HAL
open science

Meta Grayscale Adaptive Network for 3D Integrated Renal Structures Segmentation

Yuting He, Guanyu Yang, Jian Yang, Rongjun Ge, Youyong Kong, Xiaomei Zhu, Shaobo Zhang, Pengfei Shao, Huazhong Shu, Jean-Louis Dillenseger, et al.

To cite this version:

Yuting He, Guanyu Yang, Jian Yang, Rongjun Ge, Youyong Kong, et al.. Meta Grayscale Adaptive Network for 3D Integrated Renal Structures Segmentation. Medical Image Analysis, 2021, 71, pp.102055. 10.1016/j.media.2021.102055 . hal-03195538

HAL Id: hal-03195538

https://hal.science/hal-03195538v1

Submitted on 11 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Meta Grayscale Adaptive Network for 3D Integrated Renal Structures Segmentation

Yuting He[a], Guanyu Yang[a,e,*], Jian Yang[c], Rongjun Ge[a], Youyong Kong[a,e], Xiaomei Zhu[d], Shaobo Zhang[f], Pengfei Shao[f], Huazhong Shu[a,e], Jean-Louis Dillenseger[b,e], Jean-Louis Coatrieux[b], Shuo Li[g,*]

[a]*LIST, Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing, China*
[b]*Univ. Rennes 1, Inserm, LTSI - UMR 1099, F-35000 Rennes, France*
[c]*Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Electronics, Beijing Institute of Technology, Beijing, China*
[d]*Dept. of Radiology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, China*
[e]*Centre de Recherche en Information Biomédicale Sino-Français (CRIBs)*
[f]*Dept. of Urology, the First Affiliated Hospital of Nanjing Medical University, Nanjing, China*
[g]*Dept. of Medical Biophysics, University of Western Ontario, London, ON, Canada*

Abstract

Three-dimensional (3D) integrated renal structures (IRS) segmentation targets segmenting the kidneys, renal tumors, arteries, and veins in one inference. Clinicians will benefit from the 3D IRS visual model for accurate preoperative planning and intraoperative guidance of laparoscopic partial nephrectomy (LPN). However, no success has been reported in 3D IRS segmentation due to the inherent challenges in grayscale distribution: low contrast caused by the narrow task-dependent distribution range of regions of interest (ROIs), and the networks representation preferences caused by the distribution variation inter-images. In this paper, we propose the Meta Greyscale Adaptive Network (MGANet), the first deep learning framework to simultaneously segment the kidney, renal tumors, arteries and veins on CTA images in one inference. It makes innovations in two collaborate aspects: **1)** The *Grayscale Interest Search* (GIS) adaptively focuses segmentation networks on task-dependent grayscale distributions via scaling the window width and center with two cross-correlated coefficients for the first time, thus learning the fine-grained representation for fine segmentation. **2)** The *Meta Grayscale Adaptive* (MGA) learning makes an image-level meta-learning strategy. It represents diverse robust features from multiple distributions, perceives the distribution characteristic, and generates the model parameters to fuse features dynamically according to image's distribution, thus adapting the grayscale distribution variation. This study enrolls 123 patients and the average Dice coefficients of the renal structures are up to 87.9%. Fine selection of the task-dependent grayscale distribution ranges and personalized fusion of multiple representations on different distributions will lead to better 3D IRS segmentation quality. Extensive experiments with promising results on renal structures reveal powerful segmentation accuracy and great clinical significance in renal cancer treatment.

Keywords: Integrated renal structures, segmentation, Ensemble learning, Meta learning, Automatic hyper-parameter search, Meta grayscale adaptive network, Grayscale interest search, Meta grayscale ensemble

1. Introduction

Three-dimensional (3D) integrated renal structures (IRS) segmentation on computed tomography angiography (CTA) images is one of the most important tasks for laparoscopic partial nephrectomy (LPN) (Shao et al., 2011; Simone et al., 2015; Ljungberg et al., 2019). It targets achieving 3D kidney, renal tumors, arteries, and veins in one inference (Fig. 1(a)), once successful, clinicians will benefit from the 3D visual model of renal structures for accurate preoperative planning (Porpiglia et al., 2018; Bianchi et al., 2020). Preoperatively, the renal arteries will help estimate renal perfusion model (Zhang et al., 2019), so that the clinicians will select the tumor-feeding arterial branches and locate the arterial clamping position easily (Shao et al., 2012). The tumor and kidney models will visually show the lesion regions, thus helping the pre-plan of the tumor resection surface (Fig. 1(b)). Intraoperatively, the preoperative plan will be displayed on screen together with laparoscopic videos to guide the LPN (Nicolau et al., 2011; Herrell et al., 2014; Hamada et al., 2020). Renal vessels (veins, arteries) outside hilum will show a clear arterial clamping region visually, thus the clinicians will select arterial clamping branches quickly. The 3D IRS visual model will also guide the clinicians making appropriate decisions (Fig. 1(c)). Therefore, the costs of treatment will be reduced, the quality of LPN will be improved, and the pain of patients will be relieved.

Although the automatic 3D IRS segmentation is clinically urgent, there is no solution yet. Most of the recent studies focus on partial renal structure segmentation (Wang et al., 2020a; Jin et al., 2016; He et al., 2019) which lacks the countermeasures to segment integrated renal structures. Besides, the separate segmentation of each structure will bring the overlapping problem, especially in the renal hilum where veins, arteries are staggered. Some other studies are dedicated to multiple renal structure segmentation (Li et al., 2018; Taha et al., 2018), but the lack of fine details limiting their clinical downstream tasks. Especially, the inadequate segmentation of renal arteries loses the important ability of perfusion region estimation.

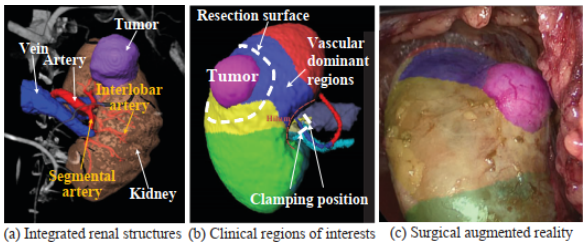

Figure 1: 3D IRS segmentation will provide accurate preoperative planning and intraoperative guidance for LPN

Inherent challenges of 3D IRS segmentation are limiting the implementation of automated method: **Challenge** 1: Narrow task-dependent distribution making low contrast. The CT image has a large grayscale range, but the renal structures are only in a narrow range, making the original CTA images appear unclear soft tissues and boundaries. As shown in Fig. 2, the irrelevant distributions occupy the large grayscale ranges squeezing the renal structures in narrow grayscale ranges, so that the CTA image has coarse-grained distributions and low contrast (a). Therefore, the network will lose the sensitivity to the adjacent structures with close grayscale distribution and weaken the ability to extract detailed features such as the boundaries between the structures. **Challenge 2**: Distribution variation inter-images making representation preferences. The subtypes of the renal tumors have different volumes and abilities to metabolize contrast agents (Young et al., 2013), so that it will make the differences in the distribution of contrast agent overall, thus making a large grayscale distribution variation inter-images. As shown in Fig. 3, the clear cell carcinomas present bright texture and large size, while the papillary has dark texture and small size. The grey distributions of five tumor subtypes have various centers and peaks. These large variations make it the network difficult to stably learn uniform low-order knowledge (style, pattern, etc.) resulting in representation preferences and large fluctuations of performance inter-images,

*Corresponding author.
Email addresses: yang.list@seu.edu.cn (Guanyu Yang), slishuo@gmail.com (Shuo Li)

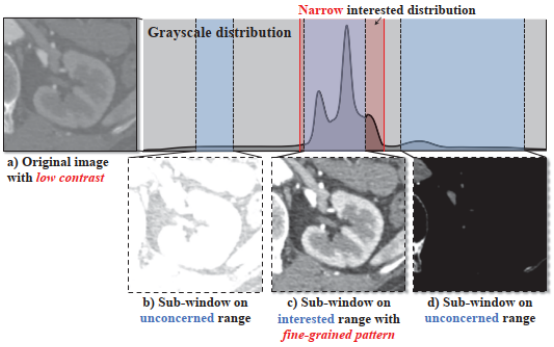

Figure 2: Narrow task-dependent distribution makes original CTA image low contrast and windowing on the distribution makes sub-window image fine-grained pattern. (a) The original CTA image has narrow distribution making low contrast. (b,d) When the window covers the unconcerned distribution, the renal structures will be lost. (c) When the window covers the task-dependent distribution, the sub-window image will have a fine-grained pattern

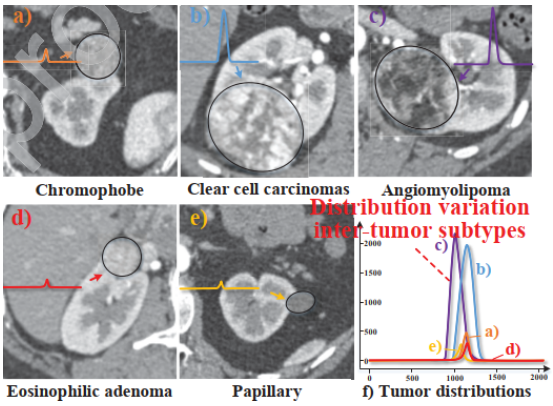

Figure 3: Different renal tumor subtypes make the images have large distribution variation. (a-e) The renal tumor subtypes have different volumes and abilities to metabolize contrast agents making distribution variations (Young et al., 2013). f) The grayscale distributions of the tumor subtypes have a large variation. The y-axis is the number of voxels and the x-axis is the grayscale value

especially for the vein structures which are inapparent and sensitive to the variation.

