IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MULTI ORGAN SEGMENTATION VIA DEEP MULTI PLANAR CO TRAINING

VINISHYA.P.S

ASSISTANT PROFESSOR ,

ARASU ENGINEERING COLLEGE,

CHENNAI MAIN ROAD,

KUMBAKONAM.

SURUTHI.S

UG SCHOLAR,

ARASU ENGINEERING COLLEGE,

CHENNAI MAIN ROAD, KUMBAKONAM.

ABSTRACT

In multi-organ segmentation of abdominal CT scans, most existing fully supervised deep learning algorithms re-quire lots of voxel-wise annotations.. Current mainstream works to address semi-supervised biomedical image segmentation problem are mostly graph-based. In this work, we propose Deep Multi-Planar Co- Training (DMPCT), whose contributions can be divided into two folds:1) The deep model is learned in a co-training style which can mine consensus information from multiple planes like the sagittal, coronal, and axial planes; 2) Multi- planar fusion is applied to generate more reliable pseudo- labels, which alleviates the errors occurring in the pseudo- labels and thus can help to train better segmentation net- works. Experiments are done on our newly collected large dataset with 100 unlabeled cases as well as 210 labeled cases where 16 anatomical structures are manually anno- tated by four radiologists and confirmed by a senior expert. The results suggest that DMPCT significantly outperforms the fully supervised method by more than 4% especially when only a small set of annotations is uses

INTRODUCTION

MULTI ORGAN SEGMENTATION

DL based multi organ segmentation tecniques represent a innovation in daily practice of radiation therapy .Multi-organ segmentation of radiology images is a crit- ical task which is essential to many clinical applications such as computer-aided diagnosis, computer-aided surgery, and radiation therapy.CT image provide accurate anatomical information and electron density for treatment planning but are of soft tissue contract.Fully supervised approaches can usually achieve high ac- curacy with a large labeled training set which consists of pairs of radiology images as well as their corresponding pixelwise label maps. However, it is quite time-consuming and costly to obtain such a large training set especially in the medical imaging domain due to the following reasons:

- precise annotations of radiology images must be hand an- notated by experienced radiologists and carefully checked by additional expert
- contouring organs or tissues in 3D volumes requires tedious manual input. By contrast, large unannotated datasets of CT images are much easier to obtain.

In the biomedical imaging domain, traditional methods for semisupervised learning usually adopt graph based methods with a clustering assumption to segment pixels (voxels) into meaningful regions, *e.g.*, superpixels. These methods were studied for tissue or anatomical structures segmentation in 3D brain MR images, ultrasound images, *etc*.

Thereby our study mainly focuses on multi-organ segmentation in a semi supervised fashion,With the recent advance of deep learning and its applications,fully convolutional networks (FCNs) have been successfully applied to many biomedical segmentation tasks such as neuronal structures segmenta- tion single organ segmentation and multi-organ segmentation in a fully super- vised manner. Their impressive performances have shown that we are now equipped with much more powerful techniques than traditional methods. The current usage of deep learning for semi-supervised multiorgan segmentation in the biomedical imaging domain is to train an FCN on both labeled and unlabeled data, and alternately update automated segmentations (pseudo-labels) for unlabeled data and the network parameters

.

RESULT

Clinical Characteristics

Structure Sets of 120 Patients With hepatocellular carcinoma were used for training,validity,and testing.The median age of the patients was 59 years(range,37-83)and males were dominant (81.7%) (Table 1).Ninety-three patients (77.5%) had liver functions of child-pugh class A and 13.35% had mild-to-moderate degree of ascites.Macroscopic vascular invasion was observed in 88(73.3%)patients.Eight (6.7%)Patients were treatment native and other patients had received various courses of treatment prior to radiotherapy.as a result, various changes such as iodized oils,low density cavity after ablative therapy and volume loss after hepatic resection existed in the liver of most patients.

