# Multi-Scale Adversarial Learning and Difficult Supervision for Kidney and Kidney Tumor Segmentation Methods

Shenhai Zheng
zhengsh@cqupt.edu.cn
Qiuyu Sun
sunqiuyuoo@163.com
Xin Ye
yexin990905@163.com
Weisheng Li
liws@cqupt.edu.cn
Laquan Li✉
lilq@cqupt.edu.cn

Chongqing University of Posts and Telecommunications,
Chongqing, China

**Abstract**

Automatic segmentation of kidney and kidney tumor areas plays an important role in radiotherapy and clinical practice. In recent years, deep learning methods have been widely used in the segmentation task and have achieved remarkable achievements. However, automatic segmentation of kidney and kidney tumors is still challenging due to their diverse shapes, complex types and unpredictable locations. Inspired by the deep supervision strategy, in this paper, we propose a cascade based approach with multi-scale adversarial learning and difficult supervision to address these challenges. On the whole, the proposed method follows the typical cascade strategy, where coarse segmentation is performed first and then fine segmentation is implemented. For the coarse segmentation part, we use Res-UNet to obtain regions of interest for kidneys and masses (include tumor and cyst). In the fine segmentation part, we propose a Multi-Scale Adversarial Learning Difficulty Supervised UNet (MSALDS-UNet) as our fine-segmented network, which consists of a segmentation network and multiple discriminators. It applies adversarial learning strategies at multiple scales of the segmentation network to improve the final segmentation performance. This is similar to the motivation of deep supervision. In addition, we also propose a difficult region supervised loss applied in MSALDS-UNet to utilize the structured information to better segment hard-to-segment regions such as fuzzy boundaries. This proposed approach can transform multi-class segmentation tasks into multiple simple binary segmentation problems. A thorough validation on the dataset provided by the 2021 Kidney and Kidney Tumor Segmentation Challenge (KiTS21) shows that our model achieves satisfactory results in kidney and kidney tumor segmentation.

# 1 Introduction

Kidney cancer is a major urinary system disease worldwide[12]. Kidney tumors have different shapes, complex types and unpredictable locations. It is still difficult for radiologists and surgeons to manually segment the kidney and its masses. Semantic segmentation of kidney and tumor tissue is the promising first step toward treatment outcomes improving[6, 14, 27]. Currently, medical imaging technologies such as computed tomography (CT) are widely used in tumor-related research. It is of great significance to predict the location, shape, benign, and malignant of kidney tumors through deep learning methods. The accuracy of the existing deep learning methods for kidney segmentation is very close to the expert level, but there is still a certain gap between the segmentation of tumors and the expert segmentation. Due to the large variation in size, location, confused intensity, and texture of kidney tumors, small tumors are difficult to distinguish from normal tissues, and accurate segmentation of kidney tumors remains a challenging task.

The traditional method of manually segmenting tumors is not only time-consuming and labor-intensive, but also has unsatisfactory clinical application results [18, 19]. Recent developments in deep learning have greatly facilitated state-of-the-art segmentation methods [17, 23] and have achieved great success in the field of tumor segmentation. In the tumor segmentation task, there are two main methods based on deep learning: one-stage and two-stage methods. One-stage methods [9, 21, 24] directly perform multi-classification tasks, obtaining multiple objects from the entire image. Oktay et al. [21] proposed the Attention U-Net, which used a novel attention gate (AG) model for medical imaging. Guo et al. [9] proposed an end-to-end model named RAU-Net, which was based on the Attention U-Net and replaced all convolutional blocks of U-Net with residual blocks. Two-stage methods [3, 7, 15, 29] mostly adopt the cascade design idea. These methods locate the region of interest in the first stage, then crop the region of interest, and use the region of interest to further segment the target. A two-stage cascade network was proposed by Lin et al. [15] which employed a cascaded framework and was designed to decompose the four-class segmentation problem into two segmentation subtasks. George, Y. [7] presented a coarse-to-fine cascaded network based on 3D U-Net architecture, which was first trained on the downsampled CT volumes and then on the full resolution images.

