Biomedical Signal Processing and Control 80 (2023) 104339

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Learning multi-organ segmentation via partial- and mutual-prior from single-organ datasets

Sheng Lian [a,b,c,1], Lei Li [d,1], Zhiming Luo [b,*], Zhun Zhong [e], Beizhan Wang [d], Shaozi Li [b,*]

[a] The College of Computer and Data Science, Fuzhou University, Fujian, China
[b] The Department of Artificial Intelligence, Xiamen University, Fujian, China
[c] Fujian Key Laboratory of Network Computing and Intelligent Information Processing (Fuzhou University), Fujian, China
[d] The Department of Software Engineering, Xiamen University, Fujian, China
[e] University of Trento, Italy

ARTICLE INFO

Keywords:
Anatomical prior
Multi-organ segmentation
Partial supervision

ABSTRACT

Automatic multi-organ segmentation in medical images is crucial for many clinical applications. The art methods have reported promising results but rely on massive annotated data. However, such data is hard to obtain due to the need for considerable expertise. In contrast, obtaining a single-organ dataset is relatively easier, and many well-annotated ones are publicly available. To this end, this work raises the partially supervised problem: can we use these single-organ datasets to learn a multi-organ segmentation model? In this paper, we propose the **P**artial- and **M**utual-**P**rior incorporated framework (**PRIMP**) to learn a robust multi-organ segmentation model by deriving knowledge from single-organ datasets. Unlike existing methods that largely ignore the organs' anatomical prior knowledge, our PRIMP is designed with two key prior shared across different subjects and datasets: (1) partial-prior, each organ has its own character (*e.g.*, size and shape) and (2) mutual-prior, the relative position between different organs follows the comparatively fixed anatomical structure. Specifically, we propose to incorporate partial-prior of each organ by learning from the single-organ statistics, and inject mutual-prior of organs by learning from the multi-organ statistics. By doing so, the model is encouraged to capture organs' anatomical invariance across different subjects and datasets, thus guaranteeing the anatomical reasonableness of the predictions, narrowing down the problem of domain gaps, capturing spatial information among different slices, thereby improving organs' segmentation performance. Experiments on four publicly available datasets (LiTS, Pancreas, KiTS, BTCV) show that our PRIMP can improve the performance on both the multi-organ and single-organ datasets (17.40% and 3.06% above the baseline model on DSC, respectively) and can surpass the comparative approaches.

1. Introduction

Multi-organ segmentation in abdominal CT scans (*e.g.*, liver, pancreas, and kidney) is a critical prerequisite for many clinical applications, such as computer-aided diagnosis (CAD), radiotherapy planning, and computer-assisted surgery (CAS) [1–3]. Recently, deep learning-based methods [4,5] have achieved promising results on this task. However, training the deep models heavily rely on massive data with multiple organs annotated, which are difficult to obtain since the annotating process requires considerable expertise and is extremely time-consuming.

On the other hand, it is relatively easier to obtain a single-organ dataset. Many well-annotated single-organ datasets have been released to the public by different hospitals or research institutes, *e.g.*, KiTS [6] with kidney annotated, and LiTS [7] with liver annotated. Although these single-organ datasets can be collected together to form a diverse training set, existing methods cannot effectively learn a multi-organ segmentation model with it, since each sample only has the annotation of one organ but lacks the others. Therefore, it is easier for us to collect several such kinds of datasets, while existing methods cannot effectively train a multi-organ segmentation model based on these datasets. Then it raises a new task of **learning a robust multi-organ segmentation model from several single-organ datasets**, which we refer to as **partially supervised scenario**. In addition to the inherent difficulties of abdominal CT image segmentation itself, such task is challenging due to factors as below: (1) Each single-organ dataset only contains annotation for specific organ, while leaves the remaining area

* Corresponding authors.
E-mail addresses: zhiming.luo@xmu.edu.cn (Z. Luo), szlig@xmu.edu.cn (S. Li).
[1] Equal contribution.

https://doi.org/10.1016/j.bspc.2022.104339
Received 12 July 2022; Received in revised form 28 September 2022; Accepted 23 October 2022
Available online 9 November 2022

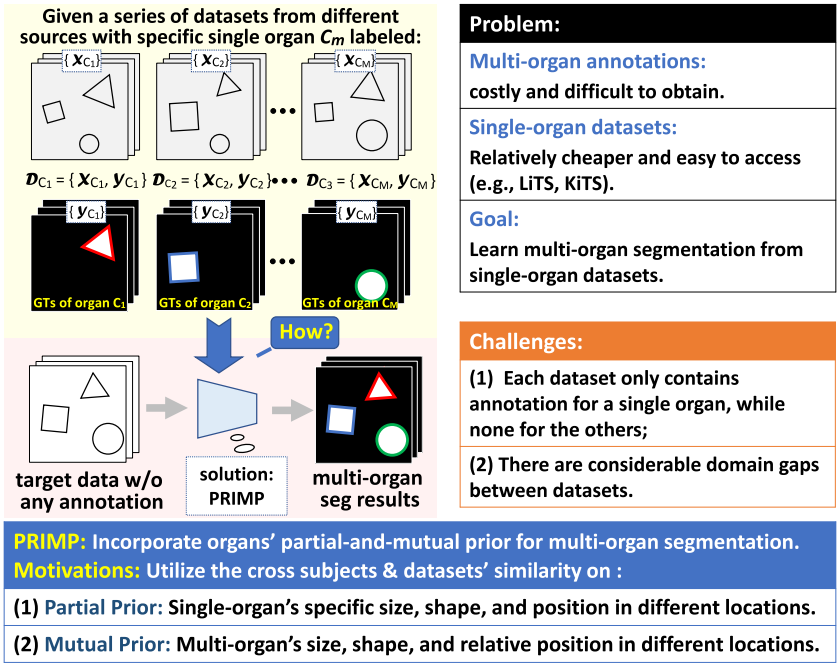

Fig. 1. The overview definition of our task: Given a series of datasets, each of which only has single organ annotations, the goal is to learn a robust multi-organ segmentation model from these datasets. The learned model should also generalize well on a new target dataset without using extra annotations.

(including other organ areas) as background. (2) There are considerable domain gaps between the datasets, including the ones between single-organ datasets and the ones between single/multi-organ datasets. An overview illustration of this task is shown in Fig. 1.

In the literature, very few studies [8–10] attempt to solve such partially supervised problem and train a multi-organ segmentation model with partially labeled data. Dmitriev and Kaufman [8] use a conditional CNN to generate the estimation of each organ by using the class-label as an additional input. Huang et al. [10] adopt a co-training framework to rectify the errors in pseudo-label generated from multiple single-organ segmentation model. Zhou et al. [9] train a model by considering a domain-invariant prior knowledge that different datasets have a similar overall class distribution. Chen et al. [11] design a multi-head network containing a shared encoder and multiple task-specific decoders. Zhang et al. [12] use the task encoding vector to generate dynamic header and produce task-specific segmentation results. Although these methods can learn a multi-organ segmentation model, they largely ignore the anatomical structure of organs, which is critical in learning a robust model. Also, they do not explicitly consider the problem of domain gaps between datasets during learning. On the other aspect, the organs across different subjects and datasets follow the comparatively fixed anatomical structure, regardless of the source of the datasets. For example, (1) each type of organ usually shares the similarity in terms of size, shape and position, and (2) the relative positions between different organs are some kind of consistent. In this study, we regard the first property as the "partial-prior" for single-organ and the second one as "mutual-prior" for multi-organ. These two anatomical priors can provide meaningful guidance for learning single-organ or multi-organ segmentation models, which are under-explored in the existing organ segmentation methods.

Inspired by the two key anatomical priors, we propose a **P**artial-and-**M**utual **P**rior incorporation framework (**PRIMP**) for learning multi-organ segmentation. We formulate the partial-prior as the averaged statistical masks of specific organ in the different axial plane, while formulating the mutual-prior as the averaged statistical masks of multi-organ in the different axial plane. Specifically, we propose to incorporate partial-prior of each organ by learning from the single-organ statistics, and inject mutual-prior of multi-organ by learning from their mutual statistics This partial- and mutual-prior is incorporated in PRIMP's learning on single and multiple organs, respectively, to capture the multi-organ's cross subject/dataset similarity on size, shape, and relative position. The proposed PRIMP framework mainly consists of the following four steps. (1) We train the robust single-organ segmentation model for each organ by injecting its own partial-prior. (2) We generate the multi-class pseudo-label for each dataset by using the previous single-organ model, and then calculate the mutual-prior from the generated pseudo-label over the training set. (3) We learn the multi-organ segmentation model based on pseudo-labels by incorporating mutual-prior. (4) An additional feature alignment module is added to bridge the domain gaps between the source and target datasets. By explicitly considering the proposed partial- and mutual-prior, our PRIMP can capture the organs' similarities across different individuals/datasets, and thus (1) Generating anatomically plausible, robust and accurate predictions; (2) Narrowing down the domain gaps and corresponding learning difficulties; (3) Capturing spatial information while with less computational complex than 3D models, thereby improving organs' segmentation performance.

