Medical Image Segmentation With Deep Atlas Prior

Huimin Huang, Han Zheng, Lanfen Lin, Ming Cai, Hongjie Hu, Qiaowei Zhang, Qingqing Chen, Yutaro Iwamoto, *Member, IEEE*, Xianhua Han, *Member, IEEE*, Yen-Wei Chen, *Member, IEEE*, and Ruofeng Tong

***Abstract*— Organ segmentation from medical images is one of the most important pre-processing steps in computer-aided diagnosis, but it is a challenging task because of limited annotated data, low-contrast and non-homogenous textures. Compared with natural images, organs in the medical images have obvious anatomical prior knowledge (e.g., organ shape and position), which can be used to improve the segmentation accuracy. In this paper, we propose a novel segmentation framework which integrates the medical image anatomical prior through loss into the deep learning models. The proposed prior loss function is based on probabilistic atlas, which is called as deep atlas prior (DAP). It includes prior location and shape information of organs, which are important prior information for accurate organ segmentation. Further, we combine the proposed deep atlas prior loss with the conventional likelihood losses such as Dice loss and focal loss into an adaptive Bayesian loss in a Bayesian framework, which consists of a prior and a likelihood. The adaptive Bayesian loss dynamically adjusts the ratio of the DAP loss and the likelihood loss in the training epoch for better learning. The proposed loss function is universal and can be combined with a wide variety of existing deep segmentation models to further enhance their performance. We verify the significance of our proposed framework with some state-of-the-art models, including fully-supervised and semi-supervised segmentation models on a public dataset (ISBI LiTS 2017 Challenge) for liver segmentation and a private dataset for spleen segmentation.**

***Index Terms*— Medical image segmentation, deep atlas prior, probabilistic atlas, adaptive bayesian loss.**

I. Introduction

ORGAN segmentation from medical images such as computed tomography (CT) images is a critical step in computer-assisted clinical interventions and an important reference for disease diagnosis. Consequently, a variety of computer vision techniques have been developed for medical image segmentation. These methods can be summarized as: active contours [1]–[3], region-based approaches [4], [5], graph-cut-based or random walk-based interactive methods [10]–[13], anatomic models-based methods [6]–[9].

Recently, a growing interest has been seen in deep learning-based semantic segmentation, such as FCN [14], U-Net [15] and its variants [16], [17], DeepLab [18]–[21], et al. Goodfellow [25] presented the generative adversarial network (GAN) model which have achieved great success in deep learning and machine learning by mutual adversarial learning between generative model and discriminative model and Luc *et al.* [26] improved the GAN model for image segmentation. Some state-of-the-art networks [22]–[24] including semi-supervised networks [27]–[30] are extensions of these models. These models achieved satisfied segmentation results for natural images.

However, compared with natural images, organ segmentation is a challenging task since the medical images are gray-scale images with low-contrast (even with some potential invisible parts) and some neighboring organs have similar intensities and textures. Thus, we can employ anatomical prior knowledge (e.g., organ shape and position) to facilitate accurate segmentation. In traditional medical image segmentation methods [10], [13], the probabilistic atlas, i.e., statistical information about the spatial distribution of the organ, was used as a common prior for medical image segmentation.

Manuscript received May 7, 2021; accepted June 6, 2021. Date of publication June 15, 2021; date of current version November 30, 2021. This work was supported in part by the Major Scientific Research Project of Zhejiang Laboratory under Grant 2020ND8AD01 and in part by the Grant-in-Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under Grant 20KK0234, Grant 21H03470, and Grant 20K21821. *(Huimin Huang and Han Zheng are co-first authors.) (Corresponding authors: Lanfen Lin; Hongjie Hu; Yen-Wei Chen.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ritsumeikan University Ethics Review Committee for Medical and Health Research Involving Human Subjects under No. BKC-LSMH-2021-037, and performed in line with the principles of the Declaration of Helsinki.

Huimin Huang, Han Zheng, Lanfen Lin, and Ming Cai are with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China (e-mail: 11921060@zju.edu.cn; 21721112@zju.edu.cn; llf@zju.edu.cn; cm@zju.edu.cn).

Hongjie Hu, Qiaowei Zhang, and Qingqing Chen are with the Department of Radiology, Sir Run Run Shaw Hospital, Hangzhou 310009, China (e-mail: hongjiehu@zju.edu.cn; radiologist@163.com; qingqingchen@zju.edu.cn).

Yutaro Iwamoto is with the College of Information Science and Engineering, Ritsumeikan University, Kyoto 603-8577, Japan (e-mail: yiwamoto@fc.ritsumei.ac.jp).

Xianhua Han is with the Graduate School of Science and Technology for Innovation, Yamaguchi University, Yamaguchi 753-8511, Japan (e-mail: hanxhua@yamaguchi-u.ac.jp).