Solution 1 to focus on the task-dependent grayscale distribution range. Windowing (Goldman, 2007) removes irrelevant grayscale ranges and expands the task-dependent distributions which will augment the contrast for ROIs, making networks focus on a wider distribution of the renal structures to perceive fine-grained patterns. **However**, it is challenging to find the narrow task-dependent distribution in a large original grayscale range manually according to human experience. Especially for our renal structures segmentation task, it has to be dependent on multiple distributions and has big distribution variation inter-images. As shown in Fig. 2, in the grayscale distribution of the example image, when the grayscale distribution range window (sub-window) covers the renal structures' distributions (c), the sub-window image has fine-grained patterns and the regions of renal structures are significant. When the sub-window is out of the renal structure's distributions (b,d), the renal structures will be lost. **Therefore**, we propose the Grayscale Interest Search (GIS) strategy which focuses segmentation networks on the task-dependent grayscale distribution in CTA images for the first time. It innovatively simulates doctor's observation of different structures in different sub-windows to automatically search the task-dependent distribution via scaling the window width and center with two effective cross-correlated coefficients. Therefore, salient regions with fine-grain patterns are available and networks will learn the fine-grained representation for fine segmentation in details.

Solution 2 to adapt to the grayscale distribution variation. Meta-learning for parameter generation learns cross-tasks knowledge to rapidly generate parameters of models (Hospedales et al., 2020), thus dynamically adapting to the variation inter-tasks (task-level). **However**, there is no meta-learning study for the adaptation of grayscale distribution variation inter-images (image-level). **In this study**, we propose the *Meta Grayscale Adaptation (MGA)* learning (Fig. 4(c)) which makes each image as a sub-learning task to dynamically generate the parameters for each image adapting the distribution variation inter-images. It is an image-level meta-learning strategy which learns the trained kernels via gradient descent for the meta knowledge of IRS segmentation, and generates the meta kernels via prediction for the knowledge of each sub-learning task. It

has three key elements: **1)** Represent the image robustly with distribution variation. We integrate the feature representation capabilities from multiple distributions covered by different sub-windows (Multi-windowing Group) so that more diverse features will be extracted, reducing the overfitting of representation preferences. **2)** Dynamically generate personalized parameters according to image's distribution. We make each image as a sub-learning task and further propose the Meta Perceiver to perceive its distribution characteristic, thus generating personalized feature parameters (meta kernels) to adapt to the distribution variation of each sub-learning task (image). **3)** Personalized fusion of features. We propose the Meta Residual Convolution (ResConv) to fuse the meta knowledge of our IRS segmentation task and the personalized knowledge of each sub-learning task. Trained kernels learn meta knowledge on IRS segmentation and the meta kernels generated by our Meta perceiver represent the personalized knowledge of sub-learning tasks. The meta kernels fine-tune the trained kernels for the personalized fusion parameters of each image adapting to the distribution variation inter-images and improving the generalization.

We propose the grayscale distribution optimization from the perspective of the grayscale distribution range adjustment and variation adaptation in our 3D IRS segmentation task, and propose the Meta Greyscale Adaptive Network (MGANet), the first deep learning framework to simultaneously segment the kidney, renal tumors, arteries and veins on CTA images in one inference. The contributions of our study are summarized as the following:

- To the best of our knowledge, we achieve the 3D IRS segmentation for the first time which will play an important role in accurate preoperative planning and intraoperative guidance of LPN. We review the relevant technical and clinical value of this task which will provide a valuable reference for follow-up works.

- We propose the MGANet to optimize the grayscale distribution of the model representation via grayscale distribution range adjustment and variation adaptation, so that it will obtain fine-grained feature representations and personalized feature fusion for optimal segmentation results.

- We propose the MGANet to optimize the grayscale distribution of the model representation via grayscale distribution range adjustment and variation adaptation, so that it will obtain fine-grained feature representations and personalized feature fusion for optimal segmentation results.

- We present a novel image-level meta-learning strategy, the MGA learning, which represents diverse robust features from multiple distributions, perceives the distribution characteristic, and generates the model parameters to fuse these features dynamically according to the image's distribution, thus adapting the grayscale distribution variation.

The rest of the paper is organized as follow: We review the related works about renal structures segmentation, ensemble learning in image segmentation, hyper-parameter optimization, and meta-learning for parameters generalization in Sec. 2. Then, we specifically introduce our proposed MGANet in Sec. 3, including the GIS strategy (Sec. 3.1), the MGA learning (Sec. 3.2) and the details of the network structures (Sec. 3.3). Then Sec. 4 describes the materials, experiment setting, and evaluation methods, and Sec. 5 shows and analyzes the results of the experiments. Finally, we discuss and conclude our work in Sec. 6.

2. Related works

2.1. Renal structures segmentation

Renal structure segmentation is significant in accurate preoperative planning and intraoperative guidance of LPN (Porpiglia et al., 2018; Bianchi et al., 2020), however, the automatic IRS segmentation has not been reported success. He et al. (2019) designed a semi-supervised framework and achieved fine segmentation of renal arteries which is important in the selection of arterial clamping position. Wang et al. (2020) took the graph-cut (Rother et al., 2004) and

proposed a tensor-cut framework for renal artery segmentation. Taha et al. (2018) proposed a Kid-Net for renal vessel segmentation including the arteries, veins, and ureter. The 3D UNet (Çiçek et al., 2016) was firstly proposed for Xenopus kidney segmentation. Li et al. (2018) proposed a residual U-Net and segmented the renal structures.

However, these works are limited in our task in two aspects: **1)** Some works (He et al., 2019; Taha et al., 2018; Wang et al., 2020a; Çiçek et al., 2016) only focus on the segmentation of partial renal structures lacking the countermeasures for integrated renal structures in one inference. The separate segmentation of each renal structure via these methods will bring the overlapping problem within segmentation regions, especially in renal hilum where veins, arteries, and ureter are staggered. **2)** Some other works (Li et al., 2018; Taha et al., 2018) lack fine details of the segmented structures such as the interlobar renal arteries which is important to build the renal perfusion model (Zhang et al., 2019), limiting the clinical downstream tasks.

2.2. Ensemble learning in image segmentation

The ensemble learning (Dietterich, 2002) constructs a set of learners and combines their representations for a more accurate discussion. It gives a strategy to embed specific designs for the challenges of tasks, such as error correction, class-imbalanced data, etc. (Polikar, 2012). The ensemble segmentation models (Xia et al., 2018; Wang et al., 2015; Zheng et al., 2019) fuse the superiorities from the multiple learners improving the integrated segmentation quality. Xia et al. (2018) proposed the Volumetric Fusion Net (VFN) which train three 2D segmentation network in X, Y, and Z-axis and take a 3D network to fuse the results. Zheng et al. (2019) further added a learner in the 3D version and proposed a meta-learner for the ensemble process.

Although the existing works behave promising results on their issues, they are inherently impeded to be applied for our task. They ignore the variation of grayscale distributions caused by the variation of tumor types in our task, making their networks difficult to stably learn uniform low-order knowledge (style, pattern, etc), resulting in representation preferences for common distributions and limiting the generalization for rare distributions. Our MGA learning trains multiple networks on different searched distributions and fuses their feature representations abilities according to the distribution characteristic dynamically. Thereby, their representation for different distributions will be jointed making generalization for distribution variation inter-images.

2.3. Hyper-parameter optimization

Hyper-parameter optimization (Bergstra and Bengio, 2012; Tan and Le, 2019) is widely used to achieve better accuracy. Numerous prior studies (He et al., 2016; Huang et al., 2017; Real et al., 2019) on network's depth and width proofs that the deeper and wider networks will capture richer and more complex features for more accurate performance. Tan and Le (2019) simultaneously studied the influence of the width, depth, and resolution's compound scaling on network accuracy, and searched the EfficientNet which uses less computing resourcesfor higher accuracy.