To sum up, the main contributions of this study are as follows:

- We propose a novel PRIMP framework by considering the anatomical priors, enabling us to learn the robust multi-organ segmentation model from several single-organ datasets.
- By incorporating the partial- and mutual-prior of different organs, the PRIMP is encouraged to capture anatomical invariance across different subjects and datasets, thus guaranteeing the anatomical reasonableness of the predictions, narrowing down the problem of domain gaps, capturing spatial information among different slices, thereby improving the organ segmentation performance.
- Experiments on four publicly available datasets show that our PRIMP can improve the performance on both the multi-organ and single-organ datasets and can surpass the comparative approaches.

2. Related work

In recent years, deep learning-based methods have shone in organ segmentation tasks due to their ability to automatically learn discriminative features. The following review focuses on the multi-organ

segmentation methods in the era of deep learning, mainly classified into methods in fully supervised scenarios (Section 2.1), methods in non-fully supervised scenarios (Section 2.2), and a special form of non-fully supervised learning that is the focus of this paper: methods in partially supervised scenarios (Section 2.3).

2.1. Methods in fully supervised scenarios

Deep learning-based organ segmentation methods can be classified into autoencoder (AE)-based methods [13], CNN-based methods [14], GAN-based methods [15], GCN-based methods [16], etc. Among them, FCNs [17] and their variants represented by U-Net [18] dominate, including 3D-UNet [19], KiU-Net [20], nnUNet [21], etc. Most existing methods focus on specific datasets in which all target organs have pixel-wise annotation. For example, Taghanaki et al. [22] focus on the class-imbalance issue between multiple organs and propose a curriculum learning-based loss function. Sinha and Dolz [23] propose to use a guided self-attention mechanism to capture richer contextual dependencies for different organ segmentation tasks. For the task of abdominal CT multi-organ segmentation, Liang et al. [4] proposed a multi-scale feature fusion network based on 3D attention mechanism, which effectively reduced the difficulty of network convergence and improved the accuracy.

Since the training of these fully supervised methods requires pixel-wise annotation of multiple organs, which requires professional knowledge and considerable manpower, researchers begin to pay attention to methods based on annotation-efficient data, denoted as non-fully supervised scenarios.

2.2. Methods in non-fully supervised scenarios

There are two types of non-fully supervised methods: **Semi-supervised methods** try to learn from a small amount of annotated data. For example, Chaitanya et al. [24] attempt to solve this problem by data augmentation, and proposed to apply spatial deformation field and intensity transformation field through GAN to synthesize new samples from labeled and unlabeled data. Peng et al. [25] applied the idea of co-training to the task of semi-supervised segmentation of organs. They enhanced diversity among different classifiers by generating adversarial samples using labeled and unlabeled data. **Weakly-supervised methods**, on the other hand, solves the problem of costly pixel-level annotation by training with weak labels, including annotations with image-, box-, and scribble-level, etc. For example, Liu et al. [26] explored the possibility to segment lung CT using only scribble annotation. Such method added the mean teacher network with uncertainty perception to the general segmentation framework to encourage the model to be consistent with different disturbances. For segmentation task with only point annotation, He et al. [27] proposed to construct a contrastive learning framework based on the internal similarity and difference between point annotation and unlabeled data, so as to learn the specific visual representation of the target task.

Although semi- and weakly-supervised learning can reduce the workload of labeling in multi-organ segmentation from different perspectives, they are obviously different from the partially supervised scenarios that this paper focus on in the form of data supervision. Therefore, the next subsection will separately introduce the organ segmentation method in partial supervised scenarios.

2.3. Methods in partially supervised scenarios

The task that this paper focus on is to learn a robust multi-organ segmentation model from several single-organ datasets, which is referred to as partial supervised scenarios in literature [9,28]. Studies similar to this task are very limited and can be mainly divided into two categories: conditional information based methods and pseudo label based methods.

The conditional information based methods introduce conditional control information in the training process to establish the relationship between the model parameters and target organ tasks. For example, Dmitriev and Kaufman [8] propose the conditional CNN for learning multi-organ segmentation models. The conditional CNN is a single model trained on several single-organ datasets by conditioning on class labels. Zhang et al. [12] proposed DoDNet, which inherited this idea and also introduced the additional task coding and dynamic parameter mechanism in the U-Net-like segmentation model, and limited the dynamic parameters to the segmentation head. Zhang et al. [29] adopted the current leading framework nnUNet [21] as the backbone model, adding the task encoding as auxiliary information to nnUNet's decoder, and incorporating deep supervision mechanism to further refine the output at different scales. Wu et al. [30] proposed TGNet, which introduced two task-guided attention modules. The designed modules highlight task-relevant features and suppress task-irrelevant information in the feature extraction process. Chen et al. [11] introduced a multi-branch decoder structure to solve the partial labeling problem. The model has a shared encoder and eight decoders, each corresponding to a specific task. During the training process, only the branch corresponding to the task is updated, while the other branches are not involved in the optimization process. The structure is not flexible to expand to new categories.

These conditional information-based approaches restrict the conditional information and variable modules to the local part of the network, and lack the sensitivity to the variation of the feature of different levels. Moreover, such methods can only generate single-organ predictions sequentially, and cannot obtain multi-organ segmentation results simultaneously.

The pseudo label based methods generates pseudo labels of unlabeled organs from partially supervised data, thus converting this task to a fully supervised-liked form. Zhou et al. [9] propose to learn a segmentation model with partially labeled data by using an anatomical prior-aware loss in the learning process. For dealing with the problem of noisy pseudo label, Huang et al. [10] proposed the weight averaged co-training framework to train the multi-organ segmentation model with the generated pseudo-labels. The co-training strategy can rectify the noise in the pseudo-labels for learning a more robust model. On the other hand, Dong et al. [31] explored this problem from the perspective of optimizing pseudo label generation. Such method is inspired by vicinal risk minimization, where the fully labeled vicinal examples are generated by linearly combining randomly sampled partial labels with a weight randomly sampled from a Dirichlet distribution. Moreover, Fang et al. [32] and Shi et al. [28] solved the problem from the perspective of network design and loss function design.

Expect from the above methods, there are also some methods about learning a unified model from multiple natural image datasets. For example, the CDCL method (Cross-Dataset Collaborative Learning) proposed by Wang et al. [33] attempts to build a unified semantic segmentation system for autonomous driving, in which the model is simultaneously learned from multiple traffic datasets. CDCL can also be used to learn a multi-organ segmentation model, and we trained a model by the CDCL in the experiments for comparison.

The proposed PRIMP framework differs from the above methods from the following aspect: (1) PRIMP takes into account the anatomical prior knowledge of organs during the learning process, which is important for the model to generate anatomically sound and accurate predictions. In particular, [9] used prior knowledge that considers only the simplest statistical information and required a small amount of multi-organ labeled data. (2) PRIMP explicitly consider the problem of inter-domain differences between datasets in the learning process.

