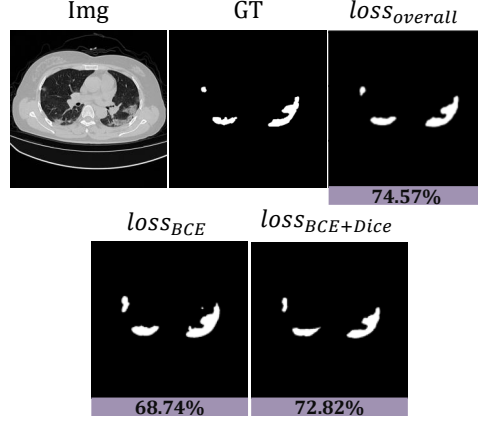
图 7: 不同损失函数训练下的模型推理结果

- ResU-Net + MRC + RMP +SAR: We use $loss_{BCE}$ and $loss_{overall}$ for our model training seperatly, which can enable us to verify the effectiveness of our model。

Dataset	Methods	Loss	Dice.	F1 score.	Iou.	Sens.	Hd.
COVID19	U-Net	BCE	0.7874	0.7874	0.6617	0.9528	5.2231
	ResU-Net	BCE	0.8105	0.8105	0.6923	0.9601	5.0248
	ResU-Net + MRC + RMP	BCE	0.8185	0.8185	0.7030	0.9295	4.8931
	ResU-Net + SAR	BCE	0.7988	0.7988	0.6846	0.7969	5.1522
	ResU-Net+SAR+MRC+RMP	BCE	0.8076	0.8076	0.6874	0.8772	5.0112
	ResU-Net+SAR+MRC+RMP	Ours	**0.8489**	**0.8489**	**0.7457**	**0.8570**	**4.313**
Cell	U-Net	BCE	0.7977	0.7977	0.6893	0.8115	5.5049
	ResU-Net	BCE	0.8314	0.8314	0.7274	0.9685	5.5861
	ResU-Net + MRC + RMP	BCE	0.8448	0.8448	0.7414	0.9714	5.0647
	ResU-Net + SAR	BCE	0.8545	0.8545	0.7534	0.9669	4.9601
	ResU-Net+SAR+MRC+RMP	BCE	0.8525	0.8525	0.7598	0.9744	4.9210
	ResU-Net+SAR+MRC+RMP	Ours	**0.8588**	**0.8588**	**0.7623**	**0.9296**	**4.6224**

表 5: Ablation study for each component on COVID19 and Cell dataset

COVID19 ablation study result 我们将上述模型在COVID19数据集上进行测试并对比其IoU, F1 score, Dice, Sensitivity and Hd value. 每次实验所使用的参数，训练集，验证集，测试集均相同。如图8和表5.2所示：ResU-Net with MRC block and RMP block 对比于ResU-Net,Iou.指标提升了1.40%, F1 score指标提升了1.34%, Dice指标提升了1.34%, Sensitivity指标提升了0.29%, Hd指标降低了0.5214。ResU-Net with SRA attention module 对比于ResU-Net,Iou.指标提升了2.60%, F1 score指标提升了2.31%, Dice指标提升了2.31%, Sensitivity指标降低了0.16%, Hd指标降低了0.6260。ResU-Net with both SRA attention module and MRC, RMP module对比于ResU-Net,Iou.指标提升了3.24%, F1 score指标提升了2.11%, Dice指标提升了2.11%, Sensitivity指标提升了0.59%, Hd指标降低了0.6651。ResU-Net with both SRA attention module and MRC with the combination of shape loss function, Compared RMP module to ResU-Net,Iou.指标提升了3.49%, F1 score指标提升了2.74%, Dice指标提升了2.74%, Sensitivity指标降低了3.89%, Hd指标降低了0.9637。

Cell ablation study result We further compare our components's performance on Cell dataset. 我们将上述模型在Cell数据集上进行测试并对比其IoU, F1 score, Dice, Sensitivity and Hd value. 每
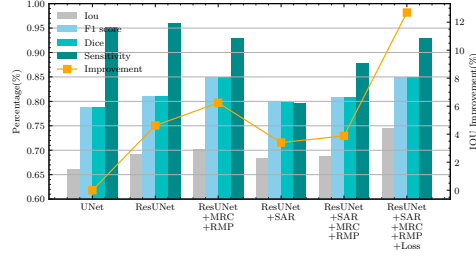
图 8: Metric results of COVID19 segmentation compared with different models.

次实验所使用的参数，训练集，验证集，测试集均相同。如图9和表5.2所示： ResU-Net with MRC block and RMP block 对比于ResU-Net,Iou.指标提升了1.40%, F1 score指标提升了1.34%，Dice指标提升了1.34%, Sensitivity指标提升了0.29%, Hd指标降低了0.5214。 ResU-Net with SRA attention module 对比于ResU-Net,Iou.指标提升了2.60%, F1 score指标提升了2.31%,Dice指标提升了2.31%, Sensitivity指标降低了0.16%, Hd指标降低了0.6260。 ResU-Net with both SRA
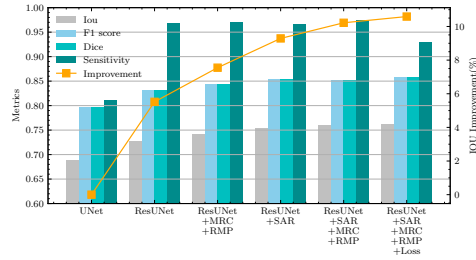

图 9: Metric results of Cell segmentation compared with different models.