TABLE 1

Patient Characteristics From:

abdominal multi organ auto segmentation using 3D patch based deep convolutional neural network

Characteristics		No. of patients (n = 120)
Age (years)	Median (range)	59 (37–83)
Sex	Male	98 (81.7%)
	Female	22 (18.3%)
Child-Pugh classification	A	93 (77.5%)
	B	27 (22.5%)
Ascites	No	104 (86.7%)
	Yes	16 (13.3%)
Stage	Early	32 (26.7%)
	Advanced	88 (73.3%)
Vascular invasion	No	32 (26.7%)
	Yes	88 (73.3%)
Previous treatments	No	8 (6.7%)
	Yes	112 (93.3%)
	Total number, range	0–19
	Surgery	0–2
	RFA	0–5
	PEI	0–1
	TACE	0–16

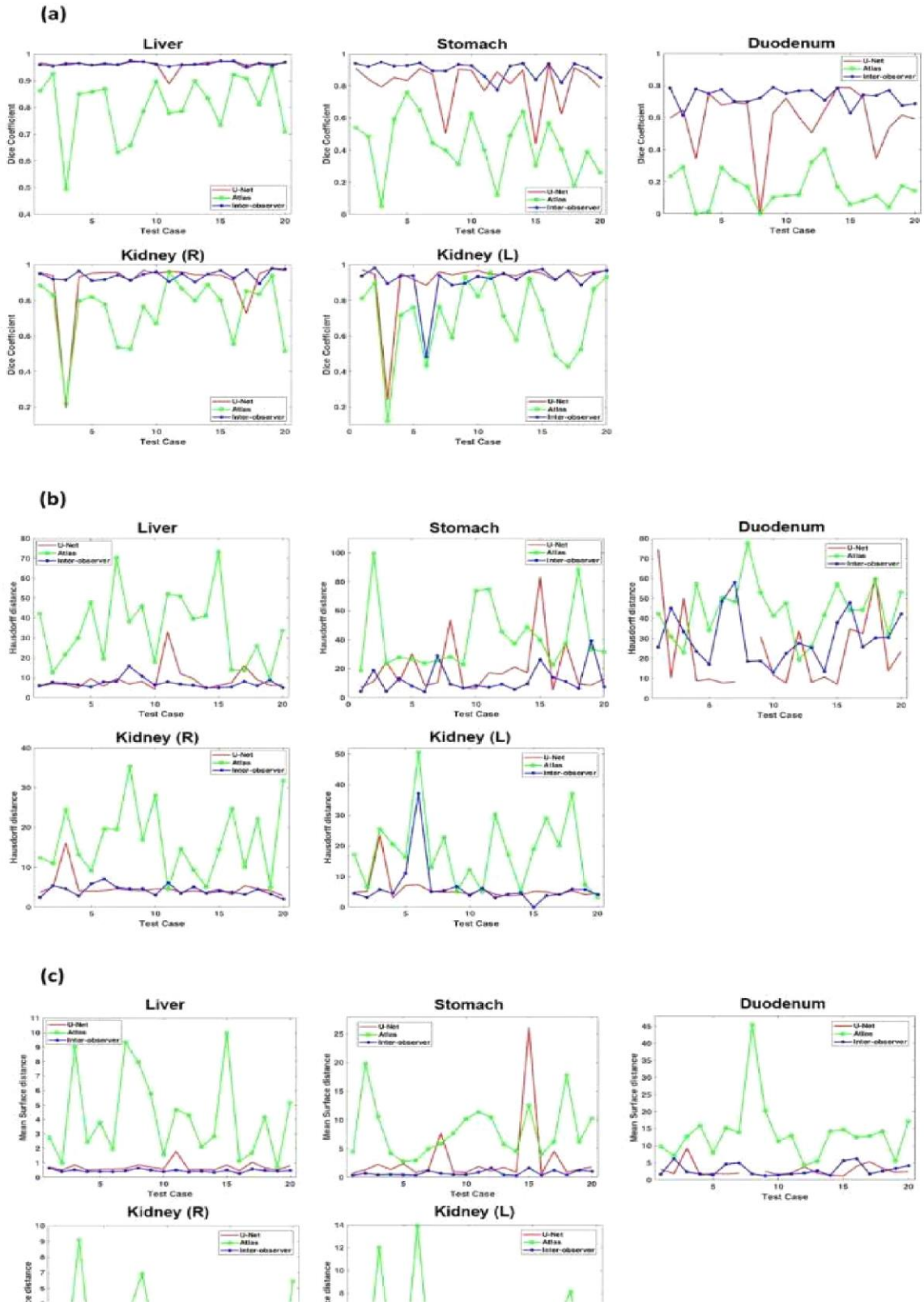
RFA-radio frequency ablation

PEI-Percutaneous ethanol injection

TACE-transarterial chemoembolization

Mean and standard deviation(in paranthesis)of dice similarity coefficient and Hausdorff distance for the five structures From:

Abdominal multiorgan auto segmentation using 3D patch based deep convolutional neural nrtwork



(a) dice similarity coefficient (b) Hausdorff distance (c) mean surface distance of the five structures (liver,stomach,duodenum, and right /left kidney) produced by U-Net-based segmentation the atlas based segmentation ,relative to the previously drawn ground truth manual contour in 20 testing cases

3. Related Work

Fully supervised multi organ segmentation early studies of abdominal organ segmentation focused on atlas-based methods. The frameworks are usually problematic because 1) they are not able to capture the large inter-subject variations of abdominal regions and 2) computational time is tightly dependent on the number of atlases. Recently, learning-based approaches with relatively large dataset have been introduced for multi organ segmentation. Especially, deep Convolutional Neural Networks (CNNs) based methods have achieved a great success in the medical image segmentation in the last few years. Compared with multi-atlas-based approaches, CNNs based methods are generally more efficient and accurate. CNNs based methods for multi-organ segmentation can be divided into two major categories: 3D CNNs based and 2D CNNs based. 