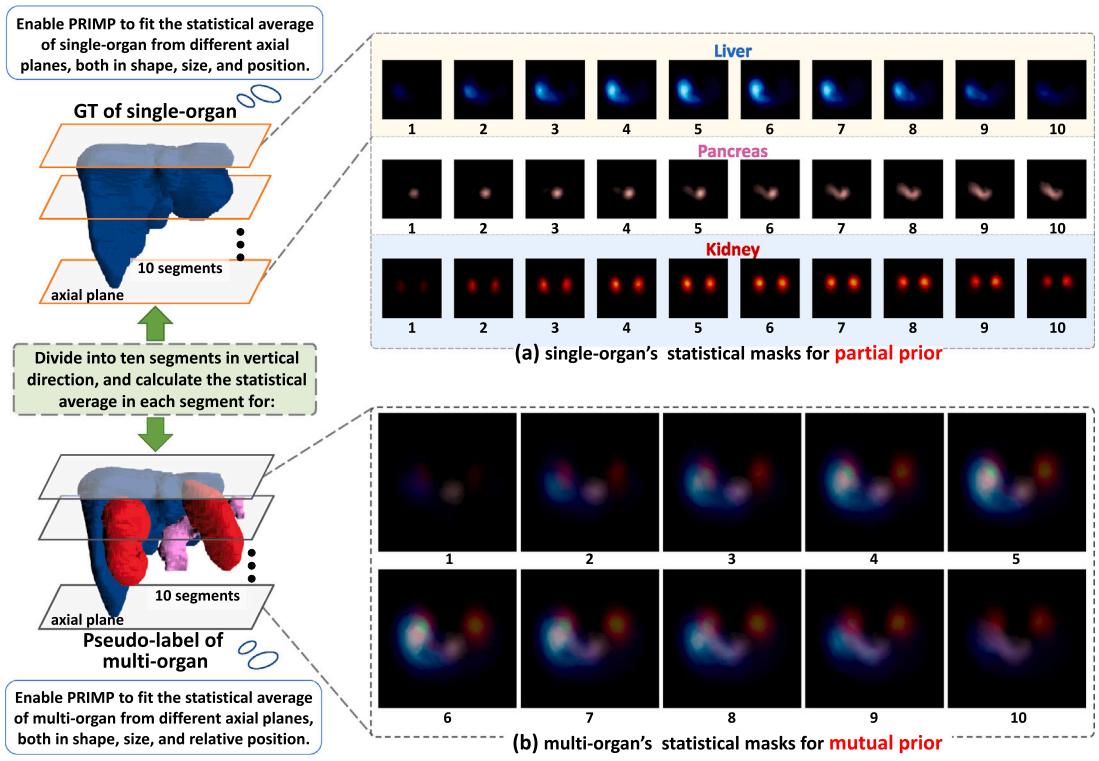

Fig. 2. Schematic diagram on generating statistical masks for both partial- and mutual-priors. The visualization results of the 10-fold single-organ and multi-organ statistical masks are listed in the right side. Here, the different-colored heatmap represent the distribution statistics of different organs, blue for liver, pink for pancreas, and red for kidney.

3. Task and prior definition

3.1. Task definition

The task definition of learning multi-organ segmentation from single-organ datasets focused in this study can be described as follows. **(1)** Suppose that there are several single-organ datasets $\{\mathcal{D}_{c_1}, \ldots, \mathcal{D}_{c_M}\}$ obtained from different sources, and each $\mathcal{D}_{c_m}$ contains annotated data of a specific organ c_m. **(2)** Then we need to learn a segmentation model based on these datasets that can simultaneously segment the multi-organ $\{c_1, \ldots, c_m\}$ (*e.g.*, $\{Liver, Pancreas, Kidney\}$ in this study). **(3)** The learned multi-organ segmentation model should also generalize to a new target dataset $\mathcal{D}_T$ without any additional annotations. The overall process of this task is illustrated in Fig. 1.

3.2. Partial- and mutual-prior

Based on the organs across different subjects and datasets following a comparatively fixed anatomical structure, we introduce the partial- and mutual-priors in our framework as follows: (1) **Partial-prior:** each type of organ usually shares the similarity in terms of size, shape, and position; (2) **Mutual-prior:** the relative positions between different organs are relatively consistent. Specifically, we formulate the partial-prior as the averaged statistical masks of an organ in different axial planes, and formulate the mutual-prior as the joint statistical masks of multiple organs. The computation process of generating statistical masks for partial- and mutual-priors is illustrated in Fig. 2.

We first locate the organs' start-and-end slice in a CT volume and then evenly divide those slices in the middle into K segments along the axial plane. Next, we compute the average label distribution map for each segment over the whole training dataset as the prior mask,

$$q^k = \frac{1}{N}\sum_{i=1}^{N} \bar{l}_i^k, \tag{1}$$

where N is the number of volumes in the training set, $\bar{l}_i^k$ is the mean label map of the kth segment in the volume i.

Notice that, the partial-prior masks $\{q^1, q^2, \ldots, q^K\}$ of organ C_m is computed by using the ground-truth from the single-organ dataset. The mutual-prior $\{q_M^1, q_M^2, \ldots, q_M^K\}$ is based on the pseudo-label on all single-organ datasets, which will be discussed later. We visualize these partial-prior and mutual-prior masks in Fig. 2. After computing the partial-prior and mutual-prior masks, we then leverage them as additional regularization terms to enforce the models' predictions following the closest prior mask's anatomical structure. Through this design, our model is able to obtain the spatial position information of organs to a certain extent.

4. Methodology

To achieve the task described in Section 3.1, the overall framework of our proposed **par**tial-and-**m**utual **p**rior incorporated model (PRIMP) is illustrated in Fig. 3, which mainly consists of the following four steps: **(1)** Learning robust single-organ models with partial-prior (Section 4.1). **(2)** Generating pseudo-label and calculating the mutual-prior (Section 4.2). **(3)** Learning multi-organ segmentation model with mutual-prior (Section 4.3). **(4)** Bridging the domain gap between single-organ datasets and target (Section 4.4).

4.1. Learning single-organ models with partial-prior

Learning single-organ segmentation models is the prerequisite of our PRIMP framework. In this step, we incorporate the partial-prior (PP) discussed in Section 3.2 for training models $\{f_L, f_P, f_K\}$ on $\{D_L, D_P, D_K\}$, respectively. The process of incorporating the proposed partial-prior is illustrated in Fig. 3 (1). For each input slice, we perform a forward step through the model f_{C_m} to obtain a segmentation mask p.

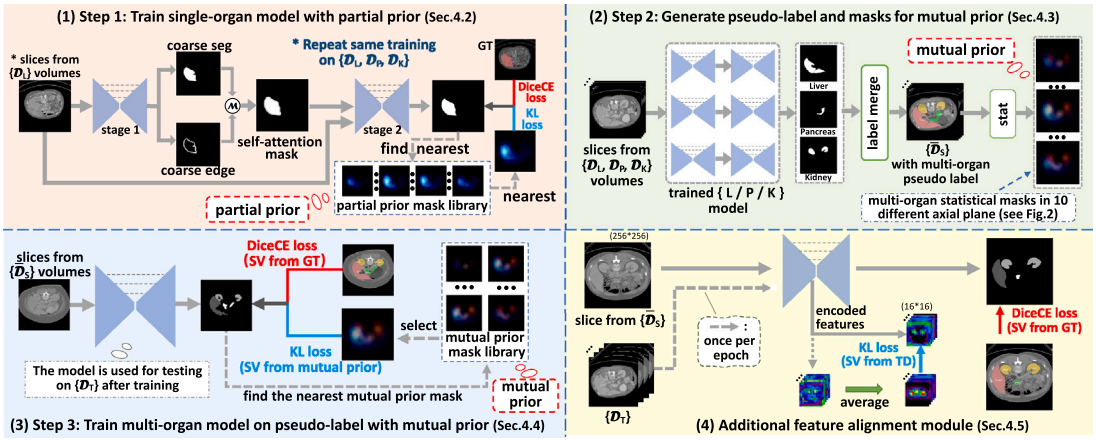

Fig. 3. The overall framework of PRIMP, which contains three steps and an additional feature alignment module. We list the abbreviation used in this figure as follows: *seg* for *segmentation*, *SV* for *supervision*, *SRC* for *source*, and *TD* for *target domain*.

We then select the nearest mask from the partial-prior set $\{q^1, \ldots, q^K\}$ of the organ C_m, by computing the euclidean distance.

$$t = \underset{q \in \{q^1, q^2, \ldots, q^K\}}{\operatorname{argmin}} \sum_{i=1}^{N} (p_i - q_i)^2, \tag{2}$$

where N is the total pixel number, p_i is the predicted softmax probability for the C_m at pixel i.

After the selection, we leverage this partial-prior mask as the auxiliary constraint that prompts the predicted mask has a similar anatomical structure as the closest statistical mask t. We implement the constraint through the KL-divergence, denoted as:

$$\mathcal{L}_{KL}^{partial} = -\sum_{i=1}^{N} t_i \log \frac{p_i}{t_i}, \tag{3}$$

where t is the selected target prior mask, and p_i is the predicted softmax probability.