Many specific segmentation models have also been extensively studied in tumor segmentation tasks. UNet [23] is a semantic segmentation network based on FCN [17]. Its U-shaped structure effectively combines high-level and low-level features in the network architecture, and is very suitable for medical image segmentation. The proposed of GAN [8] draw great attention with several improvements in implementation [1, 2, 22]. Adversarial learning methods have become especially popular in medical imaging segmentation. Nie et al. [20] proposed a difficulty-aware deep segmentation network with confidence learning for end-to-end segmentation. Cirillo et al. [4] proposed Vox2Vox which was a 3D volume-to-volume generative adversarial network for segmentation of brain tumors.

Besides the network models mentioned above, these models use many mechanisms to improve segmentation performance, such as multi-scale analysis, attention mechanisms, etc. Multi-scale analysis is also widely used in medical image segmentation [5, 26] because of the use of both large-scale and small-scale features. In medical image segmentation, the irregular distribution of medical images often leads to the easy-to-segment sample dominance phenomenon. To prevent the vast number of easy samples from overwhelming the networks during training, Lin et al. [16] proposed focal loss for detection and achieved promising results. However, the focus loss may ignore the structural information of the samples, be-

cause the focus loss only uses the predicted probability as an evaluation metric of difficulty or ease. Nie et al. [20] proposed difficulty-aware attention mechanism to solve this problem. Encouraged by this, we propose a new difficult region supervision loss function to supervise difficult-to-segment regions at different scales.

# 2 Methods

Figure 1 shows the workflow of our proposed segmentation algorithm. Our method is divided into a coarse segmentation part and a fine segmentation part. The coarse segmentation part uses Res-UNet [28] for segmentation, which provides localization of kidney and mass regions for fine segmentation. The fine segmentation part uses a cascade method, and each level contains a multi-scale adversarial learning difficulty supervised UNet (MSALDS-UNet). This proposed network combines the multi-scale adversarial learning strategy and the difficult-to-segmentation region information provided by confidence maps to finely segment kidneys, masses, and tumors in turn. Among them, the tumor is further subdivided based on the segmentation of the mass. Below we describe our implementation details.

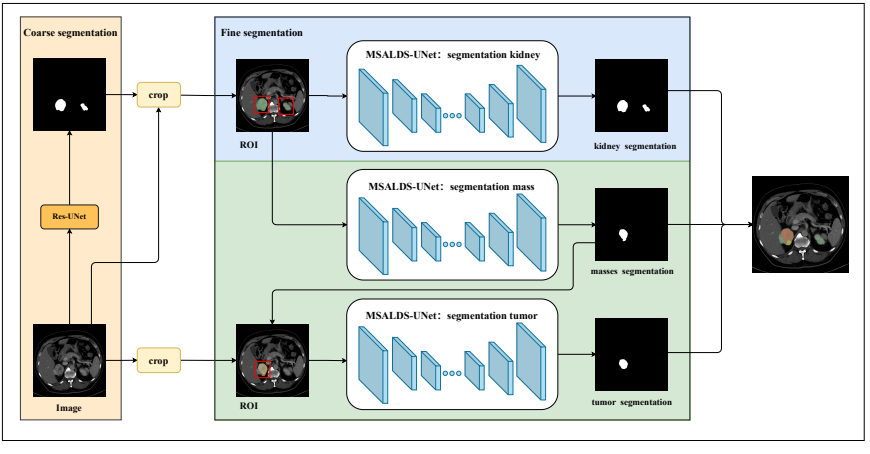


Figure 1: Architectural illustration of our proposed framework

## 2.1 Coarse segmentation

Usually, the kidney occupies only a small part of the whole CT image, while the mass and tumor are relatively smaller, and their segmentation is easily missed by irrelevant tissues. Furthermore, class imbalance makes it extremely difficult to identify and segment of whole kidney. Considering that kidney tumors are always included in the mass, and the mass is generally included in the kidney, it is feasible to first roughly locate the kidney and mass, then crop the region of interest, and then use the region of interest for fine segmentation. In this part, we use Res-UNet [28] to accomplish our task. The existing studies show that the coarse segmentation is necessary, it can remove unnecessary parts, retain specific areas and reduce background interference.

## 2.2 Fine segmentation

In this section, we first describe our cascade method, then describe our proposed MSALDS-UNet module, and finally introduce the loss function.