attention module and MRC, RMP module对比于ResU-Net,Iou.指标提升了3.24%, F1 score指标提升了2.11%, Dice指标提升了2.11%, Sensitivity指标提升了0.59%, Hd指标降低了0.6651。 ResU-Net with both SRA attention module and MRC with the combination of shape loss function, RMP module对比于ResU-Net,Iou.指标提升了3.49%, F1 score指标提升了2.74%,Dice指标提升了2.74%, Sensitivity指标降低了3.89%, Hd指标降低了0.9637。

Furthermore, 通过将上述模型的IoU指标进行对比，我们可以发现相比于ResU-Net，both MRC+RMP module and SAR attention module 可以有效地提高模型的整体性能，同时两者的结合也能起到不错的提升效果（分别提升了 1.84%, 0.64%)。同时，由于MRC block是CE-Net中DAC block的修改，所以我们也将该模型与CE-Net进行了对比，在IoU识别率仅仅降低了0.04%的情况下，我们的模型实现了30.17%的参数精简，提高了训练的速度和节省了模型的显存消耗。

6 Conclusion

U-Net is the most prominent deep network in this regard, which has been the most popular architecture in the medical imaging community.

Deep learning (DL) based semantic segmentation methods have been providing state-of-the-art performance in the last few years.

Acknowledgments

参考文献

[1] I. Arganda-Carreras, S. C. Turaga, D. R. Berger, D. Cireşan, A. Giusti, L. M. Gambardella, J. Schmidhuber, D. Laptev, S. Dwivedi, J. M. Buhmann, T. Liu, M. Seyedhosseini, T. Tasdizen, L. Kamentsky, R. Burget, V. Uher, X. Tan, C. Sun, T. D. Pham, E. Bas, M. G. Uzunbas, A. Cardona, J. Schindelin, and H. S. Seung. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Frontiers in Neuroanatomy*, 9:142, 2015.

[2] J. C. Caicedo, A. Goodman, K. W. Karhohs, and B. A. Cimini. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature Methods*, 16(12):1247–1253, Dec 2019.

[3] P. Christ. Lits – liver tumor segmentation challenge (lits17).

[4] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. B. Ünal, and W. Wells, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016 - 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, volume 9901 of *Lecture Notes in Computer Science*, pages 424–432, 2016.

[5] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, Apr 2020.

[6] H. Gao, B. Zheng, D. Pan, and X. Zeng. Covariance self-attention dual path unet for rectal tumor segmentation. *CoRR*, abs/2011.02880, 2020.

[7] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Transactions on Medical Imaging*, 38(10):2281–2292, Oct 2019.

[8] C. Guo, M. Szemenyei, Y. Yi, and W. Zhou. Channel attention residual u-net for retinal vessel segmentation. *CoRR*, abs/2004.03702, 2020.

[9] N. Ibtehaz and M. S. Rahman. Multiresunet : Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 2020.

[10] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng, W. Hao, M. Sen, Y. Xiaoyu, N. Ziwei, L. Chen, T. Lu, Z. Yuntao, Z. Qiongjie, D. Guoqiang, and H. Jian. COVID-19 CT Lung and Infection Segmentation Dataset, Apr. 2020.

[11] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. 2017.

[12] Y. Lan and X. Zhang. Real-time ultrasound image despeckling using mixed-attention mechanism based residual unet. *IEEE Access*, 8:195327–195340, 2020.

[13] C. Li, Y. Tan, W. Chen, X. Luo, Y. Gao, X. Jia, and Z. Wang. Attention unet++: A nested attention-aware u-net for liver CT image segmentation. In *IEEE International Conference on Image Processing, ICIP 2020, Abu Dhabi, United Arab Emirates, October 25-28, 2020*, pages 345–349. IEEE, 2020.

[14] W. Liu, Y. Sun, and Q. Ji. Mdan-unet: Multi-scale and dual attention enhanced nested u-net architecture for segmentation of optical coherence tomography images. *Algorithms*, 13(3):60, 2020.

[15] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038, 2014.

[16] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 3431–3440. IEEE Computer Society, 2015.

[17] C. A. Lupascu, D. Tegolo, and E. Trucco. FABC: retinal vessel segmentation using adaboost. *IEEE Trans. Inf. Technol. Biomed.*, 14(5):1267–1274, 2010.

[18] F. Milletari, N. Navab, and S. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Fourth International Conference on 3D Vision, 3DV 2016, Stanford, CA, USA, October 25-28, 2016*, pages 565–571. IEEE Computer Society, 2016.

[19] M. Noori, A. Bahri, and K. Mohammadi. Attention-guided version of 2d unet for automatic brain tumor segmentation. *CoRR*, abs/2004.02009, 2020.

[20] O. Oktay, J. Schlemper, L. L. Folgoc, M. C. H. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention u-net: Learning where to look for the pancreas. *CoRR*, abs/1804.03999, 2018.

[21] B. Peter, S. C. Norman, M. J. Graham, and C. T. M. Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PLOS ONE*, 7(3):1–12.

[22] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. W. III, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015.

[23] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[24] J. Schlemper, O. Oktay, M. Schaap, M. P. Heinrich, B. Kainz, B. Glocker, and D. Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *CoRR*, abs/1808.08114, 2018.

[25] K. Trebing and S. Mehrkanoon. Smaat-unet: Precipitation nowcasting using a small attention-unet architecture. *CoRR*, abs/2007.04417, 2020.

[26] Websites. Covid-19 ct segmentation dataset.

[27] S. Yousefi, H. Sokooti, M. S. Elmahdy, I. M. Lips, M. T. M. Shalmani, R. T. Zinkstok, F. J. Dankers, and M. Staring. Esophageal tumor segmentation in ct images using dilated dense attention unet (ddaunet). *CoRR*, abs/2012.03242, 2020.

[28] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation. *CoRR*, abs/1807.10165, 2018.

[29] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation, 2020.