Due to edge regions for each organ usually are with complex morphology and structure, we further propose to optimize the segmentation model of organ C_m in a cascaded self-attention (SA) manner [34,35] (Fig. 3 (1)), which contains two U-shaped models. In the first stage, we estimate a coarse segmentation mask $\mathcal{M}_{s1\_seg}$ and a coarse edge mask $\mathcal{M}_{s1\_edge}$, and merge them to obtain the self-attention mask. In the implementation, we dilate the $\mathcal{M}_{s1\_seg}$ and $\mathcal{M}_{s1\_edge}$ with a kernel of size $(5, 5)$. The self-attention mask is computed via:

$$\mathcal{M}_{att} = \mathcal{M}'_{s1\_seg} + 0.9 * \mathcal{M}'_{s1\_edge}, \tag{4}$$

where $\mathcal{M}'_{s1\_seg}$ and $\mathcal{M}'_{s1\_edge}$ are dilated masks. Notice that, the GT masks for training the coarse edge masks are obtained by dilating the edge map with a kernel of size $(7, 7)$.

In the second stage, the attention mask $\mathcal{M}_{att}$ is concatenated with the original image to form the input for the second U-shaped model, and thus we can get the enhanced final segmentation results around the edge regions. The DiceCE loss is used for the segmentation model learning of both *stage 1* and *stage 2*, which goes as,

$$\mathcal{L}_{DCE} = \lambda_{Dice}\mathcal{L}_{Dice} + \lambda_{CE}\mathcal{L}_{CE}, \tag{5}$$

where λ_{Dice} and λ_{CE} are weights for dice and cross-entropy loss, respectively, and they are both set to 1. Specifically, $\mathcal{L}_{Dice} = -\sum_{i=1}^{n}(1 - 2 \cdot \frac{t_i p_i}{t_i + p_i})$ and $\mathcal{L}_{\mathrm{CE}} = -\sum_{i=1}^{n} t_i \log(p_i)$, where t_i is the groundtruth (GT) label, and p_i is the predicted softmax probability for the ith class.

To sum up, we denote the DiceCE loss in *stage 1-edge*, *stage 1-seg*, and *stage 2* as $\mathcal{L}_{DCE}^{s1-edge}$, $\mathcal{L}_{DCE}^{s1-seg}$ and $\mathcal{L}_{DCE}^{s2}$, respectively. The overall loss function for learning single-organ model of the organ C_m (Fig. 3 (1)) is

$$\mathcal{L}_{single} = \lambda_s^1 \mathcal{L}_{DCE}^{s1-edge} + \lambda_s^2 \mathcal{L}_{DCE}^{s1-seg} + \lambda_s^3 \mathcal{L}_{DCE}^{s2} + \lambda_s^4 \mathcal{L}_{KL}^{partial}, \tag{6}$$

where $\lambda_s^1, \lambda_s^2, \lambda_s^3$, and λ_s^4 are weights for different components in the single stage. In the same way, we can train the single-organ segmentation model $\{f_L, f_P, f_K\}$ for liver, pancreas, and kidney, respectively.

4.2. Generating pseudo-label and calculating mutual-prior

After learning the single-organ models $\{f_L, f_P, f_K\}$, we adopt them to generate the multi-organ pseudo-label for the single-organ datasets $\{D_L, D_P, D_K\}$. Specially, we use the $\{f_L, f_P\}$ to estimate pseudo-label of the liver and pancreas of D_K, and combine them with kidney GT label. Then, we will have the $\overline{D}_K$ with multi-organ labels, and can get the $\overline{D}_L$ and $\overline{D}_P$ in the same way. We denote $\{\overline{D}_S\} = \{\overline{D}_L, \overline{D}_P, \overline{D}_K\}$.

We further generate statistical masks of multi-organ on different axial plane for incorporating mutual-prior, which is illustrated in Section 3.2 and Fig. 2 (b). Notice that, after pseudo-labels generation, we first locate the start and end slice for each volume in $\{\overline{D}_S\}$, based on the pseudo-label. Then we equally divide those slices in the middle to 10 folds and compute the mutual-prior masks across the $\{\overline{D}_S\}$, denoted as $\{q_M^1, q_M^2, \ldots, q_M^K\}$. This locating step will assign the distribution of multiple organs from different volumes in each fold roughly the same. Subsequently, these masks will be used as mutual-priors in Sections 4.3 and 4.4 to promote the learning of the multi-organ segmentation models.

4.3. Learning multi-organ model with mutual-prior

To learn a robust multi-organ segmentation model as shown in Fig. 3 (3), we incorporate the multi-organs' mutual-prior (MP) (Fig. 2 (b)) calculated in Section 4.2 into the model training. Similar to the process of using the partial-prior in Section 4.1, we first compute the multi-class segmentation map for an input slice and then select the nearest mutual-prior mask from $\{q_M^1, q_M^2, \ldots, q_M^K\}$ based on the euclidean distance. Next, we apply the KL-divergence to enforce the segmentation map following the selected mutual-prior mask's anatomical structure.

In this multi-organ segmentation learning step, we denote the DiceCE loss for basic segmentation (same as Eq. (5)) as $\mathcal{L}_{DCE}^{multi}$, and the KL loss for mutual-prior constraint (same as Eq. (3)) as $\mathcal{L}_{KL}^{mutual}$. The loss function for this step is:

$$\mathcal{L}_{multi} = \lambda_m^1 \mathcal{L}_{DCE}^{multi} + \lambda_m^2 \mathcal{L}_{KL}^{mutual}. \tag{7}$$

In this manner, our method will encourage the current prediction close to the nearest multi-organ distribution prior, thus ensuring the accuracy and anatomical rationality of the predictions.

Table 1
The statistics of datasets adopted in our study.

Datasets	Labeled organs	#Volumes		
		Training	Testing	Total
LiTS ($\mathcal{D}_L$)	Liver	104	26	130
Panc ($\mathcal{D}_P$)	Pancreas	225	57	282
KiTS ($\mathcal{D}_K$)	Kidney	168	42	210
BTCV ($\mathcal{D}_T$)	{L, P, K}[a]	6[b]	24	30

[a]{L, P, K} are short for {Liver, Pancreas, Kidney}, other labeled organs or tissues in this dataset are treated as background.
[b]Six volumes are randomly selected for domain alignment (without annotation), while the rest 24 volumes for evaluation.

4.4. Bridging the domain gap between single-organ datasets and target

Although all the datasets are abdominal CT scans and share similarities in anatomical, different datasets still suffer the domain gap issue since they are obtained from various places by different CT scanners. To bridge this domain gap, we introduce an additional domain alignment (DA) module to align the overall feature distribution between the $\{\overline{\mathcal{D}}_S\}$ and $\{\mathcal{D}_T\}$. The computation details of this module are as follows. Before optimizing each epoch, we first calculate the average final encoded-feature of all samples in the target datasets $\{\mathcal{D}_T\}$ and regard it as the reference of the target, denoted as $\mathcal{F}_{tgt}$. Then during the optimization process, we use a KL regularization term to minimize the divergence between the final encoded-feature $\mathcal{F}^i_{src}$ of each sample in $\{\overline{\mathcal{D}}_S\}$ and the $\mathcal{F}_{tgt}$. This scheme is illustrated in Fig. 3 (4). Consequently, we can mitigate the domain gap between the source and target.

In summary, after adding this additional domain alignment module, the overall loss function for the multi-organ segmentation model becomes to:

$$\mathcal{L}'_{multi} = \lambda^1_m \mathcal{L}^{multi}_{DCE} + \lambda^2_m \mathcal{L}^{mutual}_{KL} + \lambda^3_m \mathcal{L}^{align}_{KL}. \tag{8}$$

5. Experiments

5.1. Datasets and evaluation metrics

Datasets: We utilize two groups of datasets for training and testing PRIMP, including *the partially labeled single-organ datasets* and *the target multi-organ dataset*. We summarize the statistical details of the datasets in Table 1.

5.1.1. The partially labeled single-organ datasets

The source datasets consist of the LiTS ($\mathcal{D}_L$) [7], Pancreas ($\mathcal{D}_P$) [36] and KiTS ($\mathcal{D}_K$) [37]. In detail, we only use the training sets of these three datasets for training our model. The rest testing sets are used for evaluation. Note that, the tumor areas related to an organ are treated as corresponding organ areas in this study.