**Cascade segmentation.** To overcome the problem of sample imbalance during kidney, mass and tumor segmentation, we use a cascaded segmentation method to transform multi-classification problem into some binary-classification problems. Because in the multivariate segmentation, the cyst area consists of only a few pixels. However, in the binary segmentation, we consider the area of mass and tumor. The pixels of objects and backgrounds are nearly balanced in the ROI. Specifically, we use a three-level cascade. In the first and second layers, the regions of interest of the kidney and the mass obtained by the coarse segmentation are used as the input of the fine segmentation model to segment the kidney and the mass respectively. On the segmentation results of the second layer, the region of interest of the mass is cropped and used as the input of the fine segmentation model to further segment the tumor.

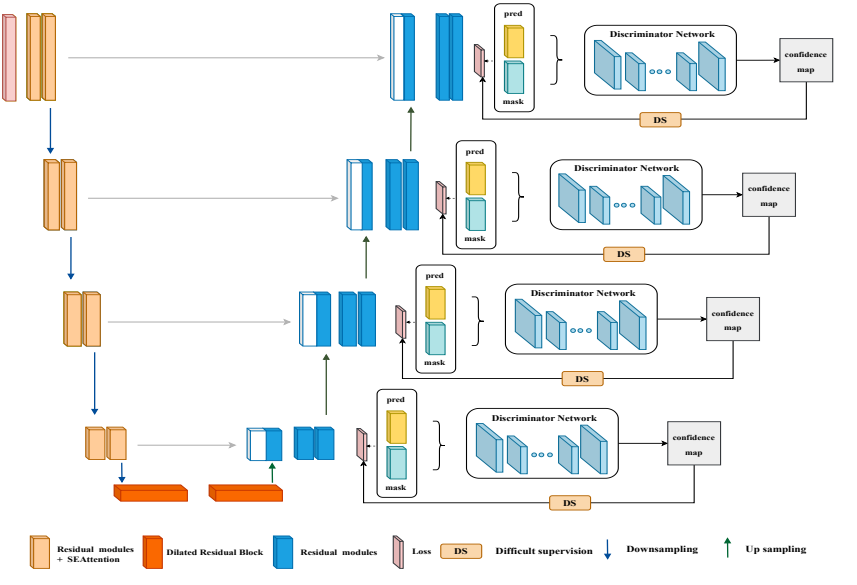


Figure 2: The overall architecture of our proposed MSALDS-UNet. The segmentation network is on the left, and the discriminator network is used on the right to perform multi-scale adversarial learning at each decoding layer of the segmentation network and supervise difficult regions.

**MSALDS-UNet approach.** As shown in Figure 2, our MSALDS-UNet adds a discriminator to each decoding layer of the segmented network to carry out adversarial learning at multiple scales of the segmented network. This is different from the traditional generative adversarial network that only carries out adversarial learning at the last layer. We use enhanced UNet as our segmentation network. Specifically, we replace the convolutional layers in UNet with residual modules [10], and apply dilated residual modules [30] in the intermedia layers between encoder and decoder to expand the receptive field and capture multi-scale contextual information. The SEAttention [11] is added to the encoding layer. It

uses the squeeze operation to perform feature compression in the spatial dimension to obtain a $1\times1\times C$ channel description with global receptive field, and then uses the excitation operation to learn the degree of dependence on each channel and adjust the feature map according to the different degree of dependence. This realize the enhancement of important features and the weakening of unimportant features.

To save memory, the discriminator network is a simplified version of UNet, keeping only one convolution layer at each stage and halving the number of feature maps in the convolution layers of the entire network. Different from traditional generative adversarial networks, the discriminator only outputs probability of an input image belonging to be the real [25], and its output is also called a confidence map, which provide structural information to regularize the output of the segmentation network [13]. Otherwise, the feature map obtained by each decoding layer of the segmentation network and the real label are spliced with the original image respectively. They are sent to the discriminator in turn, and then the discriminator outputs spatial probability maps (confidence maps) with a size of $H\times W\times 1$. Each pixel of the discriminator outputs map represents whether that pixel is sampled from the ground truth label (p = 1) or from the segmentation network (p = 0).