3D CNNs usually adopt the sliding-window strategy to avoid the *out of memory* problem, leading to high time complexity. Compared with 3D CNNs, 2D CNNs based algorithms can be directly end-to-end trained using 2D deep networks, which is less time-consuming. Semi-supervised learning. The most commonly used techniques for semi-supervised learning include selftraining co-training multi-view learning and graph-based methods In selftraining, the classifier is iteratively re-trained using the training set augmented by adding the unlabeled data with their own predictions. The procedure repeated until some convergence criteria are satisfied. In such case, one can imagine that a classification mistake can reinforce itself. Self-training has achieved great performances in many computer vision problems and recently has been applied to deep learning based semi-supervised learning in the biomedical imaging domain . Co-training assumes that (1) features can be split into two independent sets and (2) each sub-feature set is sufficient to train a good classifier. During the learning process, each classifier is retrained with the additional training examples given by the other classifier. Co-training utilizes multiple sets of independent features which describe the same data, and therefore tends to yield more accurate and robust results than self-training . Multi-view learning , in general, defines learning paradigms that utilize the agreement among different learners. Co-training is one of the earliest schemes for multi-view learning.Graph-based semi-supervised methods define a graph where the nodes are labeled and unlabeled examples in the dataset, and edges reflect the similarity of examples. These methods have been widely adopted in non-deeplearning based semi-supervised learning algorithms in the biomedical imaging domain . Different from other methods, our work tactfully embeds the multiview property of 3D medical data into the co-training framework, which is simple and effective.

Deep Multi-Planar Co-Training

We propose Deep Multi-Planar Co-Training (DMPCT), a semi-supervised multi-organ segmentation method which exploits multi-planar information to generate pseudo-labels for unlabeled 3D CT volumes. Assume that we are given a 3D CT volume dataset S containing K organs. This includes labeled volumes $SL = \{(\mathbf{I}m, \mathbf{Y}m)\}lm=1$ and unla-