5.1.2. The target multi-organ dataset

The target dataset used in this study is the BTCV ($\mathcal{D}_T$) [38] dataset. We adopt this dataset to evaluate the generalization ability of our multi-organ segmentation model, which is trained from the previous three single-organ datasets.

Evaluation Metrics: We employ the Dice Similarity Coefficient (DSC) as the evaluation metrics to measure the similarity between the predictions and the ground-truth segmentation masks. In detail, DSC is a statistic for gauging the similarity of two samples, which defined as: $\mathrm{DSC}(\mathcal{P}, \mathcal{G}) = \frac{2\times|\mathcal{P}\cap\mathcal{G}|}{|\mathcal{P}|+|\mathcal{G}|}$, where $\mathcal{P}$ is the binary prediction and $\mathcal{G}$ is the ground-truth.

5.2. Implementation details

We implement the proposed model with the *PyTorch* library [39] in a device with an NVIDIA 2080TI GPU. We use the soft tissue CT window with the HU in a range of [−512, 512]. The input image is resized to 512 × 512 and then randomly cropped to 256 × 256 for training. The random flip is used as data augmentation. The network is a U-shaped structure with a resnext50_32x4d [40] as encoder, and is implemented by the segmentation-models-pytorch *(SMP)* toolbox.[2] We train the network with a maximum training epoch of 12, and the batch size is set to 8. The Adam optimizer [41] is used for optimization with an initial learning rate of 1e-4. The learning rate is decreased by 0.2 every four epochs. The cross-entropy loss and dice loss used in this study are the same as nn-UNet [21]. The hyper-parameters of the single-organ model learning are $\lambda^1_s = 2$, $\lambda^2_s = 2$, $\lambda^3_s = 5$ and $\lambda^4_s = 5$. For the multi-organ learning, the hyper-parameters are $\lambda^1_m = 5$, $\lambda^2_m = 5$ and $\lambda^3_m = 0.01$.

5.3. Ablation study

5.3.1. Components analysis in single-organ learning

We conduct evaluation for the components introduced in the single-organ segmentation stage (Fig. 3 (1)), including: (a) Adopting the vanilla cascaded U-shaped model; without self-attention (SA) scheme and single-organ statistical prior (partial-prior, PP); (b) Applying the self-attention (SA) scheme; (c) Employing the single-organ statistical masks as partial-prior (PP); (d) Using both SA and the PP scheme. We summarize the results of ablation study on single-organ models in Table 2.

The left part of Table 2 shows the results on the testing set of the source datasets ($\{\mathcal{D}_L, \mathcal{D}_P, \mathcal{D}_K\}$), and the right part is the direct testing results on the target dataset $\{\mathcal{D}_T\}$. By analyzing different components, we have the following observations.

(1) The proposed SA and PP are beneficial for the source datasets. From the left part of Table 2, we find that: *First*, incorporating SA can slightly increase the average DSC from 87.09% to 87.48%, where $\mathcal{D}_L$ witnesses a notable rise of 1.47%. *Second*, adding PP can significantly boost the average DSC from 87.09% to 89.14%. *Third*, when jointly incorporating both SA and PP, our model achieves a further improvement, with the average DSC rising from 87.09% to 90.15%. Specifically, leveraging PP to $\{\mathcal{D}_P\}$ can bring a remarkable increase of the pancreas' DSC (from 75.23% to 79.57%). When further adding SA, the overall promotion is more remarkable (from 75.23% to 81.56%). These factors demonstrate that the proposed SA and PP are mutually beneficial for the source datasets, especially for datasets with smaller tissues and more unbalanced data, such as pancreas. It is worth noting that after adding SA, the DSC on $\mathcal{D}_K$ slightly decreased from 94.80% to 94.48%. The reason may come from over-fitting and false positives caused by the special structure of kidney.

(2) Directly applying single-organ models to the target dataset would lead to significant performance drop. There are innegligible domain gap between source and target datasets. We can observe a striking performance degradation for all source single-organ models when directly applied to the target dataset. For example, when using vanilla U-Net, the average DSC drop from 87.09% to 67.00% when applied to the target.

(3) The merits brought by SA and PP can be generalized directly to the target domain. After incorporating SA and PP in training, the performance will achieve consistent improvements for the three organs in the target dataset. Especially, the average DSC will rise from 67.00% to 71.58% when incorporating both SA and PP. For the pancreas with a smaller size and more unbalanced data, we have a remarkable boost similar to the one in the source datasets (from 36.19% to 48.17%).

[2] https://smp.readthedocs.io/.

Table 2
The results of ablation study on single organ model.

	SA	PP	DSC (%) on $\{D_L, D_P, D_K\}$				DSC (%) on $\{D_T\}$			
			Liv.	Pan.	Kid.	Avg.	Liv.	Pan.	Kid.	Avg.
Single organ model			91.22	75.23	94.80	87.09	84.74	36.19	80.07	67.00
	✓		92.69	75.26	94.48	87.48	**88.70**	42.74	77.30	69.58
		✓	92.74	79.57	95.12	89.14	87.41	44.94	**81.41**	71.26
	✓	✓	**93.64**	**81.56**	**95.27**	**90.15**	86.14	**48.17**	80.42	**71.58**

SA refers to cascaded self-attention scheme. *PP* refers to partial-prior. *liv., pan., kid., and avg.* are short for liver, pancreas, kidney, and average, respectively.

Table 3
The results of ablation study on multi-organ model. *w/ PL* refers to training with pseudo label. *MP* and *DA* refer to multi-organ statistical masks as mutual-prior, and source-target feature alignment, respectively.

Organ model	w/ PL	MP	DA	DSC(%) on D_T			
				Liv.	Pan.	Kid.	Avg.
Single				86.14	48.17	80.42	71.58
Multi	✓			92.08	71.43	83.52	82.34
	✓	✓		92.49	72.80	83.89	83.06
	✓		✓	92.61	72.06	84.43	83.03
	✓	✓	✓	**93.73**	**74.46**	**85.00**	**84.40**

Liv., pan., kid., avg. are short for liver, pancreas, kidney, average, respectively.

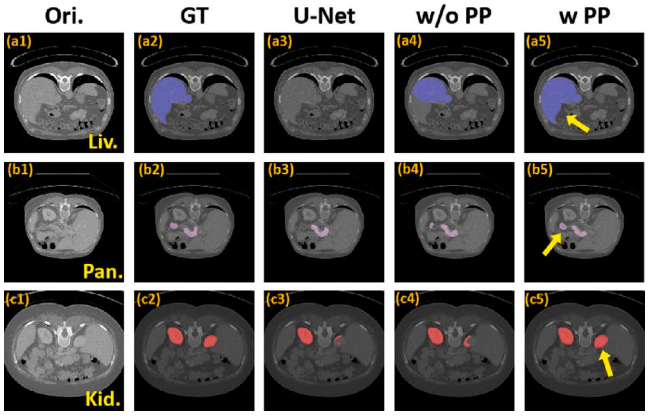

Fig. 4. Single-organ prediction results comparisons on U-Net, ours without partial-prior, and ours with partial-prior. *PP* is short for partial-prior. blue for liver, pink for pancreas, and red for kidney.

These results indicate that merits brought by SA and PP in source datasets also can generalize to the target dataset without additional operations.

Qualitative comparison of different single-organ models. The visualized results on different organs of $\{D_L, D_P, D_K\}$ are displayed in Fig. 4, where segmentation results on *vanilla U-Net*, *ours without partial-prior*, and *ours with partial-prior* are listed. From the third column, we observe that training with vanilla U-Net suffers the issue of identifying liver (a3) and kidney (c3). While in the fourth column, with only the SA module involved, our model can better locate the three organs (a4–c4), however still fails to maintain organs' anatomical morphology. When further incorporating PP, as indicated in (a5, b5, c5) with yellow arrows, our model can better locate organ regions and maintain the integrity and anatomical reasonableness of organs, producing more accurate segmentation results.