**Difficult area supervision.** Due to the irregular shape and small size of kidney tumors, the grayscale of kidney tumors is similar to that of the surrounding tissues and organs, and the accurate segmentation of their boundary regions has always been a challenging difficulty and a hot spot in the world. Boundary regions are difficult regions that are difficult to segment in kidney tumors, and the confidence map output by the discriminator in adversarial learning provides the confidence of correctly segmenting each local region. This information is invaluable for helping us to segment difficult regions. More importantly, the confidence maps contains information from the original input image and the predicted probability mask. It can provide structural information of easy or difficult regions. To this end, we propose a hard region supervision loss using the predicted probability mask and confidence map, which is shown in Eq. (1):

$$L_{dce} = \lambda L_{ce}(Y,P) \tag{1}$$

where $Y$ is the real mask, $P$ is the output of the decoding layer of the segmentation network. $\lambda$ as the spatial weight, it is multiplied by the cross-entropy loss $L_{ce}$ to supervise the difficult-to-segment areas, that is, the weight of the difficult-to-segment area is large, and the weight of the easy-to-segment area is small. It consists of the absolute and squared errors of the ground-truth probability mask and the confidence map. It is defined as shown in Eq. (2):

$$\lambda = (1-M)^2 + |1-M| \tag{2}$$

where $M$ is the confidence map, and $1-M$ is the difficulty of the segmentation of each pixel, and the larger the value, the more difficult the segmentation. Inspired by the $L_1$ loss and the $L_2$ loss, we propose a difficult region supervision parameter $\lambda$ in the hope of improving the supervision of difficult regions.

**Loss function.** In the segmentation network, we use a multi-task loss function as the training loss, and each decoding layer has a segmentation loss $L_{seg}$, which is defined below:

$$L_{seg} = L_{dice} + L_{dce} + \lambda_1 L_{adv} \tag{3}$$

where $L_{dice}$, $L_{dce}$, $L_{adv}$ denote the dice loss, the difficult region supervision loss, and the adversarial loss. $\lambda_1$ is the scale factor of the adversarial learning regularization term, it usually takes a very small value, in our experiments its value is 0.05 derived from experimental experience.

We also adopt the adversarial learning through the adversarial loss $L_{adv}$, defined as:

$$L_{adv}(X) = L_{bce}(D(S(X)), 1) \tag{4}$$

where $X$ is the input image, $S$ is the segmentation network, and $D$ is the discriminator. By adversarial loss, the segmentation network $S$ generates segmentation that are closer to the ground truth to fool the discriminator.

To train the discriminator network, we use a binary cross-entropy loss, defined as:

$$L_D = \frac{1}{2}(L_{bce}(D(X,P),0) + L_{bce}(D(X,Y),1)) \tag{5}$$

It worth noting that, we change from the traditional supervised segmentation network with only one scale to supervised segmentation network with multiple scales, so each decoding layer needs a loss.

# 3 Experiments

## 3.1 Datasets and Evaluation Metrics

To evaluate our proposed method, we conduct rich contrast experiments and ablation experiments on the dataset of KiTS21[1](KiTS21). It contains 300 annotated cases of abdominal CT scans with annotated regions including kidneys, tumors, and cysts. Our segmentation tasks are to segment kidney, kidney masses (tumor + cyst), and tumor from abdominal CT images. It is worth noting that KiTS21 officially provides a separate binary annotation for each region and an aggregation of cyst and tumor annotated regions (i.e., masses), which allows us to directly use the officially provided data without special processing. Furthermore, we quantitatively evaluate the segmentation results using evaluation metrics commonly used in medical image segmentation tasks, including Dice Similarity Coefficient (DSC) and Jaccard Similarity Coefficient (JC).

## 3.2 Pre-processing and Post-processing

Our data pre-processing includes three steps of normalization, resampling and data augmentation. We adjust the CT scan to a window width of 540 and a window level of 140, and normalize it to [0, 255], and then convert the 3D image into a 2D PNG format slice according to the cross-sectional direction, finally resize all slices to [128, 128]. During training, we use different angles of rotation, horizontal and vertical mirroring, and add random salt and pepper noise and Gaussian noise for data augmentation.