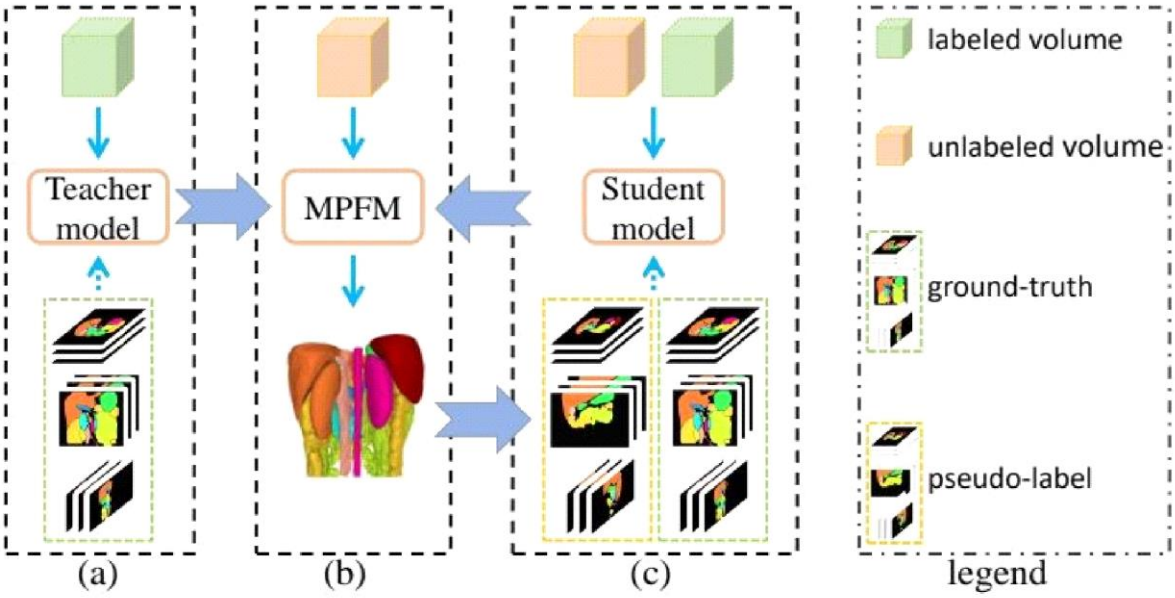


Figure 1. Illustration of the Deep Multi-Planar Co-Training (DMPCT) framework. (a) We first train a teacher model on the labeled dataset.
(b) The trained model is the used to assign pseudo-labels to the unlabeled data using our multi-planar fusion module as demonstrated in
Figure 2. (c) Finally, we train a student model over the union of both the labeled and the unlabeled data. Step (b) and (c) are performed in an iterative manner. beled volumes $SU = \{\mathbf{I}m\}Mm=l+1$, where $\mathbf{I}m$ and $\mathbf{Y}m$ denote a 3D input volume and its corresponding ground-trut and unlabeled volumes, respectively. Typically $l \quad M$. As shown in Figure 1, DMPCT involves the following steps:

- Step 1: train a *teacher model* on the manually labeled data SL in the fully supervised setting
- Step 2: the trained model is then used to assign pseudo-labels $\{\hat{}\mathbf{Y}m\}Mm =l+1$ to the unlabeled data SU by fusing the estimations from all planes
- Step 3: train a *student model* on the union of the manually labeled data and automatically labeled data $SL \cup \{(\mathbf{I}m, \hat{}\mathbf{Y}m)\}Mm=l+1$
- Step 4: perform step 2 & 3 in an iterative manner.

Teacher Model

We train the teacher model on the labeled dataset SL. By splitting each volume and its corresponding label mask from the sagittal (S), coronal (C), and axial (A) planes, we can get three sets of 2D slices, *i.e.*, $SV\,\mathrm{L} = \{(\mathbf{I}Vn\ ,\mathbf{Y}V\ n\ )\}NV\ n=1, V \in \{\mathrm{S, C, A}\}$, where NV is the number of 2D slices obtained from plane V . We train a 2DFCN model (we use as our reference CNN model throughout this paper) to perform segmentation from each plane individually. Without loss of generality, let $\mathbf{I}V \in \mathrm{R}W\times H$ and $\mathbf{Y}V = \{yVi\}W\times H\ i=1$ denote a 2D slice and its corresponding label mask in SVL , where $yV\,i \in \{0, 1, \ldots, K\}$ is the organ label (0 means background) of the i-th pixel in $\mathbf{I}V$. Consider a segmentation model $MV : \hat{}\mathbf{Y} = \mathbf{f}\mathbf{I}V\ ;\ \theta$, where θ denotes the model parameters and $\hat{}\mathbf{Y}$ denotes the prediction for $\mathbf{I}V$.