In medical image analysis, the grayscale distribution ranges of the medical images are another important hyper-parameter that lacks the attention of researchers. The CT imaging process makes a large grayscale range so that the original image will appear in low contrast and unclear soft tissues. When observing medical images, radiologists use the windowing method (Goldman, 2007) to removes irrelevant grayscale ranges and expands the distributions of the ROIs, thus improving the contrast and bringing clearer details. However, there is no study on the optimization of this hyper-parameter: **1)** numerous medical image analysis studies (He et al., 2019; Çiçek et al., 2016; Li et al., 2018; Taha et al., 2018) ignore the optimization of this hyper-parameter and train their models on original grayscale ranges, limiting the segmentation quality on their regions of interest (ROIs); **2)** Some other studies (Isensee et al., 2018; Wang et al., 2020b) only set the grayscale range manually based on researchers' experience or clinical criteria directly without further optimization for the better grayscale range in their task. We propose the GIS strategy which adaptively searches the task-dependent grayscale range so that the network will

learn more fine-grained feature representation for the interested distributions.

2.3.1. Meta learning for parameters generalization

Meta-learning for parameters generalization directly predict model parameters (Hospedales et al., 2020) rather than the gradient in back-propagation, thus generating the personalized parameters and providing dynamic adaptation to the variation. The Hypernetworks (Ha et al., 2017; Brock et al., 2018) train a network to generate the parameters of another network for fast learning, which are common applications in model compression or multi-task learning. The Meta networks (Munkhdalai and Yu, 2017) combine the parameters predicted by meta learner and the weights trained cross tasks making rapid generalization. The WeightNet (Ma et al., 2020) unifies the SENet (Hu et al., 2018) and CondConv (Yang et al., 2019), trains the network in the kernel space and takes the network to generate weights for the classification task.

However, these studies only focus on task-level parameter generation and there is no meta-learning study for the adaptation of grayscale distribution variation inter-images (image-level). Our MGA learning makes each image as a sub-learning task to dynamically generate parameters for each image, thus making an image-level meta-learning framework to adapt the distribution variation. It represents diverse robust features from multiple distributions and generates the model parameters to fuse these features dynamically according to the input image's distribution in a stable residual structure, thus adapting the grayscale distribution variation inter-images.

3. Methodology

As shown in Fig. 4, our MGANet adaptively optimizes the grayscale distribution of medical images and achieves the 3D IRS segmentation. It makes innovations in two collaborate components: **1)** The Grayscale Interest Search (GIS, Sec. 3.1, Fig. 4 (a)) adaptively optimizes the focused grayscale distribution range via scaling the window width and center with two cross-correlated coefficients, thus adaptively discovering the task-dependent distribution. It further expands the searched distribution via windowing making salient regions with fine-grain patterns, so that segmentation networks will learn fine-grained feature representation for renal structures. **2)** The Meta Grayscale Adaptive (MGA, Sec. 3.2, Fig. 4 (b)) learning makes each image as a sub-learning task to adapt the distribution variation via three key elements: the Multi-windowing Group extracts features from multiple grayscale distributions for robust feature representation with distribution variation; the Meta Perceiver perceive the distribution characteristic to generate personalized parameters (meta kernels) to adapt to the distribution variation; the Meta ResConv fuses the meta knowledge of our IRS segmentation task and the personalized knowledge of each sub-learning task, improving the generalization.

3.1. Grayscale interest search for task-dependent distributions

Our GIS (Fig. 4 (a)) scales the dimensions of window width and center using two cross-correlated coefficients discovering our task-dependent distributions adaptively and expands the searched distribution via windowing (Goldman, 2007) for fine-grained feature representation of renal structures. Therefore, it focuses segmentation networks on the task-dependent grayscale distribution in CTA images to learn the fine-grained representation for fine segmentation.

3.1.1. Problem Formulation

As shown in Fig. 2, the automatic search for the grayscale range of interest has to scale the dimensions of window width w and center c. The width w determines the size of the grayscale range covered by the window, the wider the window, the richer distribution information will be covered, but a wider window will weaken the discrimination between the distributions of the structures thus reducing the contrast of medical images. The center c determines the position of the window, the closer the window center is to the task-dependent distribution, the clearer the details of these structures will appear in images. We define the images in sub-window as $X^{w,c} = \mathcal{W}^{w,c}\left(X_{[g_{min},g_{max}]}\right)$ where the $\mathcal{W}$ is the windowing

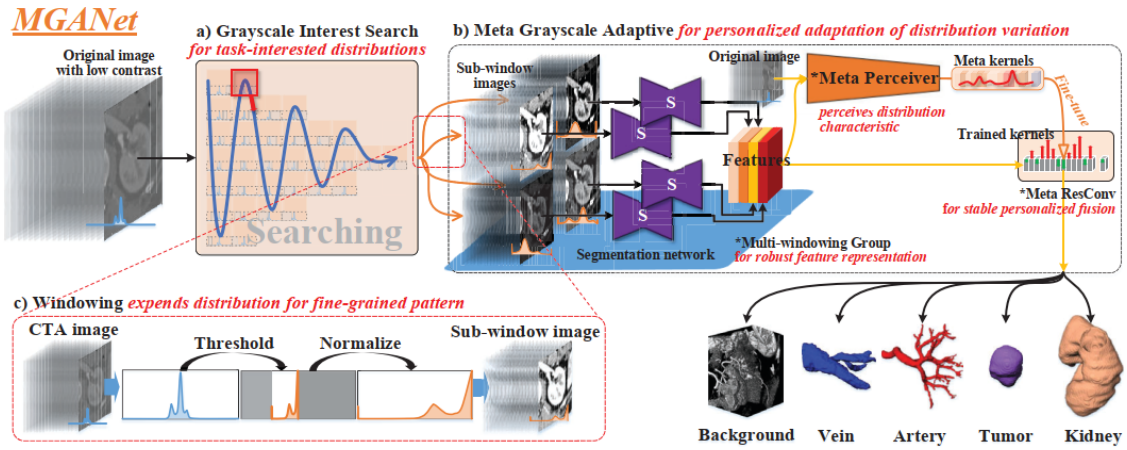

Figure 4: Our MGANet adaptively optimizes the grayscale distribution of medical images via our GIS strategy and MGA learning, achieving the 3D IRS segmentation

process (Sec. 3.1.2), the g_{min} and g_{max} are the minimum and maximum values of the defined original grayscale range. In our IRS segmentation task, the $g_{min} = 0$ and $g_{max} = 2048$. Segmentation networks $\mathcal{N}^{w,c}$ trained on the sub-window images predicts segmentation results $\mathcal{Y}(w,c)$, thus this process is defined as $\mathcal{Y}(w,c) = \mathcal{N}^{w,c}\left(\mathcal{W}^{w,c}\left(X_{[g_{min},g_{max}]}\right)\right)$. The boundaries of the window cannot exceed the original grayscale ranges. We set the Dice coefficient (DSC) (Taha and Hanbury, 2015) of all targeted renal structures as the metric to evaluate these segmentation results on the searched sub-window images, thus the searching process of the task-dependent grayscale distribution is as follow:

$$\begin{aligned} \max_{w,c} \quad & DSC\left(\mathcal{Y}(w,c)\right) \\ s.t. \quad & \mathcal{Y}(w,c) = \mathcal{N}^{w,c}\left(\mathcal{W}^{w,c}\left(X_{[g_{min},g_{max}]}\right)\right) \\ & c - \frac{w}{2} \geq g_{min} \\ & c + \frac{w}{2} \leq g_{max} \end{aligned} \tag{1}$$

We take the grid search method (Bergstra and Bengio, 2012) to optimize these two hyper-parameters, and scale an initial window width $w_0 = \phi$ and center $c_0 = \phi$ in integer multiples using two cross-correlated coefficients, α and β, to reduce the search space:

$$\begin{aligned} \text{width}: \quad & w_\alpha = \alpha\phi \\ \text{center}: \quad & c_\beta = \beta\phi \\ s.t. \quad & \alpha \in \mathbb{Z}^+ \\ & \beta \in \mathbb{Z}^+ \\ & \alpha + 2\beta \leq \frac{2g_{max}}{\phi} \\ & \alpha + 2\beta \geq \frac{2g_{min}}{\phi} \end{aligned} \tag{2}$$

where the α and β are the cross-correlated scaling coefficients which have to match the condition $\frac{2g_{min}}{\phi} \leq \alpha + 2\beta \leq \frac{2g_{max}}{\phi}$ for the width and center. In our experiment, we set $\phi = 256$ to make the grayscale window grid (Fig. 5), so that 32 sub-windows on our grayscale window grid are searched. In particular, we find the best values are $\alpha = 3$ and $\beta = 5$.

3.1.2. Adaptively search the task-dependent distribution range

As shown in Fig. 5, we make a grayscale window grid adaptively via scaling the dimensions of window width and center, and our GIS strategy searches the task-dependent distribution on this grid.