For post-processing, we first concatenate the segmentation masks obtained from each cascade part to obtain a complete three-category segmentation mask, then resize the three-category segmentation mask to the original CT image size, and finally integrate the resized segmentation mask into the original image to get the final displayed segmentation result.

## 3.3 Training and Implementation Details

All our experiments are based on Python 3.7, PyTorch 1.7.1, and Ubuntu 20.04 LTS, and are performed under the hardware conditions of AMDő Ryzen 9 9500x 12-core processor and

[1] https://kits21.kits-challenge.org/

NVIDIA GeForce RTX 3090 graphics card. We divide the dataset into training, validation and test sets in a ratio of 8:1:1. All models are trained from scratch using the Adam optimizer with an initial learning rate of 0.0001, and the learning rate during training is adjusted according to Eq.(6)

$$lr = lr_0 \times \gamma^{\frac{last\_epoch}{step\_size}} \tag{6}$$

where the *step_size* represents the step size of the learning rate update, which is updated every *step_size* epoch (1 in our experiments). $\gamma$ represents the decay rate of the learning rate, after each *step_size* epochs (0.1 in our case), the learning rate becomes $lr * \gamma$. After *last_epoch* epochs, the learning rate reverts to the initial learning rate.

We record the training loss and validation loss for each epoch during the training process, and define the maximum number of iterations as 200 epochs. After every 50 epochs, we save the model parameters, and save the model with the smallest validation loss as the optimal model. All our experimental results are based on the reasoning of this optimal model on the test set.

# 4 Experiment Results

## 4.1 Comparisons with state-of-the-art methods

To further demonstrate the superiority of our proposed algorithm, we compare it with several other state-of-the-art segmentation methods on the KiTS21 dataset. The compared methods and quantitative analysis results are shown in Table 1. It can be found that all methods have higher scores on kidneys. However, due to the smaller size and different shapes of tumors and cysts, the DSC and JC scores of both masses (tumor + cyst) and tumors are significantly lower than those of kidneys. While as shown in Table 1 (the best result are shown in bold), our proposed method achieves the best DSC and JC values in the segmentation of kidneys, masses, and tumors. And our method outperforms other methods by about 2% and 3% (RAU-Net) on the DSC scores of masses and tumors. This shows the superiority of our method in masses and tumors segmentation.

| Method | DSC | | | JC | | |
|---|---|---|---|---|---|---|
| | kidney | masses | tumor | kidney | masses | tumor |
| Vox2Vox [4] | 0.8970 | 0.6945 | 0.7464 | 0.8309 | 0.5942 | 0.6502 |
| Lin et al. [15] | 0.9211 | 0.6772 | 0.6927 | 0.8781 | 0.5883 | 0.6035 |
| George et al. [7] | 0.9425 | 0.6783 | 0.7780 | 0.9008 | 0.5873 | 0.6863 |
| Res-UNet [28] | 0.9233 | 0.7617 | 0.7819 | 0.8763 | 0.6768 | 0.6985 |
| Nie et al. [20] | 0.9566 | 0.7856 | 0.7923 | 0.9237 | 0.7048 | 0.7109 |
| Attention-UNet [21] | 0.9469 | 0.8235 | 0.8341 | 0.9103 | 0.7455 | 0.7584 |
| RAU-Net [9] | 0.9504 | 0.8284 | 0.8409 | 0.9138 | 0.7510 | 0.7655 |
| **Ours** | **0.9610** | **0.8449** | **0.8746** | **0.9278** | **0.7714** | **0.8051** |

Table 1: Quantitative comparison between our proposed method and other state-of-the-art methods on the KiTS21 testing dataset.

Figure 3 presents the violin plots to shows the data distribution of DSC and JC scores. The thin line in the middle of each violin plot represents the data range, and the data beyond the thin line is the outlier. The black box inside the violin body shows the interquartile range

of the data, and the white point inside the black box is the median. The width of the violin body represents the probability density of the data distribution at that point, and the wider the width, the denser the data distribution at that point. As can be seen from the figure 3, the median of our method leads both under the DSC and JC evaluation metrics. At the same time, the results obtained by our method show a more concentrated data distribution with significantly fewer outliers than other methods. This shows that our method has stronger stability while obtaining more accurate segmentation results, and has more advantages for the segmentation of difficult samples.