Our objective function is $L(\mathbf{I}V\ ,\mathbf{Y}V\ ;\ \theta) = \_\ 1\ W \times H\ i=1k=01(yVi = k)\ \log\ pVi,k$ where pVi,k denotes the probability of the i-th pixel been classified as label k on 2D slice $\mathbf{I}V$ and

1(·) is the indicator function. We train the teacher model by optimizing L w.r.t. θ by stochastic gradient descent.

Multi-Planar Fusion Module

Given a well-trained teacher model $\{MV \mid V \in \{S, C, A\}\}$, our goal of the multiplanar fusion module is to generate the pseudo-labels $\{\hat{\mathbf{Y}}m\}Mm=l+1$ for the unlabeled data *SU*.We first make predictions on the 2D slices from each plane and then reconstruct the 3D volume by stacking all slices back together. Several previous studies suggest that combining predictions from multiple views can often improve the accuracy and the robustness of the final decision since complementary information can be exploited from multiple views simultaneously.Thereby, the fused prediction from multiple planes is superior to any estimation of a single plane. The overall module is shown in Figure 2. More specifically, majority voting is applied to fuse the

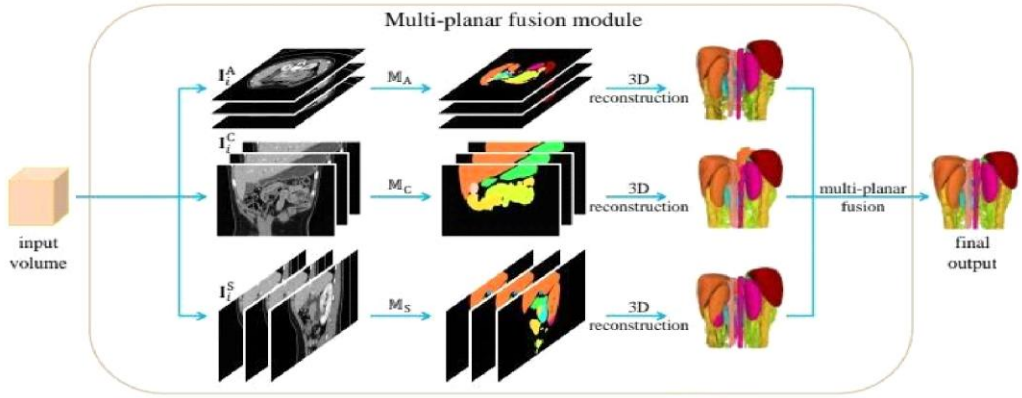


Figure 2. Illustration of the multi-planar fusion module, where the input 3D volume is first parsed into 3 sets of slices along the sagittal, coronal, and axial planes to be evaluated respectively. Then the final 3D estimation is obtained by fusing predictions from each individual plane.

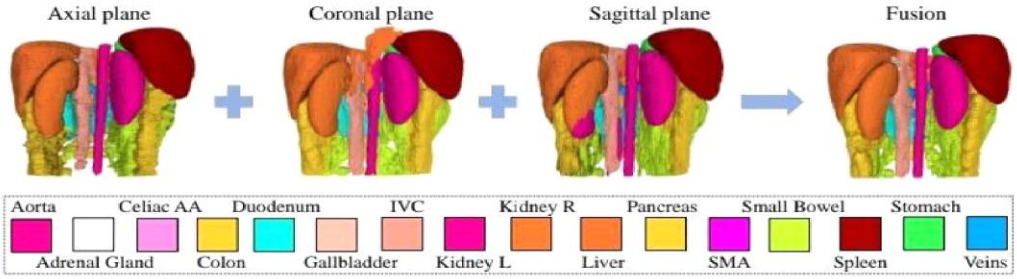


Figure 2. Illustration of the multi-planar fusion module, where the input 3D volume is first parsed into 3 sets of slices along the sagittal,coronal, and axial planes to be evaluated respectively. Then the final 3D estimation is obtained by fusing predictions from each individual plane.