5.3.2. Components analysis in the multi-organ learning

We further conduct experiments to evaluate the effect of the components introduced in the multi-organ segmentation learning stage (Fig. 3 (3) and (4)) based on the generated pseudo labels, including: (a) Directly testing single organ models on $\{D_T\}$ with SA and PP incorporated (same as the last line of Table 2's right part). (b) Adopting the vanilla U-shaped model based on $\{\overline{D}_S\}$. (c) Utilizing multi-organ's statistical masks as mutual-prior (MP) (Section 4.3). (d) Adding domain alignment (DA) module between $\{D_L, D_P, D_K\}$ and $\{D_T\}$ to narrow down the domain gaps. (e) Applying both MP and DA module. We report the ablation studies on multi-organ models in Table 3, and have the following conclusions.

(1) Learning with pseudo-labels can boost the model performance. From the first two lines in Table 3, we can find that learning the multi-organ model with pseudo-labels (generated as Fig. 3 (2)) can significantly boost the performance of multi-organ segmentation. For example, the average DSC rises from 71.58% to 82.34%, and DSC of pancreas in $\{D_T\}$ witnesses a remarkable boost of 23.26%.

(2) The proposed MP and DA are beneficial for segmentation on $\{D_T\}$. Compared with the baseline model (line 2 of Table 3), incorporating either the proposed MP and DA can bring consistent increases in DSC for different organs. Specifically, involving MP brings an improvement of 0.72% on average DSC, while applying DA between $\{\overline{D}_S\}$ and $\{D_T\}$ achieves the rise of 0.69% in average DSC. Simultaneously adopting MP and DA boosts the average DSC from 82.34% to 84.40%, where DSC improvement for liver, pancreas, and kidney are 1.65%, 3.03%, and 1.48%, respectively. These factors demonstrate the effectiveness and complementary of the proposed MP and DA.

(3) The choice for K segment when generating prior. We also evaluate different choices of K as introduced in Section 3.2. Specifically, K is the number of segments when generating prior masks. The comparing results summarized in Fig. 5. For pancreas and liver, when $K = 10$, our model achieves best DSC. However, since liver is larger, the mutual-prior generated with $K = 20$ yields slightly higher results. For the average DSC, our model achieves best DSC when $K = 10$. For the sake of uniformity, we choose $K = 10$ in this study.

Qualitative results on multi-organ models. In Fig. 6, we visualize the multi-organ segmentation results of different models. From the third column, we can find that only learning with pseudo labels still suffers the main issue of identifying the pancreas as in (a2), (b2), and also produce some anatomically impossible false-positive prediction of the liver (c2). After incorporating the MP or DA, the segmentation of pancreas and liver will be improved, while still have false-positive (a3, a4, b3, b4, c3, c4). In the last column, by simultaneously adopting MP and DA, the anatomically irrational false positive are suppressed, with the organ-specific morphology better maintained.

5.4. Comparisons with state-of-the-arts

We compare the proposed PRIMP with a series of state-of-the-art models, and report the comparison results on source datasets of $\{D_L, D_P, D_K\}$ and target dataset $\{D_T\}$ in Table 4. The comparison methods include: (1) U-Net$_{single}$: Several vanilla U-Nets trained on $\{D_L, D_P, D_K\}$. (2) PaNN: The prior-aware network proposed by Zhou et al. [9], which uses the category statistic distribution of different organs as prior. (3) Co-training: The co-training weight-averaged model proposed by Huang et al. [10] that leverages a co-training framework to rectify the pseudo label during the training. (4) CDCL: The cross-dataset collaborative learning method proposed by Wang et al. [33] for Semantic Segmentation in Autonomous Driving. (5) Med3D: a multi-head network proposed by Chen et al. [11], which contains a shared

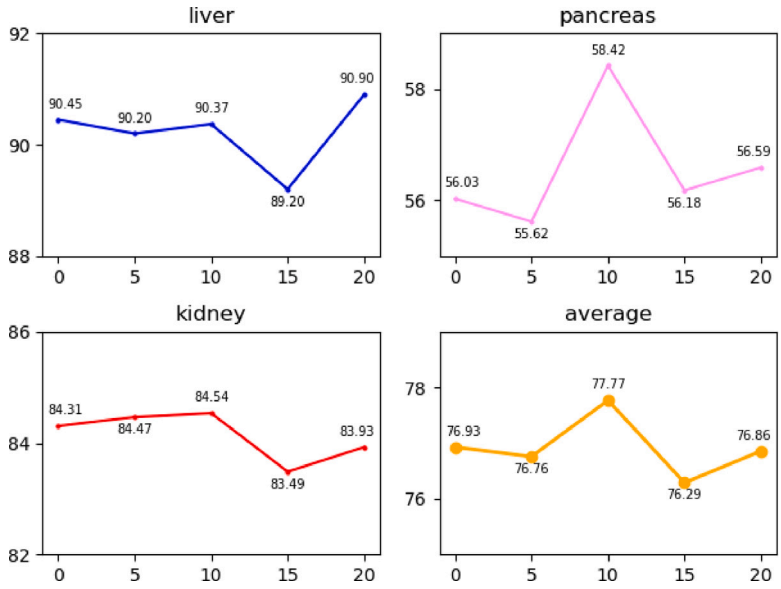

Fig. 5. The DSC results when selecting different K for MP as defined in Section 3.2.

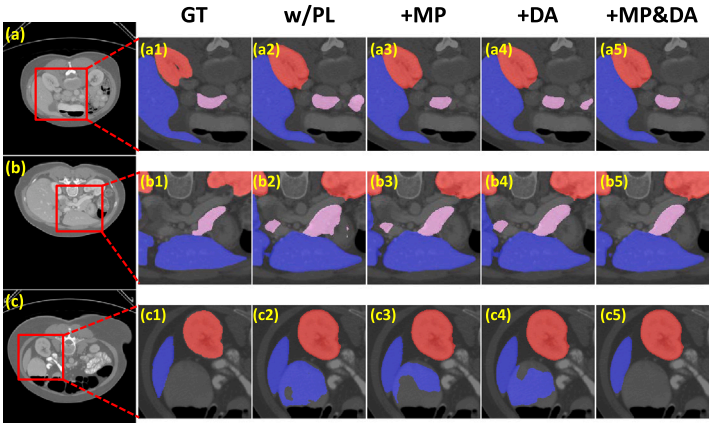

Fig. 6. Multi-organ prediction results compared on U-Net, ours without mutual-prior, and ours with mutual-prior. *MP* is short for mutual-prior. blue for liver, pink for pancreas, and red for kidney.

Table 4
Comparisons with state-of-the-art approaches. The subscript $_{single}$ means the model is trained with only $\{D_P, D_K, D_L\}$, without any pseudo-label.

Methods	Type	DSC (%) on $\{D_L, D_P, D_K\}$				DSC (%) on $\{D_T\}$				Speed (s/case)
		Liv.	Pan.	Kid.	Avg.	Liv.	Pan.	Kid.	Avg.	
U-Net$_{single}$	2D	91.22	75.23	94.80	87.09	84.74	36.19	80.07	67.00	**4.61**
PaNN [9]	2D	92.99	71.55	90.95	85.16	91.99	60.81	**86.93**	79.91	8.02
Co-training [10]	2D	93.09	71.04	90.79	84.97	90.99	66.29	80.24	79.17	8.69
CDCL [33]	2D	92.91	74.83	94.14	87.29	88.93	68.42	75.28	77.55	6.08
Med3D [11]	3D	**96.13**	80.54	93.75	90.14	89.57	**75.85**	79.11	81.51	238.63
DoDNet [12]	3D	95.70	**82.09**	**95.94**	**91.24**	91.94	73.32	85.68	83.65	250.68
TGNet [30]	3D	95.10	81.27	94.94	90.44	88.47	75.27	84.80	82.85	422.22
Ours$_{single}$	2D	***93.64***	***81.56***	***95.27***	***90.15***	86.14	48.17	80.42	71.58	15.61
Ours	2D	–	–	–	–	**93.73**	74.46	85.00	**84.40**	13.44

encoder and multiple task-specific decoders. (6) DoDNet: dynamic on-demand network proposed by Zhang et al. [12], which introduce an additional task coding and dynamic parameters to the segmentation head. (7)TGNet: The task-guided network proposed by Wu et al. [30] Specifically, the learning and testing process of [10] are separated between $\{D_{c1}, D_{c2}, D_{c3}\}$ and $\{D_T\}$, and [42] require fully supervised data in $\{D_T\}$, which is out of touch with the application scenario. To have a fair comparison, we re-implement PaNN [9], Co-training [10] and CDCL [33] method with the same backbone network as ours. Note that U-Net$_{single}$, PaNN, Co-training, and our PRIMP use 2D models, while Med3D, DoDNet, and TGNet adopt 3D models and with more computational complexity and time requirements. From Table 4, we have the following findings.