Windowing expends distribution for fine-grained pattern. As shown in Fig. 4(c), a windowing process

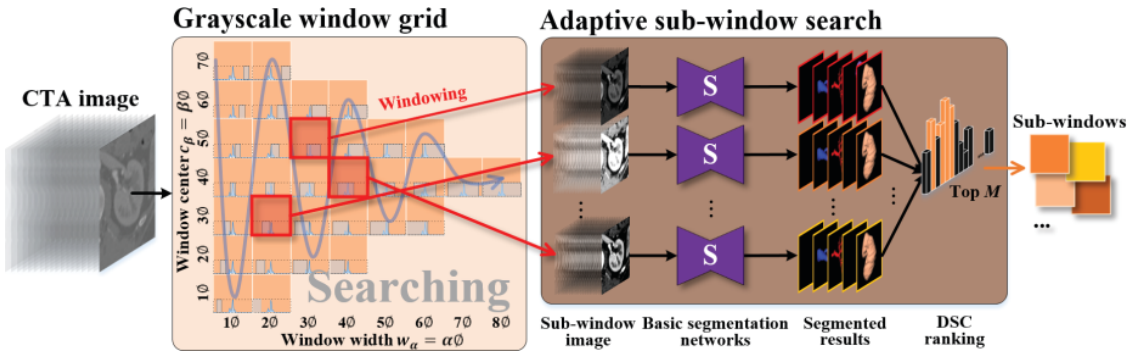

Figure 5: Our GIS strategy for our task-dependent grayscale distribution making the networks fine-grained feature representation

$\mathcal{W}$ removes irrelevant grayscale ranges and expands the distributions of the renal structures thus improving the contrast and bring salient ROIs, so that the network will focus on a wider distribution of the renal structures to perceive their fine-grained patterns. It takes the window width w_α and center c_β from the grayscale window grid firstly, and then thresholds the original image x on the boundaries, $c_\beta - \frac{w_\alpha}{2}$ and $c_\beta + \frac{w_\alpha}{2}$, of the sub-window for a thresholded image x_{thr}. Finally, the thresholded image is normalized for sub-window images $x^{w,c}$ to [0, 1] making more salient renal structures. Therefore, the whole windowing process is as follow:

$$\begin{aligned} x^{w,c} &= \mathcal{W}^{w,c}(x) \\ &= \frac{\max(\min(x, c_\beta - \frac{w_\alpha}{2}), c_\beta + \frac{w_\alpha}{2}) - (c_\beta - \frac{w_\alpha}{2})}{w_\alpha} \end{aligned} \tag{3}$$

Adaptive sub-window search for task-dependent distribution. Our adaptive sub-window search trains multiple segmentation networks iteratively on our grayscale window grid and evaluates via the DSC metric (Taha and Hanbury, 2015) for the best window width w and center c which covers the task-dependent distribution. As shown in Fig. 5, multiple sub-window images on the grayscale window grid train their corresponding segmentation networks iteratively to adapt to the segmentation of renal structures on different sub-windows. Therefore, segmented results on these windows are available to calculate their average DSCs on the four renal structures. The DSC ranking step ranked these average DSCs. For our following Meta Grayscale Adaptive learning, our framework selects top M windows as the sub-windows. Evaluated by our experiments, we set the $M = 4$, considering model performance and computational efficiency. We follow DenseBiasNet (He et al., 2020) as our segmentation networks in our GIS process. The details of DenseBiasNet is introduced in Sec. 3.3.1.

Summary of the advantages. **1)** Our GIS strategy adaptively discovers our IRS segmentation task-dependent distribution, so that it will automatically provide the deep learning model with a more appropriate hyper-parameter of the grayscale distribution range. **2)** Our GIS strategy removes irrelevant grayscale ranges and expands the task-dependent distribution so that the interruption of task-unconcerned distributions' variations will be weakened, and the contrast of the ROIs will be augmented, thus focusing segmentation networks on a wider distribution of the renal structures to perceive fine-grained patterns.

3.2. Meta grayscale adaptive learning for personalized adaptation of distribution variation

Our MGA learning (Fig. 4 (b)) makes an image-level meta-learning to extract diverse robust features from multiple distributions via our Multi-windowing

Group, perceive the distribution characteristics of each image via our Meta Perceiver, and fuse the meta knowledge (trained kernels) and the personalized knowledge (meta kernels) for a personalized fusion process via our Meta ResConv, thus adapting the distribution variation.

3.2.1. Multi-windowing Group for robust feature representation

Our Multi-windowing Group extracts features from multiple distributions covered by different sub-windows, thus representing the images in multi-views making the features more diverse to reduce the overfitting of representation preferences to a single distribution. For a CTA image, it obtains M sub-window images $\{x_0, ..., x_M\}$ with different distributions from the searched M sub-windows via windowing, improving the diversity of image distributions. Independent segmentation networks $\{\mathcal{N}_0, ..., \mathcal{N}_M\}$ trained in the searched sub-windows during our GIS process learn the representation of the distributions covered by these sub-windows, and the corresponding sub-window images are put into these networks for the features before the output convolutional layers. These features are concatenated for the features $\mathcal{F}$ which integrate the representation of multiple distributions:

$$\mathcal{F} = \text{cat}(\mathcal{N}_0(x_0), ..., \mathcal{N}_M(x_M)) \tag{4}$$

where the 'cat' means the concatenation process. Therefore, the features from different distributions will have richer representation capabilities and can be more robustly deal with distribution changes.

3.2.2. Meta perceiver perceives distribution characteristic

Our meta perceiver $\mathcal{P}$ perceives the distribution characteristic of each image to dynamically fuse the features from the trained segmentation networks of the M grayscale distributions covered by the searched sub-windows. The features from our Multi-windowing Group are put into our meta perceiver $\mathcal{P}$ together with the original image x to generate meta kernels $\hat{\mathbf{W}}_{meta}, \hat{b}_{meta}$. The meta kernels will fine-tune the trained kernels initialized by the trained weights in the output layers of the segmentation networks for personalized fusion parameters in our Meta ResConv, thus dynamically fusing the features according to the image's distribution:

$$\hat{\mathbf{W}}_{meta}, \hat{b}_{meta} = \mathcal{P}(\mathcal{F}, x) \tag{5}$$

3.2.3. Meta residual convolution for stable personalized fusion

The personalized fusion parameters from our Meta ResConv has two components: trained kernels $\mathbf{W}_{conv}, b_{conv}$ and meta kernels $\hat{\mathbf{W}}_{meta}, \hat{b}_{meta}$. The trained kernels represent the meta knowledge of our IRS segmentation task and the meta kernels from our Meta Perceiver represent the personalized knowledge of each image. Therefore, the meta kernels dynamically fine-tune the trained kernels via residual (He et al., 2016), generating the personalized fusion parameters and improving the generalization for varied distributions inter-images. As shown in Fig. 6, we use the weights of output convolutional layers in the trained segmentation networks to initialize the trained kernels. The meta kernels $\hat{\mathbf{W}}_{meta}, \hat{b}_{meta}$ from our Meta Perceiver are added to the trained kernels $\mathbf{W}_{conv}, b_{conv}$ making a dynamic meta Conv $\hat{\mathbf{W}}, \hat{b}$ adjust the attention degree of different components in the features $\mathcal{F}$ dynamically. Therefore, the representations of different distributions will be dynamically fused according to the distribution characteristic of each image. The formulation of this process is:

$$\begin{aligned} \hat{\mathbf{W}} &= \hat{\mathbf{W}}_{meta} + \mathbf{W}_{conv} \\ \hat{b} &= \hat{b}_{meta} + b_{conv} \\ \hat{y} &= \sigma(\hat{\mathbf{W}}^T \times \mathcal{F}_n + \hat{b}) \end{aligned} \tag{6}$$

Here, the σ is a softmax activation function for our segmented IRS mask, the $n \in N$ is the position of feature vectors in three-dimensional feature maps. The $\times$ is the matrix multiplication and the $\hat{\mathbf{W}}^T$ is the transpose of the matrix $\hat{\mathbf{W}}$.

Our Meta ResConv is a dynamic fine-tuning of the output convolutional kernels of the segmentation networks which have been trained in our searched sub-windows. When our meta kernels tend to zero $\hat{\mathbf{W}}_{meta} \to 0, \hat{b}_{meta} \to 0$, the Meta Conv degenerates

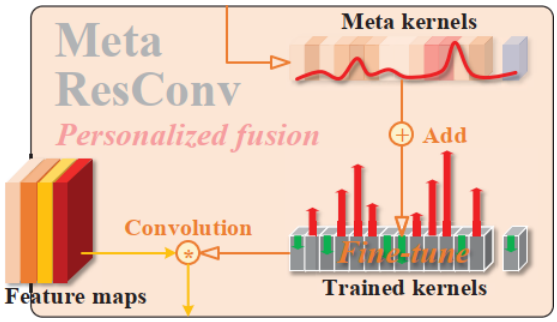

Figure 6: Our Meta ResConv takes the trained kernels as basic parameters and fine-tunes them via our meta kernels from our Meta Perceiver for the adaptation of each image.