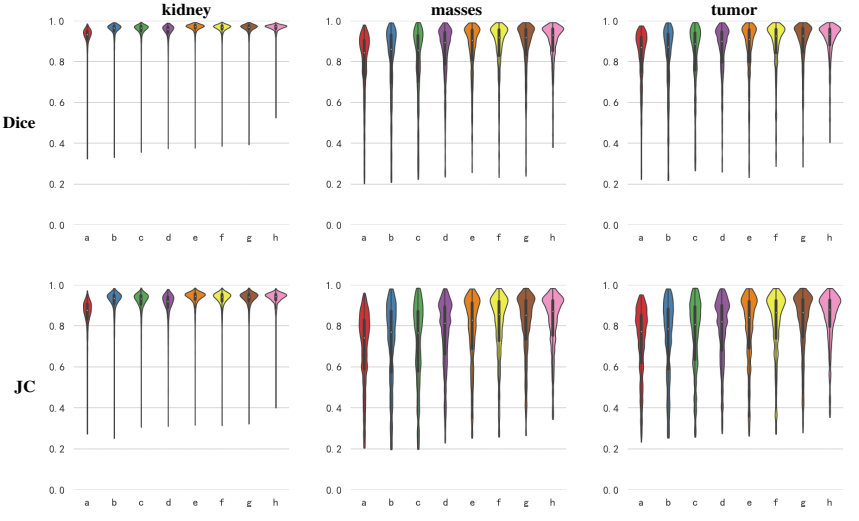


Figure 3: Violin plots obtained by different segmentation methods on the test set. a-h correspond to the methods in Table 1 one-to-one (the order from top to bottom).

Figure 4 shows the visualization results of samples under different segmentation methods. We select four different patients in the test set for comparison to show more intuitive segmentation results. It can be seen from the figure that all methods have achieved relatively good results in the segmentation of kidneys, but in the segmentation of masses and tumors, some methods have the phenomenon of under-segmentation or over-segmentation, especially in the boundary area. In contrast, our segmentation results are closer to the real labels, and the segmentation of boundary regions is better.

## 4.2 Ablation experiment

To demonstrate the performance of the proposed method and illustrate the relevance of the different modules, in this section, we performe ablation studies on the KiTS21 dataset. Table 2 shows the results of ablation experiments of our proposed method on the KiTS21 dataset and highlights the impact of each component applied to the model. We sequentially validate the effects of coarse-to-fine structure, multi-scale adversarial learning and difficult supervised loss on the model. "Without coarse segmentation" means that only fine segmentation is used instead of coarse segmentation. It is used to verify the influence of coarse to fine

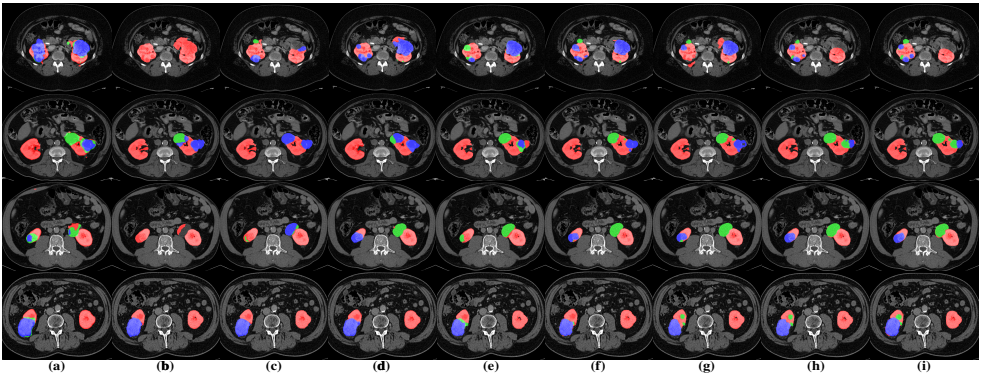


Figure 4: Example of kidney, masses and tumor segmentation results of our method and other methods on the test dataset. The red, blue and green areas represent kidneys, tumors and cysts, respectively. (a)-(h) correspond to the methods in Table 1 one-to-one (the order from top to bottom), the (i) represents the ground truth.