(1) On the source datasets, our single models are superior to other 2D methods, and comparable to other 3D methods. As indicated in the left part of Table 4, our single models with partial-prior and self-attention scheme have a significant improvement on all organs compared to other 2D methods in source datasets. Specifically,

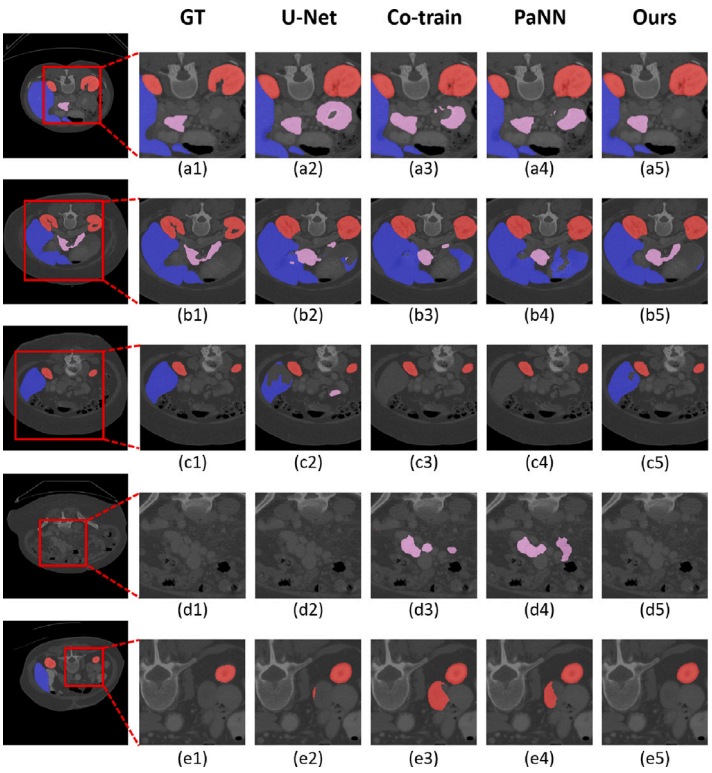

Fig. 7. Multi-organ segmentation results compared with the state-of-the-art models. We show the results of ground-truth (GT), U-Net [18], Co-training [10], PaNN [9], and the proposed PRIMP. blue for liver, pink for pancreas, and red for kidney.

our single model achieves an average DSC of 3.06%, 4.99%, 5.18%, 2.96% higher than U-Net$_{single}$, PaNN [9], Co-training [10], CDCL [33] respectively. It is worth noting that pancreas, which with relatively smaller size and more complex shape, witness the most significant improvement. These improvements are mainly due to (a) The incorporation of single-organ anatomical prior through PP, and (b) The enhanced model learning ability for organ edges by SA. Also, our single model is comparable to the other 3D models, e.g., 0.02 ahead of Med3D, 0.28/1.08 below TGNet/DoDNet. This is because through partial prior and corresponding position prediction operations, our model is able to obtain the spatial position information of organs to a certain extent. Although the overall Dice is slightly lower than TGNet/DoDNet, the computational complexity and time requirements of our 2D model are much lower than that of the 3D models. **(2) Our multi-organ model has a higher generalization ability than the others on target dataset.** From the right side of Table 4, we observe that: ***First***, despite the enhanced capability of the single-organ models, directly applying them on $\{D_T\}$ achieves a little improvement, with average DSC rises from 67.00% to 71.58%. ***Second***, PaNN, Co-training and CDCL, Med3D, DoDNet, TGNet are designed for this task and can achieve more significant gains: 8.33%, 7.59%, 5.97%, 9.93%, 12.07%, 11.27% higher than Ours$_{single}$, respectively. ***Third***, with MP and DA incorporated, PRIMP achieves consistent and remarkable boosts in three organs' accuracy on $\{D_T\}$. As a result, in D_T regarding the average DSC, our PRIMP achieves state-of-the-art results, which is 12.82%, 4.49%, 5.23%, 6.85%, 2.89%, 0.75%, 1.55% higher than Ours$_{single}$, PaNN, Co-training, CDCL, Med3D, DoDNet, and TGNet, respectively. On the other aspect, we can still observe a 7% and 10% accuracy gap of the pancreas and kidney between the segmentation on the source and target of our method. ***Fourth***, Our PRIMP achieves the best combination of inference time and segmentation accuracy. As can be seen from the last column of the table, although 3D-based methods achieved similar average Dice, they required much more interference time. For example, due to the corresponding 3D conv operations, Med3D, DoDNet, and TGNet require over 200 s for predicting one case on our device. Due to the need for feature map comparison in the mutual prior step, our model is slightly slower than other 2D models (e.g., PRIMP: 13.44 s vs. PaNN: 8.02 s). However, compared to 3D based models, the improvement in segmentation accuracy of our model is significant, and the increase in prediction time is acceptable.

Qualitative comparisons between PRIMP and SOTAs. We also visualize the segmentation results of our model and the comparison methods in Fig. 7, and have the following findings. *First*, compared to other SOTA models, PRIMP can better maintain the organ-specific morphology and position, *e.g.*, pancreas in (b5) and liver in (c5). *Second*, PRIMP is able to suppress the anatomical irrational predictions and the corresponding false positives. For example, compared to other results, the false positives in (a5, d5) for pancreas, (b5) for liver, and (e5) for kidney, are eliminated by PRIMP. Generally, the proposed partial- and mutual-prior helps the model maintaining the anatomical structure and generating more accurate segmentation results on the new target dataset.

6. Discussion

In this section, we briefly discuss the main strengths, the versatility, the limitations of our PRIMP model.

Strength: Unlike existing methods which do not explicitly consider the anatomical prior knowledge, our PRIMP innovatively incorporates organs' partial and mutual anatomical prior, and thus (1) Generating anatomically plausible, robust and accurate predictions; (2) Narrowing down the domain gaps and corresponding learning difficulties; (3) Capturing spatial information while with less computational complex than 3D models, thereby improving the organ segmentation performance. Note that all the experiments covered in this manuscript used public medical image segmentation datasets (for example, LiTS and KiTS), and there is no clinical validation and no pathological data.

PRIMP's versatility: Our research helps to overcome the practical problems in partially supervised scenarios, and achieve intelligent and robust abdominal CT multi-organ segmentation. More importantly, this

research intends to be a possible paradigm for solving partially supervised medical image segmentation tasks, and inspires other research on similar tasks (for example, optic disc, vessels, and pathological areas segmentation in ophthalmology research).

Limitations: We summarize the limitations of PRIMP as follows. (1) When generating statistical masks for both partial- and mutual-prior, PRIMP requires the orientation of all frames to be aligned, otherwise additional errors will be introduced. (2) Feature comparison and positioning operation in partial- and mutual-prior calculation is relatively time-consuming compared to other 2D models. (3) The inter-frame spatial information is not explicitly modeled, which may lead to further performance gains.

Future work: There are two main directions we considered for the future direction of this problem: (1) VAE-based compressed prior knowledge acquisition: We consider the use of VAE to obtain compressed knowledge representations in latent space for different single/multi-organs, and promote the segmentation model to follow the learned anatomical priors. (2) Self-supervised mutual reconstruction for narrowing domain gaps: In addition, we consider to introduce a self-supervised mechanism, where the semantic-level erasure and mutual reconstruction of target regions in different datasets is conduct. In this way, the domain gaps between different datasets are narrowed, and the generalization performance of the segmentation models are improved.

7. Conclusion

In this study, we propose the PRIMP framework to learn robust multi-organ segmentation model from several single-organ datasets, The partial- and mutual-priors are formulated as the statistical information over the datasets and encourage the PRIMP to give anatomically proper segmentation, both in organs' size, shape, and location. The ablation study and the comparison with the state-of-the-art solutions demonstrate the effectiveness of leveraging priors both when learning single-organ and multi-organ segmentation models. For future work, we will further investigate the anatomical prior from different aspects, such as more efficient prior knowledge representation and more flexible transfer/incremental learning among different modalities/datasets.