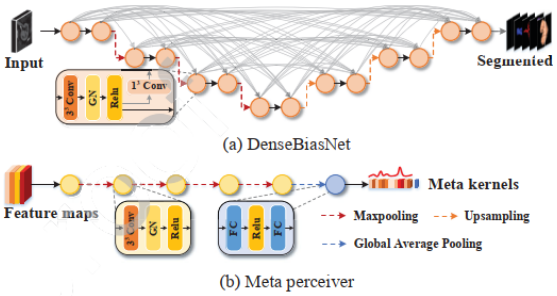

Figure 7: The detailed basic segmentation network structures (DenseBiasNet) for IRS segmentation and meta perceiver for high-order meta kernels.

to the original kernels of the trained output convolution layers $\hat{\mathbf{W}} \approx \mathbf{W}_{conv}, \hat{b} \approx \hat{b}_{conv}$, so our MGA learning will degenerate to an average ensemble strategy (Polikar, 2012). Therefore, our MGA learning is a dynamical enhancement of average ensemble which further considers the personalized optimization of distribution variation for each image, so it will bring the generalization of varied distributions inter-images and stable performance.

Summary of the advantages. **1)** Our Multi-windowing Group integrates feature representation capabilities from multiple distributions, thus extracting more diverse features and reducing the overfitting of representation preferences to a single distribution. **2)** Our Meta Perceiver perceives distribution characteristics of each image and our Meta ResConv generates personalized model parameters to fuse the features from different sub-windows, thus adapting the distribution variation in our 3D IRS segmentation task. **3)** Our Meta ResConv block uses residuals to generate new fusion parameters based on trained kernels, thus greatly reducing the output space of meta kernels, so that the generated fusion parameters will not make large fluctuations and the personalized fusion for distribution variation will be stable.

3.3. Details of the networks

We follow DenseBiasNet structure (He et al., 2020) as our segmentation network for our IRS segmentation and take continuous convolutions, pooling, activations, and full connections as our meta perceiver for the perception of discriminative compositions in features.

3.3.1. DenseBiasNet for IRS segmentation

As shown in Fig. 7 (a), we follow DenseBiasNet (He et al., 2020) as our segmentation network for IRS segmentation, it fuses multi-receptive fields and multi resolution features for the adaptation of scale changes and has achieved success in renal artery segmentation. We take seven resolution stages and each stage has two $3\times3\times3$ convolutional layers followed by a group normalization (Wu and He, 2018)(GN) and a ReLU activation. Max-pooling and up-sampling layers are used to change the resolution. After the final stage, a $1\times1\times1$ convolutional layer followed by a softmax outputs our segmented IRS. The feature maps from each Conv-GN-ReLU layer are compressed via a $1\times1\times1$ convolutional layer and transmitted to every forward layer thus making a full dense connection network. The number of transmitted feature maps in this network is a hyper-parameter k, and in our experiments, it is 4. To train DenseBias-Net, a cross-entropy (ce) loss function $\mathcal{L}_{ce}$ (Taha et al., 2018,?), which has good convergence (Goodfellow et al., 2016) and is widely used in medical image segmentation tasks, is calculated between the IRS labels y and their segmented results $\hat{y}$.

3.3.2. Meta Perceiver perceives distribution characteristic

As shown in Fig. 7 (b), our Meta Perceiver takes continuous convolutions, pooling, activations, and full connections for meta kernels which will give a higher response to the discriminative compositions of the feature vectors in feature maps. It has five resolution stages, each stage has one $3 \times 3 \times 3$ convolutional layer followed by a GN and a ReLU activation. Max-pooling follows each stage for down-sampling of the feature maps. Finally, the feature maps are compressed into a one-dimensional feature vector via a global average pooling layer (Lin et al., 2014) and this feature vector is put into two full connections for our meta kernels which will enhance the task-beneficial features. These meta kernels are trained to optimize the trained kernels initialized by the weights of the trained output convolution layers following our Sec. 3.2 via the $\mathcal{L}_{ce}$.

4. Experiments Configurations

Dataset. Abdominal CTA images from 123 patients who have a tumor in their unilateral kidney were retrospectively selected from the radiology department of Jiangsu Province Hospital in this study. Each image is acquired on a Siemens dual-source 64-slice CT scanner and contrast media was injected during the CT image acquisition. This dataset has five different kidney tumor types including clear renal cell carcinomas, papillary, chromophobe, angiomyolipoma, and eosinophilic adenoma resulting in large heterogeneity and distribution variation. Pixel sizes of these images are between 0.47 mm/pixel and 0.74 mm/pixel, and their slices thicknesses are 0.5 mm/pixel. The cancerous kidney ROIs whose size is $150 \times 150 \times /200$ are pre-extracted automatically via a multi-atlas-based approach following (Yang et al., 2014). Kidney, tumor, veins, and arteries are finely labeled by two radiologists with cross-checking.

Data augmentation. To avoid overfitting and improve the generalization of models, the dataset is augmented via batch-generators[1]. The combination of random rotation ($-20^\circ \leftrightarrow 20^\circ$), cropping ($128 \times 128 \times 128$), mirror (in the axis of x, y and z), scaling ($0.75 \leftrightarrow 1.25$ times) and elastic deformation, enlarges the dataset 1,000 times, thus making 123,000 samples.

Implementation. Our MGANet has two training stages to search the task-dependent grayscale distributions and learn to fuse the features from multiple distributions dynamically. In the first stage, the initial window width w_0 and center c_0 are 256 and 256, thus 32 sub-windows on our grayscale window grid are searched. The optimization setting of the corresponding segmentation networks (DenseBiasNet (He et al., 2020)) trained in these sub-windows is the Adam with the batch size of 1, learning rate of 1×10^{-4} and iterations of 1.6×10^5. The DSC of their results is ranked for the four best sub-windows. In the second stage, the weights of DenseBiasNets on the searched four sub-windows are fixed to extract features, and the kernels of their output convolutional layers are used to initialize the trained kernels during the training of our Meta ResConv. Same as the first stage, The optimization setting of our Meta perceiver is the Adam with the batch size of 1, learning rate of 1×10^{-4}, and iterations of 1.6×10^5.

The framework is implemented with PyTorch[2] and runs on NVIDIA TITAN Xp GPUs. Five-fold cross-validation is performed for comprehensive evaluation. All models in our experiments are optimized for 1.6×10^5 iterations.

Comparison settings. To illustrate the superiority of our framework, we compare the proposed MGANet with V-Net (Milletari et al., 2016), 3D U-Net (Çiçek et al., 2016), Res-U-Net (Li et al., 2018), Kid-Net (Taha et al., 2018), Dense-BiasNet (He et al., 2020), and an ensemble model, VFN (Xia et al., 2018) on CTA images. To further study the nature of our GIS strategy, we analyze the performance of the segmentation network in different sub-windows, and the variation of their ensemble performance with the number

[1] https://github.com/MIC-DKFZ/batchgenerators
[2] https://pytorch.org/

of integrated learners changes. All methods use a same implementation.

Evaluation metric. The segmentation performance of each method is evaluated in two aspects (Taha and Hanbury, 2015) following He et al. (2020). 1) The area-based metric: we take the Dice coefficient (DSC) to evaluate the area-based overlap index. 2) The distance-based metric: we take the average Hausdorff distance (AVD) to evaluate the coincidence of the surface for stable and less sensitive to outliers. To further compare the segmentation quality of outliers, we also take the Hausdorff distance (HD) which is sensitive to outliers. Average DSC (%), AVD (mm), and HD (mm) of all renal structures are also reported together with their standard deviation (±std).

5. Results and Analysis

Our MGANet achieves fine-grained feature representation and bridges the representation preferences caused by the distribution variations thus achieving an excellent 3D integrated renal structures segmentation. In this part, we will thoroughly evaluate and analyze the effectiveness of our proposed MGANet: **1)** The quantitative evaluation and qualitative evaluation in the comparative study (Sec. 5.1) will show the superiorities of our framework compared with other models. **2)** The ablation study (Sec. 5.2) will demonstrate the contribution of our innovations and the influence of different hyper-parameters in our framework. **3)** The GIS process, the number of fused subwindows, and the meta kernels will be analyzed in framework analysis (Sec. 5.3)

5.1. Comparative study shows our superiority

5.1.1. Quantitative evaluation for metric superiority

As demonstrated in Tab. 1, our MGANet achieves excellent performance compared with other methods. In our MGANet, the artery achieves 89.0% DSC, 22.67 mm HD, and 0.60 mm AVD which will strongly support the perfusion regions estimation and the arterial clamping position selection. The kidney gets 95.1% DSC, 12.28 mm HD, and 0.53 mm AVD, the tumor gets 86.4% DSC, 29.85 mm, and 2.76 mm AVD, and the veins get 81.0% DSC, 18.94 mm HD, and 0.90 mm AVD which will provide the operation guidance. The ensemble model, VFN, fuses 2D information in different perspectives, achieving good performance on tumors (86.3% DSC) and fewer outliers on veins (18.14 mm HD). However, it has poor performance on arteries (81.8% DSC) because of the low contrast in original images and the small arteries in 2D slices. The Kid-Net takes deep supervision thus getting competitive performance with smaller average surface distances on tumors (2.42 mm AVD), but it only gets 78.0% DSC on arteries and 75.4% DSC on veins. Dense-BiasNet which is our basic segmentation network has a decent performance in four structures and achieves 84.6% AVG DSC on the original window. When taking our GIS and MGA, our MGANet achieves a significant average improvement of 3.3% DSC in each renal structure.