structure on the model. Baseline represents the use of coarse segmentation, and in fine segmentation, the discriminator is only used in the last layer of the segmentation network. This structure is also the structure of the traditional GAN. "MSAL" is multi-scale adversarial learning. "Baseline+MSAL" is based on Baseline, adding discriminators to other decoding layers, and performing adversarial learning on multiple scales. "Baseline + MSAL + DS" is our proposed method, where DS is our proposed difficult regions supervision mechanism.

| Method | DSC | | | JC | | |
|---|---|---|---|---|---|---|
| | kidney | masses | tumor | kidney | masses | tumor |
| Without Coarse segmentation | 0.9567 | 0.7983 | 0.8352 | 0.9225 | 0.7139 | 0.7572 |
| Baseline | 0.9618 | 0.8151 | 0.8511 | 0.9299 | 0.7384 | 0.7786 |
| Baseline + MSAL | **0.9620** | 0.8370 | 0.8620 | **0.9307** | 0.7648 | 0.7916 |
| **Baseline + MSAL + DS(Ours)** | 0.9610 | **0.8449** | **0.8746** | 0.9278 | **0.7714** | **0.8051** |

Table 2: Values of DSC and JC in ablation experiments.

It can be seen from the Table 2 that the segmentation results of masses and tumors are significantly improved after using coarse segmentation (the best result are shown in bold). In addition, after introducing multi-scale adversarial learning(MSAL), the DSC scores of our proposed method for kidney, masses, and tumor are 0.9620, 0.8370, and 0.8620, respectively. And the DSC scores for kidney, masses, and tumor after introducing difficult region supervision loss(DS) are 0.9610, 0.8449, and 0.8746, respectively. As a result, it can be found that the segmentation effects of masses and tumor have been improved to varying degrees under the condition that the kidney segmentation effect is not very different.

# 5 Conclusions

In this paper, we propose a cascade approach based on multi-scale adversarial learning and difficult supervision to obtain accurate segmentation of kidney and kidney tumor. Specif-

ically, we use the cascade strategy to first roughly segment the kidney and mass regions, and then refine them based on MSALDS-UNet. In the fine segmentation, the multi-class segmentation problem is transformed into three simple binary segmentation problems to alleviate the problem of sample imbalance and reduce computation. Importantly, we propose a refine segmentation method named MSALDS-UNet, which uses a multi-scale adversarial learning strategy to better train the segmentation network and obtain more accurate segmentation results by using multi-scale information and adversarial learning. On this basis, we propose a difficult region supervision loss, which uses structured information to improve the segmentation effect on difficult-to-segment regions such as boundary regions. A limitation of this study is that we segment the tumor in the mass, and the mass minus the tumor is the cyst. If the tumor segmentation error is large, the segmentation error for the cyst will also increase accordingly. However, compared to directly segmenting cysts with extremely imbalanced samples, our method can achieve much better segmentation results. Experimental results demonstrate that our method provides an accurate and robust solution for kidney and kidney tumor segmentation compared the state-or-the-art.

## Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 61902046 and 61901074), the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K202200606), the Natural Science Foundation of Chongqing (Grant No. 2022NSCQ-MSX3746) and China Postdoctoral Science Foundation (Grant No. 2021M693771)

## References

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR. ISBN 2640-3498.

[2] David Berthelot, Thomas Schumm, and Luke Metz. Began: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717*, 2017.

[3] Jason Causey, Jonathan Stubblefield, Jake Qualls, Jennifer Fowler, Lingrui Cai, Karl Walker, Yuanfang Guan, and Xiuzhen Huang. An ensemble of u-net models for kidney tumor segmentation with ct images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021. ISSN 1545-5963.

[4] Marco Domenico Cirillo, David Abramian, and Anders Eklund. Vox2vox: 3d-gan for brain tumour segmentation. In *International MICCAI Brainlesion Workshop*, pages 274–284. Springer.

[5] Tongle Fan, Guanglei Wang, Yan Li, and Hongrui Wang. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020. ISSN 2169-3536.

[6] Vincenzo Ficarra, Giacomo Novara, Silvia Secco, Veronica Macchi, Andrea Porzionato, Raffaele De Caro, and Walter Artibani. Preoperative aspects and dimensions used

for an anatomical (padua) classification of renal tumours in patients who are candidates for nephron-sparing surgery. *European urology*, 56(5):786–793, 2009. ISSN 0302-2838.