Figure 3. An example of 3D predictions reconstructed from the sagittal,coronal, and axial planes as well as their fusion output. Estimations from single planes are already reasonably well, whereas the single fusion outcome is superior to estimation from any single plane. hard estimations by seeking an agreement among different planes. If the predictions from all planes do not agree on a voxel, then we select the prediction for that voxel with the maximum confidence. As simple as this strategy might sound, this method has been shown to result in highly robust and efficient outcome in various previous studies . The final decision for the *i*-th voxel $y\,i$ of $\hat{\mathbf{Y}}m$ is: $y\,i = yVi$, if $\exists V, V \in \{S, C, A\}$, $V = V \mid yVi = yVi \quad yV\ i$, otherwise, (2)

where V = arg max Pi Vj. Pi s j,Pi,j c and Pi.j A denote $V \in \{S,C,A\}$ the probabilities of the *i*-th pixel classified as label *j* from the sagittal, coronal, and axial planes, respectively. *yVi* denotes the hard estimation for the *i*-th pixel on plane V , *i.e.*,*yVi* = arg max Pi,j V our multi-planar fusion module improves both over- and under-estimation by fusing aspects from different planes and therefore yields a much better outcome. Note that other rules can also be easily adapted to this module. We do not focus on discussing the influence of the fusion module in this paper, although intuitively better fusion module should lead to higher performance.

Student Model

After generating the pseudo-labels $\{\hat{\mathbf{Y}}m\}Mm=l+1$ for the unlabeled dataset SU, the training set can be then enlarged by taking the union of both the labeled and the unlabeled dataset, *i.e.*, $S = SL \cup \{(\mathbf{I}m, \hat{\mathbf{Y}}m)\}Mm = l+1$. The student model is trained on this augmented dataset S the same way we train the teacher model as described . The overall training procedure is summarized . In the training stage, we first train a teacher model in a supervised manner and then use it to generate the pseudo-labels for the unlabeled dataset. Then we alternate the training of the student model and the pseudo-label generation procedures in an iterative manner to optimize the student model T times. In the testing stage, we follow the method .

Computation time

In our experiments, the teacher model training process takes about 4.94 hours on an NVIDIA TITAN Xp GPU card for 80, 000 iterations over all the training cases. The average computation time for generating pseudo-label as well as testing per volume depends on the volume of the target structure, and the average computation time for 16 organs is approximately 4.5 minutes, which is comparable to other recent methods even for single structure inference. The student model training process takes about 9.88 hours for 160, 000 iterations.

Conclusion

In this paper, we designed a systematic framework DMPCT for multi-organ segmentation in abdominal CT scans,which is motivated by the traditional cotraining strategy to incorporate multi-planar information for the unlabeled data during training. The pseudo-labels are iteratively update by inferencing comprehensively on multiple configurations of unlabeled data with a multiplanar fusion module. We evaluate our approach on our own large newly
collected high-quality dataset. The results show that

- our method outperforms the fully supervised learning approach by a large margin
- it outperforms the single planar method, which further

demonstrates the benefit of multi-planar fusion;

- it can learn better if more unlabeled data provided especially when the scale of labeled data is small. Our framework can be practical in assisting radiologists for clinical applications since the annotation of multiple organs in 3D volumes requires massive labor from radiologists. Our framework is not specific to a certain structure, but shows robust results in multiple complex anatomical structures within efficient computational time. We believe that our algorithm may achieve even higher accuracy if a more powerful backbone network or an advanced fusion algorithm is employed, which we leave as the future work.