5.1.2. Qualitative evaluation for visual superiority

As illustrated in Fig. 8, our proposed MGANet has great visual superiority which will provide visual guidance for renal surgery. In case 1, our MGANet achieved fine IRS segmentation results visually, while DenseBiasNet, Kid-Net, and Res-U-Net have serious mis-segmentations (yellow arrows) between the kidney and the tumor. V-Net and VFN have rough segmentation in tumor regions due to the unclear boundaries caused by the narrow grayscale range and the variable grayscale distribution of tumors. In case 2, the enlarged details show fine segmentation quality on the boundaries of the tumor. The comparison methods segment the tumor regions outside the kidney into the kidney (yellow arrows) owing to the unclear boundaries caused by low contrast in the original window and the representation preferences of the varied tumor distributions. In case 3, the enlarged hilum regions illustrate the excellent segmentation quality of our method in complex regions. The kidneys, veins, and arteries structures are crowded in the narrow renal hilum region, which will make the mis-segmentation of these structures. Our MGANet has fine and complete segmentation of veins, while the other comparison methods have serious mis-segmentation of vein regions that have low contrast in such a complex environment.

Table 1: The quantitative evaluation demonstrates the superiority of our MGANet for our 3D IRS segmentation task. Our MGANet achieves better performance than the comparison methods (V-Net, 3D U-Net, Res-U-Net, Kid-Net, VFN, and Dense-BiasNet) on all four renal structures. The EN means that the model is an ensemble method and the AVG DSC means the average DSC on the four structures.

Method	Kidney			Tumor		
	DSC(%)	HD(mm)	AVD(mm)	DSC(%)	HD(mm)	AVD(mm)
V-Net	94.3±1.6	14.77±8.59	0.69±0.26	81.5±19.7	23.81±25.63	3.40±4.35
3D U-Net	94.9±1.6	19.40±11.09	0.79±0.50	82.0±19.4	31.22±27.98	2.78±3.27
Res-U-Net	94.4±1.2	14.52±9.52	0.62±0.17	80.3±14.0	53.87±30.38	5.69±5.71
Kid-Net	94.3±1.5	14.33±8.70	0.66±0.20	82.7±19.1	22.04±25.31	**2.42±3.32**
VFN (EN)	94.5±1.5	15.70±9.17	0.65±0.36	86.3±10.3	**16.34±23.77**	2.45±3.82
DenseBiasNet	94.2±1.7	20.44±12.70	0.76±0.50	82.6±15.1	35.60±32.25	4.23±5.32
Our MGANet (EN)	**95.1±1.5**	**12.28±7.92**	**0.53±0.14**	**86.4±11.8**	29.85±31.17	2.76±3.12

Method	Vein			Artery			AVG DSC(%)
	DSC(%)	HD(mm)	AVD(mm)	DSC(%)	HD(mm)	AVD(mm)	
V-Net	76.4±9.2	21.99±18.66	1.20±0.71	84.3±4.5	53.51±14.86	0.76±0.77	84.1±6.5
3D U-Net	73.7±13.5	30.82±21.06	2.12±2.69	80.2±8.2	50.82±14.49	1.47±1.83	82.7±7.7
Res-U-Net	76.8±7.9	19.04±10.84	1.27±0.93	84.5±8.6	27.60±14.36	1.20±1.89	84.0±6.6
Kid-Net	75.4±10.4	18.25±10.61	1.19±0.96	78.0±7.1	35.15±16.76	0.91±1.02	82.6±7.3
VFN (EN)	76.9±9.0	**18.14±12.45**	1.20±1.40	81.8±6.1	23.33±15.55	0.84±1.16	84.9±6.5
DenseBiasNet	75.3±11.3	23.47±15.67	1.61±2.53	86.1±8.0	25.68±14.77	1.27±2.43	84.6±6.8
Our MGANet (EN)	**81.0±7.2**	18.94±13.86	**0.90±0.70**	**89.0±5.1**	**22.67±13.33**	**0.60±0.70**	**87.9±5.1**

5.2. Ablation study shows improvements of the innovations

As illustrated in Tab. 2, our innovations bring significant enhancements. The DenseBiasNet in the original window achieves 94.2%, 82.6%, 75.3%, and 86.1% DSC on kidneys, tumors, veins, and arteries. When taking our GIS strategy, the best sub-window ($w = 768, c = 1280$) augments the contrast of the CTA image and make the network focus on a wider distribution of the renal structures to perceive fine-grained patterns, so that the veins and arteries achieve 3.9% and 1.9% DSC improvements. The AVG EN-4 in Tab. 2 is an ensemble models which takes an average ensemble strategy (Polikar, 2012) to calculate the average of the segmentation results from the networks trained in the top four sub-windows. When we take the average ensemble strategy (AVG EN-4), it achieves 95.0%, 85.5%, 79.5%, and 88.1% DSC on kidneys, tumors, veins, and arteries. When using our MGA learning to make the personalized fusion of networks' preferences for different distributions, our MGANet gets great improvements, especially on veins, it achieves the significant 1.5% DSC improvement compared with the average ensemble strategy and finally obtains the 87.9% average DSC for our 3D IRS segmentation task.

5.3. Framework analysis

5.3.1. Grayscale interest search analysis

Our GIS strategy achieves multiple sub-windows which will perform better than original CTA images in renal structures. As shown in Fig. 9, we calculate the average DSC of the segmented renal structures is, highlight and select the highest four windows as the sub-windows in our MGANet. **1)** The networks' performance present different characteristics of the renal structures when the sub-windows are changed. The kidney has a more stable performance when the sub-window changes than other structures, because it has a relatively large volume, regular shape, and wide grayscale range making the stable learning of feature representation. The vein, tumor, and artery are sensitive to the window center and width, because their volumes are small and grayscale ranges are narrow making the searched sub-windows difficult to cover in some extreme sub-windows. **2)** Since the renal

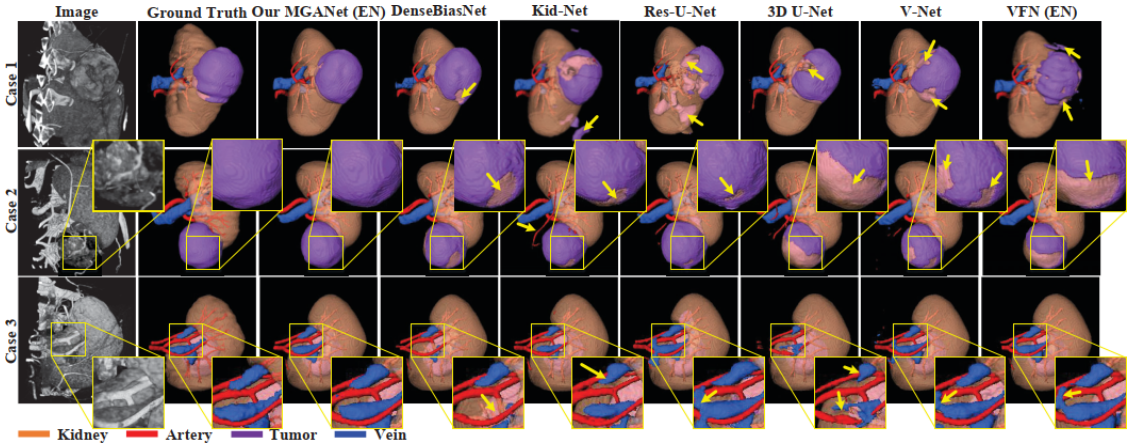

Figure 8: The qualitative evaluation illustrates the visual superiority of our proposed MGANet

Table 2: The ablation study analyses the contributions of our innovations. The AVG EN-4 takes an average ensemble strategy to fuse the networks trained in the top four sub-windows, achieving further improvements in tumors.

GIS	AVG EN-4	MGA-4	DSC(%)±std				
			Kidney	Tumor	Vein	Artery	AVG
			94.2±1.7	82.6±15.1	75.3±11.3	86.1±8.0	84.6±6.8
✓			94.5±2.0	83.6±14.1	79.2±6.7	88.0±5.3	86.3±5.7
✓	✓		95.0±1.4	85.5±13.0	79.5±8.2	88.1±7.0	87.0±5.6
✓		✓	**95.1±1.5**	**86.4±11.8**	**81.0±7.2**	**89.0±5.1**	**87.9±5.1**

structures have different grayscale distributions, they have different interested sub-windows due to their different distribution characteristics. The best sub-window of the veins, kidneys, tumors, and arteries are $(w = 1280, c = 1024)$, $(w = 1536, c = 1024)$, $(w = 1280, c = 1280)$ and $(w = 768, c = 1280)$. Considering the superiority of overall performance, we calculate the average DSC of these structures and select the top four sub-windows, $\{(w = 768, c = 1280)$, $(w = 1280, c = 1024)$, $(w = 512, c = 1280)$, $(w = 1280, c = 1280)\}$, for our MGANet. **3)** The networks trained in the sub-windows which cover the task-dependent distributions will achieve better performance than those trained in the original window. The network trained in the original window achieves 84.6% average DSC which is 1.7% DSC lower than the network trained in the best sub-window.