Yen-Wei Chen is with the College of Information Science and Engineering, Ritsumeikan University, Kyoto 603-8577, Japan, also with the Research Center for Healthcare Data Science, Zhejiang Lab (Phase I), Hangzhou 311121, China, and also with the College of Computer Science, Zhejiang University, Hangzhou 310058, China (e-mail: chen@is.ritsumei.ac.jp).

Ruofeng Tong is with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China, and also with the Research Center for Healthcare Data Science, Zhejiang Laboratory, Hangzhou 311121, China (e-mail: trf@zju.edu.cn).

Digital Object Identifier 10.1109/TMI.2021.3089661

Recently, a few studies have combined the deep learning with the atlas. AtlasNet [31] aligned multiple anatomies to reduce biological variability via non-linear transformers to obtain an atlas. The multi-network architecture improved the accuracy but also increased network parameters. Huo *et al.* [32] proposed the spatially localized atlas network tiles (SLANT) method spatialized for high-resolution whole-brain segmentation. The SLANT method employed multiple spatially distributed networks to learn contextual information of a fixed spatial location. The recent work [45], [46] also used the atlas to combine registration and segmentation processes. However, the aforementioned methods directly incorporated the atlas prior into network architecture, which may increase network parameters and/or only well-designed for specific organs. Different from these methods, we focus the research on combining the anatomic prior knowledge with a loss function as a framework of medical image segmentation. The proposed loss function is universal and can be combined with a wide variety of existing deep segmentation models to further enhance their performance.

Defining a loss function is a crucial issue in machine learning and deep learning techniques. For classification tasks, the commonly used losses are zero-one loss, logistic loss, and hinge loss [33]. In contrast, squared loss, absolute loss and Huber loss are more applicable for regression tasks. In addition, cross-entropy loss and dice loss [34] are often used in segmentation tasks. Focal loss [36] was originally proposed for object detection, but also applies to the two-class classification and segmentation tasks. Note that these losses are not specific to medical images, and they employ the probabilities predicted from previous learning epochs without any prior knowledge, which are referred to as likelihood losses.

In this paper, we propose a framework that integrates the anatomical prior through loss into a deep learning model to realize accurate organ segmentation. We design a prior loss function based on the probabilistic atlas for medical image segmentation, which is referred to as deep atlas prior (DAP). Considering the imbalance between the positive and negative samples in training datasets, we also propose asymmetric DAP loss to relatively increase the weights of the small number of positive samples. Furthermore, we incorporate the proposed DAP loss and other likelihood losses into the Bayesian loss, similar to a Bayesian framework with a prior and a likelihood. Here, we propose two Bayesian losses, i.e., a fixed Bayesian loss, which uses a fixed ratio of prior loss and likelihood loss, and an adaptive Bayesian loss, which optimizes itself by adaptively adjusting the ratio of these two losses. For the adaptive Bayesian loss, deep learning models will increase or decrease the influence of prior knowledge automatically according to the current situation. If the prior knowledge is not worth learning for the given task, the ratio of DAP loss will be reduced. Otherwise, the ratio of the likelihood loss will be reduced. We verify the significance of the proposed loss by combining it with state-of-the art models including both fully-supervised and semi-supervised segmentation models on a public dataset (i.e., ISBI LiTS 2017 Challenge) for liver segmentation and a private dataset [13] for spleen segmentation.

Our primary contributions are summarized as follows:

- We propose a framework that integrates the anatomical prior through loss into a deep learning model to realize accurate organ segmentation. The proposed method can be combined with a wide variety of existing deep segmentation models to further enhance their performance.
- The probabilistic atlas is a representative prior feature in medical image, and we extracte it from the CT images after registration. We introduce a new loss function called as Deep Atlas Prior (DAP) loss for accurate organ segmentation, and, to balance positive and negative samples, we also propose an asymmetrical DAP loss.
- By incorporating the DAP loss with conventional likelihood losses, we propose a Bayesian loss for medical image segmentation in a Bayesian framework comprised of a likelihood and a prior. In addition, we also propose an adaptive Bayesian loss, which optimizes itself in the training phase to realize adaptive learning.
- We verify the significance of our proposed framework with some state-of-the-art models, including fully-supervised and semi-supervised segmentation models on a public dataset for liver segmentation and a private dataset for spleen segmentation. Experiment results demonstrate the significance of the proposed method.

Preliminary work was presented as a conference paper in the 2019 International Conference on Medical Image Computing and Computer Assisted Intervention [47]. This paper involves methodological and experimental extensions and validations. We evolve fixed Bayesian loss into adaptive Bayesian loss to automatically adjust the ratio of the DAP loss and the likelihood loss to adapt to different optimization goals. In addition, we integrate the proposed loss functions with the most competitive deep segmentation network architectures including both supervised and semi-supervised models to demonstrate the versatility of our method, and we expand experiments on different organs including the liver and the spleen.