[7] Yasmeen George. A coarse-to-fine 3d u-net network for semantic segmentation of kidney ct scans. In *International Challenge on Kidney and Kidney Tumor Segmentation*, pages 137–142. Springer.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[9] Jingna Guo, Wei Zeng, Sen Yu, and Junqiu Xiao. Rau-net: U-net model based on residual and attention for kidney and kidney tumor segmentation. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pages 353–356. IEEE. ISBN 1728183197.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

[11] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141.

[12] Junjie Huang, David Ka-Wai Leung, Erica On-Ting Chan, Veeleah Lok, Sophia Leung, Iris Wong, Xiang-Qian Lao, Zhi-Jie Zheng, Peter Ka-Fung Chiu, and Chi-Fai Ng. A global trend analysis of kidney cancer incidence and mortality and their associations with smoking, alcohol consumption, and metabolic syndrome. *European Urology Focus*, 8(1):200–209, 2022. ISSN 2405-4569.

[13] Wei-Chih Hung, Yi-Hsuan Tsai, Yan-Ting Liou, Yen-Yu Lin, and Ming-Hsuan Yang. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*, 2018.

[14] A. Kutikov and R. G. Uzzo. The r.e.n.a.l. nephrometry score: a comprehensive standardized system for quantitating renal tumor size, location and depth. *J Urol*, 182(3): 844–53, 2009. ISSN 1527-3792 (Electronic) 0022-5347 (Linking). doi: 10.1016/j.juro.2009.05.035. URL https://www.ncbi.nlm.nih.gov/pubmed/19616235.

[15] Chaonan Lin, Rongda Fu, and Shaohua Zheng. Kidney and kidney tumor segmentation using a two-stage cascade framework. Kidney and Kidney Tumor Segmentation, pages 59–70. Springer International Publishing, . ISBN 978-3-030-98385-7.

[16] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, .

[17] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

[18] Suhuai Luo, Rongxin Li, and Sébastien Ourselin. A new deformable model using dynamic gradient vector flow and adaptive balloon forces. In *APRS workshop on digital image computing, Brisbane, Australia*.

[19] Guangrui Mu, Zhiyong Lin, Miaofei Han, Guang Yao, and Yaozong Gao. Segmentation of kidney tumor by multi-resolution vb-nets. 2019.

[20] Dong Nie, Li Wang, Lei Xiang, Sihang Zhou, Ehsan Adeli, and Dinggang Shen. Difficulty-aware attention network with confidence learning for medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1085–1092. ISBN 2374-3468.

[21] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, and Bernhard Kainz. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[22] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

[24] D Sabarinathan, M Parisa Beham, and SM Mansoor Roomi. Hyper vision net: kidney tumor segmentation using coordinate convolutional layer and attention unit. In *National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics*, pages 609–618. Springer.

[25] Mohammad Sabokrou, Masoud Pourreza, Mohsen Fayyaz, Rahim Entezari, Mahmood Fathy, Jürgen Gall, and Ehsan Adeli. Avid: Adversarial visual irregularity detection. In *Asian Conference on Computer Vision*, pages 488–505. Springer.

[26] Ashish Sinha and Jose Dolz. Multi-scale self-guided attention for medical image segmentation. *IEEE journal of biomedical and health informatics*, 25(1):121–130, 2020. ISSN 2168-2194.

[27] Ahmed Taha, Pechin Lo, Junning Li, and Tao Zhao. Kid-net: convolution networks for kidney vessels segmentation from ct-volumes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 463–471. Springer.

[28] Xiao Xiao, Shen Lian, Zhiming Luo, and Shaozi Li. Weighted res-unet for high-quality retina vessel segmentation. In *2018 9th international conference on information technology in medicine and education (ITME)*, pages 327–331. IEEE. ISBN 153867744X.

[29] Xu Yan, Kun Yuan, Weibing Zhao, Sheng Wang, Zhen Li, and Shuguang Cui. An efficient hybrid model for kidney tumor segmentation in ct images. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 333–336. IEEE. ISBN 1538693305.

[30] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 472–480.