5.3.2. Fused networks' amount analysis

As shown in Fig. 10, we analyze the DSC change of the models with the amount of the fused networks trained in our GIS increasing. The networks trained during our GIS process are fused start from the highest average DSC. It illustrates the characteristic in two aspects: **1)** The fusion performance increases and then decreases. Due to the severe degradation of the networks accuracy in some singular sub-windows (noisy networks), the average DSC first increases and then decreases with the amount of the fused networks increasing. **2)** Renal structures have different sensitivity to the increasing of the fused networks. Kidney has little fluctuation due to its large volume and wide grayscale range, making its DSC have less fluctuated. Tumors are sensitive and when the models fuse more than 20 networks, the noisy networks will seriously interrupt the whole accuracy and make severe degradation. Considering model performance and com-

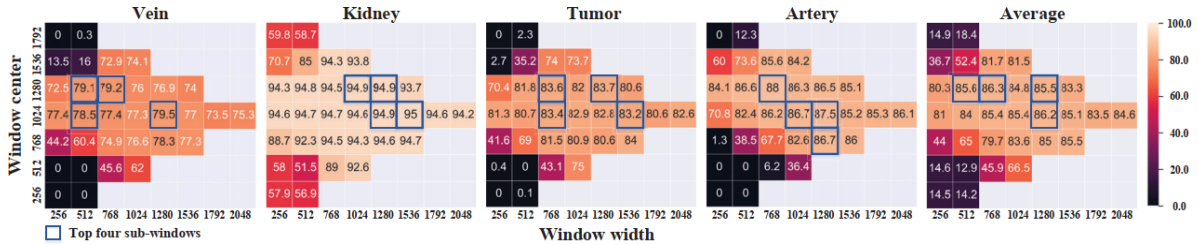

Figure 9: The heatmaps show the DSC of four renal structures and their average performance from our GIS process

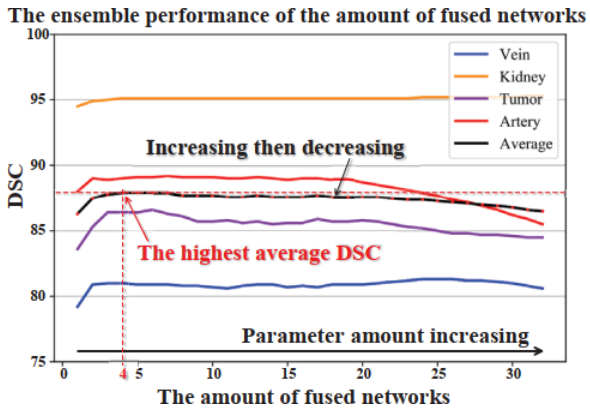

Figure 10: With the amount of the fused networks trained in our GIS process increasing, the DSC will increase and then decrease, and the renal structures will have different sensitivity to the increasing of the networks

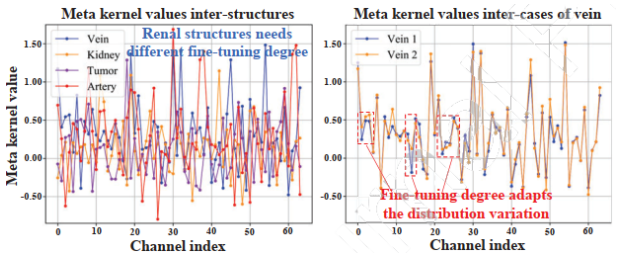

Figure 11: The meta kernels in our MGA learning fine-tune the trained kernels to dynamically adapt the variation of the grayscale distributions.

putational efficiency, we set the fused sub- window amount to 4 in our MGANet which has the highest average DSC and relatively fewer parameters.

5.3.3. Meta kernels analysis

As shown in Fig. 11, the meta kernels in our MGA learning fine-tune the trained kernels to dynamically adapt the variation of the grayscale distributions. **1)** The renal structures have different meta kernel values to fine-tune the trained kernels in different degrees. Because the renal structures have different distributions which present different characteristics, our meta perceiver perceives their characteristics and output the best fine-tuning degrees (meta kernels) for renal structures enhancing the final results. **2)** The cases have different meta kernels to fine-tune the trained kernels for the adaptation of the distribution variation. Owing to our MGA learning's significant improvement on veins, we show the meta kernels of the vein in two cases. As the distribution changes inter-images, our meta perceiver will perceive the variation and the fine-tuning degree of the trained kernels will also dynamically change.

6. Discussion and Conclusion

In this paper, we propose and systematically study the grayscale distribution optimization in the 3D IRS segmentation task, and propose the MGANet for 3D IRS segmentation which will provide great technical support for the treatment of renal cancer for the first time. Our MGANet takes the advantage of our GIS strategy and MGA learning achieving powerful generalization ability. Our GIS strategy (fig. 4 (a)) adaptively searches and expands our task-dependent distributions to help segmentation networks learn fine-grained feature representation for renal structures. Our MGA learning (fig. 4 (b)) takes image-level meta learning strategy which makes each image as

a sub-learning task. It represents diverse robust features from multiple distributions, perceives the distribution characteristic of each sub-learning task, and generates the personalized parameters to fuse these features dynamically according to the sub-learning task's distribution, thus adapting the grayscale distribution variation.

Fine optimization of grayscale distribution significantly improves the performance of medical image segmentation without adding additional model parameters. The experiments of our GIS strategy verified that: **1)** The images in suitable sub-window will obtain better performance than in the original window because the sub-window expands the task-dependent distribution making networks focus on a wider distribution of the renal structures to perceive fine-grained patterns. fig. 9 shows that the best sub-window achieves 1.7% average DSC improvement compared with the original window. **2)** The segmentation objects have different best sub-windows owing to their different distribution characteristics. For example, the vein achieves the best DSC in $(w = 1280, c = 1024)$, while the tumors best sub-window is $(w = 1280, c = 1280)$. As shown in fig. 12, our GIS strategy automatically searches multiple window centers and widths for windowing to augment the contrasts which adapt to our renal structures, thus bringing clear boundaries, salient regions, and fine-grained textures in different sub-windows. Therefore, the model obtains the discriminative representation of different structures with such stronger regional differences, and finally obtains higher accuracy. Compared with the other methods, our framework achieves higher DSC, lower HD, and AVD in Tab. 1 illustrating our significant improvement.

Our 3D IRS segmentation provides visual assistance for preoperative plans for LPN having great clinical significance. **1)** A clinical case (fig. 1 (b)) demonstrates preoperative plans via our 3D IRS segmentation. The 3D IRS model will help doctors estimate the renal perfusion model, locate the tumor resection surface and point the arterial clamping positions visually for preoperative plans. **2)** Intraoperatively, the preoperative plan will be displayed on screen together with laparoscopic videos to guide the LPN (fig. 1 (c)). Therefore, 3D IRS visual model will

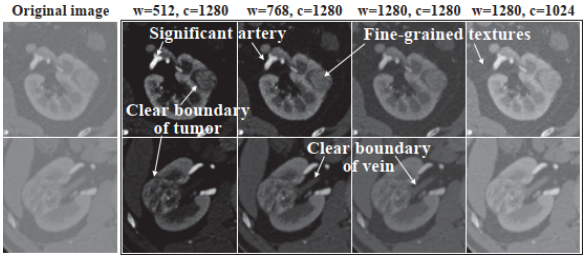

Figure 12: Compared with the original images, our GIS strategy automatically searches multiple window centers and widths for windowing to augment the contrasts which adapt to our renal structures, thus bringing clear boundaries, salient regions, and fine-grained details in different sub-windows

supplement the invisible regions in the laparoscope, so that the difficulties of LPN will be reduced and the quality of LPN will be improved. **3)** Thanks to the high computing speed of deep learning frameworks on GPUs, our MGANet has good time efficiency. It only takes 2.305 seconds on each case which will greatly reduce the time cost for doctors to obtain the renal structures 3D model.

Future works. Hyper-parameter optimization for grayscale distribution in medical image analysis is a research direction worthy of further exploration. There are many potential research topics such as the quick search of this hyper-parameter, the grayscale distribution optimization for other medical image tasks, the application of grayscale distribution optimization in cross-domain learning, etc.

Acknowledgements

This research was supported by the National Natural Science Foundation under grants (31571001, 61828101, 31800825), Southeast University-Nanjing Medical University Cooperative Research Project (2242019K3DN08) and Excellence Project Funds of Southeast University. We thank the Big Data Computing Center of Southeast University for providing the facility support on the numerical calculations in this paper.