The remainder of this paper is organized as follows. Section II introduces related work. The proposed DAP loss and adaptive Bayesian loss are described in Section III. Subsequently, Section IV discusses experiments and results, and finally Section V concludes the paper.

II. Related Work

In this section, we briefly introduce some studies related to our work, including CNN-based segmentation methods, loss functions and probabilistic atlas.

A. CNN-Based Segmentation

There are many semantic segmentation models based on CNN and the typical models are FCN [14], U-Net [15] and its variants [16], [17], DeepLab [18]–[21], etc. FCN [14] replaced the fully connected layers in the traditional convolutional neural network (CNN) with convolutional layers to classify the image at the pixel level. U-Net [15] consisted of a contracting path that captures context and a symmetric expanding path that enables precise localization. In 2015, Google proposed DeepLab [18], which combined the deep CNN and a fully connected conditional random field (CRF). Then, Google

proposed DeepLabv2 [19], v3 [20] and v3+ [21], which included improvements such as atrous convolution, atrous spatial pyramid pooling (ASPP).

These models were used for medical image segmentation with some specific pre-processing to improve the segmentation accuracy. Liu *et al.* [37] combined the simple linear iterative clustering (SLIC) of super-pixels, support vector machine (SVM) and CNN to classify the pixels in CT images. Roth *et al.* [38] presented a two-stage, coarse-to-fine approach that train two FCN models, one for roughly delineating the region of interest and the other for segmenting the details of organs. He *et al.* [39] also showed a two-stage FCN model to segment three pelvic organs. They did organ detection with a novel distinctive curve in the first stage then learn the segmentation map and distinctive curve separately in the second stage. Larsson *et al.* [40] obtained an organ localization by a robust feature registration method and segment the region of interest via a 3D convolution neural network. Recently, GAN-based segmentation models have been used for accurate medical image segmentation. For example, Luc *et al.* [26] replaced the generator in GAN with a semantic segmentation network and used the discriminator to refine the segmentation results. Hung *et al.* [27] extended it to a semi-supervised learning model that uses the distinguish ability of the discriminator to pick up a part of the unannotated volumes for training. In addition, Souly *et al.* [28] replaced the discriminator in a GAN with a segmentation network, and the generator was employed to augment the training samples.

In this study, we combined the proposed methods with various models to verify their versatility.

B. Loss Functions

In deep learning, there are many basic loss functions and different loss functions usually correspond to different learning objectives. Cross-entropy loss is used to evaluate the gap between labels and predictions and it is the most common loss in deep learning models. Squared loss uses Euclidean distance as the error metric and is always used in regression tasks. Hinge loss [33] is dedicated to the two-class problem and is an important loss used by support vector machines (SVM).

There are many variations of these basic loss functions, e.g., mean squared loss, binary cross-entropy, weighted cross-entropy, and squared hinge loss [42]. Further, some studies have proposed specific losses for specific research problems. Milletari *et al.* [34] proposed a new loss function called the Dice loss function from the Dice score coefficient, which was originally applied to evaluate the segmentation results. Further, Sudre *et al.* [35] proposed a more robust loss called the generalized Dice overlap for unbalanced tasks. Focal loss [36] is a variation of cross- entropy loss, which makes neural networks focus on hard samples and reduce ineffective learning. These losses employ the predicted probabilities of the training samples, which can be called as likelihood losses. In our study, we proposed a prior loss based on the probabilistic atlas and combine these two types of losses into an adaptive Bayesian loss.

C. Probabilistic Atlas

Probabilistic atlas is the statistical information about the spatial distribution of the organ in medical images, and there has been a lot of research on atlas in medical image segmentation. Before the advent of deep learning, Park *et al.* [10] proposed a method to construct a probabilistic atlas of four organs in the abdomen and then to incorporate the atlas into the Bayesian framework to aid the segmentation of low-contrast organs. Park's study is also a combination of probabilistic atlas and the Bayesian models, however in this research, we further apply it to deep learning, which performs better than conventional machine learning methods. In addition to the Bayesian models, the probabilistic atlas is also used in other algorithms. A single reference may cause a bias in a probabilistic atlas, Dong *et al.* [13] proposed a template matching framework with an iterative probabilistic atlas. Tong *et al.* [41] proposed an automated multi-organ segmentation method to generate a probabilistic atlas using dictionary learning and sparse coding techniques. They obtained final segmentation results after post-processing based on the graph-cuts method.