CRediT authorship contribution statement

Sheng Lian: Conceptualization, Methodology, Software, Writing – original draft. **Lei Li:** Software, Validation, Investigation, Visualization, Writing – original draft. **Zhiming Luo:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Zhun Zhong:** Data curation, Visualization, Writing – review & editing. **Beizhan Wang:** Supervision, Resources, Project administration. **Shaozi Li:** Resources, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to https://doi.org/10.1016/j.bspc.2022.104339. Sheng Lian reports financial support was provided by The National Natural Science Foundation of China. Sheng Lian reports a relationship with Western University that includes: non-financial support.

Data availability

We used open-sourced data, which is cited in the manuscript.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61876159, No. 62076116, No. 62276221), Guiding Project of Science and Technology Department of Fujian Province, China (No. 2019Y0018), and the Natural Science Foundation of Fujian Province of China (No. 2022J01002).

References

[1] B. Van Ginneken, C.M. Schaefer-Prokop, M. Prokop, Computer-aided diagnosis: how to move from the laboratory to the clinic, Radiology 261 (3) (2011) 719–732.
[2] V. Pekar, T.R. McNutt, M.R. Kaus, Automated model-based organ delineation for radiotherapy planning in prostatic region, Int. J. Radiat. Oncol.* Biol.* Phys. 60 (3) (2004) 973–980.
[3] I.J.Y. Wee, L.-J. Kuo, J.C.-Y. Ngu, A systematic review of the true benefit of robotic surgery: Ergonomics, Int. J. Med. Robot. Comput. Assist. Surg. 16 (4) (2020) e2113.
[4] X. Liang, N. Li, Z. Zhang, J. Xiong, S. Zhou, Y. Xie, Incorporating the hybrid deformable model for improving the performance of abdominal CT segmentation via multi-scale feature fusion network, Med. Image Anal. (MedIA) 73 (2021) 102156.
[5] P.-H. Conze, A.E. Kavur, E. Cornec-Le Gall, N.S. Gezer, Y. Le Meur, M.A. Selver, F. Rousseau, Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks, Artif. Intell. Med. 117 (2021) 102109.
[6] N. Heller, F. Isensee, K.H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han, et al., The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge, Med. Image Anal. (MedIA) 67 (2021) 101821.
[7] P. Bilic, P.F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, et al., The liver tumor segmentation benchmark (lits), 2019, arXiv.
[8] K. Dmitriev, A.E. Kaufman, Learning multi-class segmentations from single-class datasets, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019, pp. 9501–9511.
[9] Y. Zhou, Z. Li, S. Bai, C. Wang, X. Chen, M. Han, E. Fishman, A.L. Yuille, Prior-aware neural network for partially-supervised multi-organ segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV, 2019, pp. 10672–10681.
[10] R. Huang, Y. Zheng, Z. Hu, S. Zhang, H. Li, Multi-organ segmentation via co-training weight-averaged models from few-organ datasets, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI, Springer, 2020, pp. 146–155.
[11] S. Chen, K. Ma, Y. Zheng, Med3D: Transfer learning for 3D medical image analysis, 2019, arXiv preprint arXiv:1904.00625.
[12] J. Zhang, Y. Xie, Y. Xia, C. Shen, DoDNet: Learning to segment multi-organ and tumors from multiple partially labeled datasets, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2021, pp. 1195–1204.
[13] M. Roy, J. Kong, S. Kashyap, V.P. Pastore, F. Wang, K.C. Wong, V. Mukherjee, Convolutional autoencoder based model HistoCAE for segmentation of viable tumor regions in liver whole-slide images, Sci. Rep. 11 (1) (2021) 1–10.
[14] S. Wang, K. Sun, L. Wang, L. Qu, F. Yan, Q. Wang, D. Shen, Breast tumor segmentation in DCE-MRI with tumor sensitive synthesis, IEEE Trans. Neural Netw. Learn. Syst. (TNNLS) (2021).
[15] D. Nie, L. Wang, Y. Gao, J. Lian, D. Shen, STRAINet: Spatially varying stochastic residual AdversarIal networks for MRI pelvic organ segmentation, IEEE Trans. Neural Netw. Learn. Syst. (TNNLS) 30 (5) (2018) 1552–1564.
[16] Z. Tian, X. Li, Y. Zheng, Z. Chen, Z. Shi, L. Liu, B. Fei, Graph-convolutional-network-based interactive prostate segmentation in MR images, Med. Phys. 47 (9) (2020) 4164–4176.
[17] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 3431–3440.
[18] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI, Springer, 2015, pp. 234–241.
[19] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI, Springer, 2016, pp. 424–432.
[20] J.M.J. Valanarasu, V.A. Sindagi, I. Hacihaliloglu, V.M. Patel, Kiu-net: Over-complete convolutional architectures for biomedical image and volumetric segmentation, IEEE Trans. Med. Imaging (TMI) (2021).
[21] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2) (2021) 203–211.

[22] S.A. Taghanaki, Y. Zheng, S.K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, G. Hamarneh, Combo loss: Handling input and output imbalance in multi-organ segmentation, Comput. Med. Imaging Graph. (CMIG) 75 (2019) 24–33.

[23] A. Sinha, J. Dolz, Multi-scale self-guided attention for medical image segmentation, IEEE J. Biomed. Health Inform. (JBHI) 25 (1) (2020) 121–130.

[24] K. Chaitanya, N. Karani, C.F. Baumgartner, A. Becker, O. Donati, E. Konukoglu, Semi-supervised and task-driven data augmentation, in: International Conference on Information Processing in Medical Imaging, IPMI, Springer, 2019, pp. 29–41.

[25] J. Peng, G. Estrada, M. Pedersoli, C. Desrosiers, Deep co-training for semi-supervised image segmentation, Pattern Recognit. (PR) 107 (2020) 107269.

[26] X. Liu, Q. Yuan, Y. Gao, K. He, S. Wang, X. Tang, J. Tang, D. Shen, Weakly supervised segmentation of covid19 infection with scribble annotation on ct images, Pattern Recognit. (PR) 122 (2022) 108341.

[27] X. He, L. Fang, M. Tan, X. Chen, Intra-and inter-slice contrastive learning for point supervised OCT fluid segmentation, IEEE Trans. Image Process. (TIP) (2022).

[28] G. Shi, L. Xiao, Y. Chen, S.K. Zhou, Marginal loss and exclusion loss for partially supervised multi-organ segmentation, Med. Image Anal. (MedIA) 70 (2021) 101979.

[29] G. Zhang, Z. Yang, B. Huo, S. Chai, S. Jiang, Multiorgan segmentation from partially labeled datasets with conditional nnU-Net, Comput. Biol. Med. 136 (2021) 104658.

[30] H. Wu, S. Pang, A. Sowmya, Tgnet: A task-guided network architecture for multi-organ and tumour segmentation from partially labelled datasets, in: 2022 IEEE 19th International Symposium on Biomedical Imaging, ISBI, IEEE, 2022, pp. 1–5.

[31] N. Dong, M. Kampffmeyer, X. Liang, M. Xu, I. Voiculescu, E. Xing, Towards robust partially supervised multi-structure medical image segmentation on small-scale data, Appl. Soft Comput. 114 (2022) 108074.

[32] X. Fang, P. Yan, Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction, IEEE Trans. Med. Imaging (TMI) 39 (11) (2020) 3619–3629.

[33] L. Wang, D. Li, Y. Zhu, L. Tian, Y. Shan, Cross-dataset collaborative learning for semantic segmentation, 2021, arXiv preprint arXiv:2103.11351.

[34] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019, pp. 3146–3154.

[35] X. Li, Z. Zhong, J. Wu, Y. Yang, Z. Lin, H. Liu, Expectation-maximization attention networks for semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV, 2019, pp. 9167–9176.

[36] A.L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B.A. Landman, G. Litjens, B. Menze, et al., A large annotated medical image dataset for the development and evaluation of segmentation algorithms, 2019, arXiv preprint arXiv:1902.09063.

[37] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, et al., The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes, 2019, arXiv preprint arXiv:1904.00445.

[38] Multi-atlas labeling beyond the cranial vault workshop and challenge, https://doi.org/10.7303/syn3193805.

[39] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems (NeurIPS), Vol. 32, 2019, pp. 8026–8037.

[40] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 1492–1500.

[41] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[42] Y. Zhou, J. Bai, Multiple abdominal organ segmentation: an atlas-based fuzzy connectedness approach, IEEE Trans. Inf. Technol. Biomed. 11 (3) (2007) 348–352.