References

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research 13, 281–305.

Bianchi, L., Barbaresi, U., Cercenelli, L., Bortolani, B., Gaudiano, C., Chessa, F., Angiolini, A., Lodi, S., Porreca, A., Bianchi, F.M., Casablanca, C., Ercolino, A., Bertaccini, A., Golfieri, R., Marcelli, E., Schiavina, R., 2020. The impact of 3D digital reconstruction on the surgical planning of partial nephrectomy: a case-control study. still time for a novel surgical trend? Clinical Genitourinary Cancer 18, e669–e678.

Brock, A., Lim, T., Ritchie, J.M., Weston, N., 2018. SMASH: one-shot model architecture search through hypernetworks, in: ICLR 2018 : InternationalConference on Learning Representations.

Çiçek, z., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: MICCAI 2016. Springer. Lecture Notes in Computer Science, pp. 424–432.

Dietterich, T.G., 2002. Ensemble learning. MIT Press. chapter The handbook of brain theory and neural networks. pp. 405–408.

Goldman, L.W., 2007. Principles of CT and CT technology. Journal of Nuclear Medicine Technology 35, 115–128.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. The MIT Press.

Ha, D., Dai, A., Le, Q.V., 2017. Hypernetworks, in: ICLR 2017 : InternationalConference on Learning Representations.

Hamada, A., Sawada, A., Kono, J., Koeda, M., Onishi, K., Kobayashi, T., Yamasaki, T., Inoue, T., Noborio, H., Ogawa, O., 2020. The current status and challenges in augmented-reality navigation system for robot-assisted laparoscopic partial nephrectomy, in: International Conference on Human-Computer Interaction. Springer. Lecture Notes in Computer Science, pp. 620–629.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.

He, Y., Yang, G., Chen, A., Kong, Y., Wu, J., Tang, L., Zhu, X., Dillenseger, J.L., Shu, H., Coatrieux, J.L., Shao, P., Zhang, S., Li, S., 2019. DPA-DenseBiasNet: semi-supervised 3D fine renal artery segmentation with dense biased network and deep priori anatomy, in: MICCAI 2019. Springer. Lecture Notes in Computer Science, pp. 139–147.

He, Y., Yang, G., Yang, J., Chen, Y., Kong, Y., Wu, J., Tang, L., Zhu, X., Dillenseger, J.L., Shao, P., Zhang, S., Shu, H., Coatrieux, J.L., Li, S., 2020. Dense biased networks with deep priori anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation. Medical Image Analysis 63, 101722.

Herrell, S.D., Galloway, R.L., Miga, M.I., 2014. Image guidance in robotic-assisted renal surgery, in: Advances in Image-Guided Urologic Surgery. Springer, pp. 221–241.

Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A., 2020. Meta-learning in neural networks: a survey. CoRR abs/2004.05439v2. `arXiv:2004.05439`.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141.

Huang, G., Liu, Z., Maaten, L.V.D., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269.

Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., Maier-Hein, K.H., 2018. nnU-Net: self-adapting framework for U-Net-based medical image segmentation. CoRR abs/1809.10486. `arXiv:1809.10486`.

Jin, C., Shi, F., Xiang, D., Jiang, X., Zhang, B., Wang, X., Zhu, W., Gao, E., Chen, X., 2016. 3D fast automatic segmentation of kidney based on modified AAM and random forest. IEEE Transactions on Medical Imaging 35, 1395–1407.

Li, J., Lo, P., Taha, A., Wu, H., Zhao, T., 2018. Segmentation of renal structures for image-guided surgery, in: MICCAI 2018. Springer. Lecture Notes in Computer Science, pp. 454–462.

Lin, M., Chen, Q., Yan, S., 2014. Network in network, in: ICLR 2014 : International Conference on Learning Representations.

Ljungberg, B., Albiges, L., Abu-Ghanem, Y., Bensalah, K., Dabestani, S., Fernández-Pello, S., Giles, R.H., Hofmann, F., Hora, M., Kuczyk, M.A., Kuusk, T., Lam, T.B., Marconi, L., Merseburger, A.S., Powles, T., Staehler, M., Tahbaz, R., Volpe, A., Bex, A., 2019. European association of urology guidelines on renal cell carcinoma: the 2019 update. European Urology 75, 799–810.

Ma, N., Zhang, X., Huang, J., Sun, J., 2020. WeightNet: revisiting the design space of weight networks. CoRR abs/2007.11823. `arXiv:2007.11823`.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571.

Munkhdalai, T., Yu, H., 2017. Meta Networks, in: Proceedings of the 34th International Conference on Machine Learning, pp. 2554–2563.

Nicolau, S., Soler, L., Mutter, D., Marescaux, J., 2011. Augmented reality in laparoscopic surgical oncology. Surgical Oncology 20, 189–201.

Polikar, R., 2012. Ensemble Machine Learning. Springer US. chapter Ensemble Learning. pp. 1–34.

Porpiglia, F., Fiori, C., Checcucci, E., Amparore, D., Bertolo, R., 2018. Hyperaccuracy three-dimensional reconstruction is able to maximize the efficacy of selective clamping during robot-assisted partial nephrectomy for complex renal masses. European Urology 74, 651–660.

Real, E., Aggarwal, A., Huang, Y., Le, Q.V., 2019. Regularized evolution for image classifier architecture search, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 4780–4789.

Rother, C., Kolmogorov, V., Blake, A., 2004. GrabCut: interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics 23, 309–314.

Shao, P., Qin, C., Yin, C., Meng, X., Ju, X., Li, J., Lv, Q., Zhang, W., Xu, Z., 2011. Laparoscopic partial nephrectomy with segmental renal artery clamping: technique and clinical outcomes. European Urology 59, 849–855.

Shao, P., Tang, L., Li, P., Xu, Y., Qin, C., Cao, Q., Ju, X., Meng, X., Lv, Q., Li, J., Zhang, W., Yin, C., 2012. Precise segmental renal artery clamping under the guidance of dual-source computed tomography angiography during laparoscopic partial nephrectomy. European Urology 62, 1001–1008.

Simone, G., Gill, I.S., Mottrie, A., Kutikov, A., Patard, J.J., Alcaraz, A., Rogers, C.G., 2015. Indications, techniques, outcomes, and limitations for minimally ischemic and off-clamp partial nephrectomy: a systematic review of the literature. European Urology 68, 632–640.

Taha, A., Lo, P., Li, J., Zhao, T., 2018. Kid-Net: convolution networks for kidney vessels segmentation from CT-volumes, in: MICCAI 2018. Springer. Lecture Notes in Computer Science, pp. 463–471.

Taha, A.A., Hanbury, A., 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Medical Imaging 15.

Tan, M., Le, Q., 2019. EfficientNet: rethinking model scaling for convolutional neural networks, in: 36th International Conference on Machine Learning, pp. 6105–6114.

Wang, C., Oda, M., Hayashi, Y., Yoshino, Y., Yamamoto, T., Frangi, A.F., Mori, K., 2020a. Tensorcut: A tensor-based graph-cut blood vessel segmentation method and its application to renal artery segmentation. Medical Image Analysis 60, 101623.

Wang, G., Song, T., Dong, Q., Cui, M., Huang, N., Zhang, S., 2020b. Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. Medical Image Analysis 65, 101787.

Wang, S., Yin, Y., Cao, G., Wei, B., Zheng, Y., Yang, G., 2015. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. Neurocomputing 149, 708–717.

Wu, Y., He, K., 2018. Group normalization, in: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19.

Xia, Y., Xie, L., Liu, F., Zhu, Z., Fishman, E.K., Yuille, A.L., 2018. Bridging the gap between 2D and 3D organ segmentation with volumetric fusion net, in: MICCAI 2018. Springer. Lecture Notes in Computer Science, pp. 445–453.

Yang, B., Bender, G., Le, Q.V., Ngiam, J., 2019. CondConv: conditionally parameterized convolutions for efficient inference, in: NeurIPS 2019, pp. 1307–1318.

Yang, G., Gu, J., Chen, Y., Liu, W., Tang, L., Shu, H., Toumoulin, C., 2014. Automatic kidney segmentation in CT images based on multi-atlas image registration, in: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5538–5541.

Young, J.R., Margolis, D., Sauk, S., Pantuck, A.J., Sayre, J., Raman, S.S., 2013. Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT. Radiology 267, 444–453.

Zhang, S., Yang, G., Tang, L., Lv, Q., Li, J., Xu, Y., Zhu, X., Li, P., Shao, P., Wang, Z., 2019. Application of a functional3-dimensional perfusion model in laparoscopic partial nephrectomy with precise segmental renal artery clamping. Urology 125, 98–103.

Zheng, H., Zhang, Y., Yang, L., Liang, P., Zhao, Z., Wang, C., Chen, D.Z., 2019. A new ensemble learning framework for 3D biomedical image segmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 5909–5916.