With the increasing use of neural networks, various studies have combined deep learning techniques and the probabilistic atlas. For example, Vakalopoulou *et al.* [31] proposed AtlasNet, which obtained a different atlas by using multiple forward transformers mapping all training images to a common subspace to reduce biological variability. In addition, Zeng *et al.* [45] considered the mean shape of the liver as a prior, and used the features learned from the network to deform the mean shape and obtain the segmentation results of test samples. Ding *et al.* [46] proposed a DL-based fusion strategy to selectively pick up a reliable atlas and vote for label fusion. In the proposed framework, we extract the probabilistic atlas as a prior to facilitate the network training; however, the difference is that we incorporate the atlas into the proposed loss functions.

III. Proposed Methods

In this section, we describe the proposed methods. First, we present an overview of the proposed framework with DAP. Subsequently, we introduce how we obtain the probabilistic atlas and convert it to DAP loss. Finally, we discuss the proposed adaptive Bayesian loss, which combines prior and likelihood losses, and optimizes itself during training epochs.

A. Overview

An overview of the proposed medical image segmentation framework with DAP is shown in Fig. 1. After registration, the training volumes are sent to the segmentation network. The main obstacles of training directly with 3D volumes lie in the expensive computational cost, memory requirement and time consumption. To reduce training time, we use a 2D slice image (target image) with its multiple neighboring slices as input images to catch 3D spatial information. Therefore, the input image $\boldsymbol{X}$ has three channels, including the slice to be segmented and the upper and lower slices in the proposed framework. Given input image $\boldsymbol{X}$ of the size (H, W, 3), the segmentation network (SN) outputs a segmentation feature

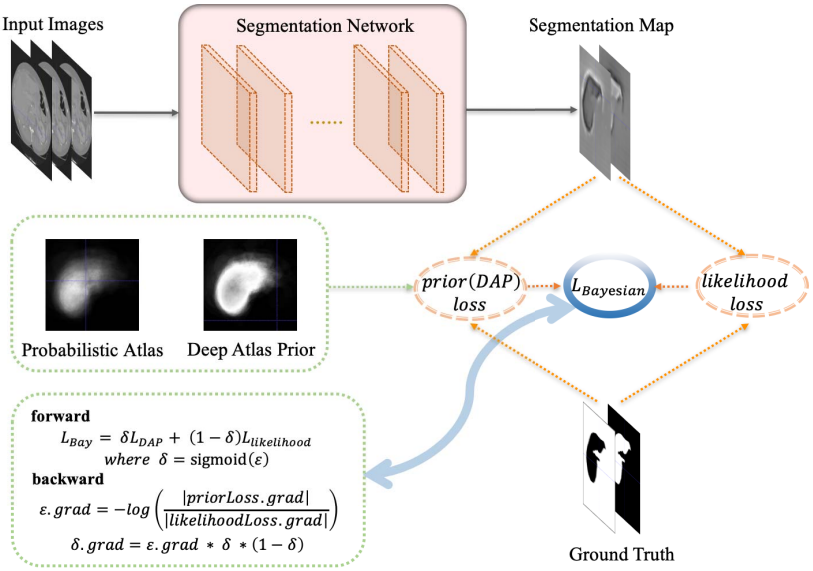

Fig. 1. Overview of the medical image segmentation framework with deep atlas prior.

map of the size (H, W, C), where C is the number of segmentation classes, and H and W are the height and width of the input image $\boldsymbol{X}$. Here, the SN can be any neural network structure, such as FCN, U-Net and DeepLab.

DAP is a prior that is transformed from the probabilistic atlas. DAP is applied in the loss function to guide model training, and we further combine it with the conventional likelihood loss to realize the adaptive Bayesian loss.

B. Probabilistic Atlas

The resolution and size of the CT volumes in our datasets are inconsistent; therefore, it is necessary to register the volumes to obtain the probabilistic atlas. Registration of the original medical image volumes is too complicated to implement; therefore, we extracted the bones and performed bone registration because bones form a rigid framework of the body.

The details of extracting and registration refer to [13]; the process of extracting bones which does not require any annotation and directly manipulate the original 3D CT volumes. We select the middle slice of the input volumes first and extract the abdominal region by Otsu thresholding [43]. After extracting the abdominal region, we calculate the threshold value T_{bone} of the bones. We assume the intensity distribution of the abdominal region as a Gaussian model $G(\mu, \sigma)$ and initialize as $T_{bone} = \mu$. We consider the area whose intensity is above T_{bone} is bones and increase T_{bone} iteratively with $0.2 * \mu$ until the ratio of the bone's area to the image is less than the threshold value Σ. Based on the experimental results, we set the threshold value Σ to 0.03.

In bone registration, the choice of the reference bone is crucial. Some of the training volumes are scanned from the shoulder to the legs, whereas other training volumes are just scanned from the liver region to the pelvis. To obtain more accurate registration results, we choose the volume with the largest range as the reference. After extracting the bones, we register by Simple ITK [49]–[51], which is an open-source, cross-platform system for the segmentation and registration. We obtain the registration transform by using bone registration.

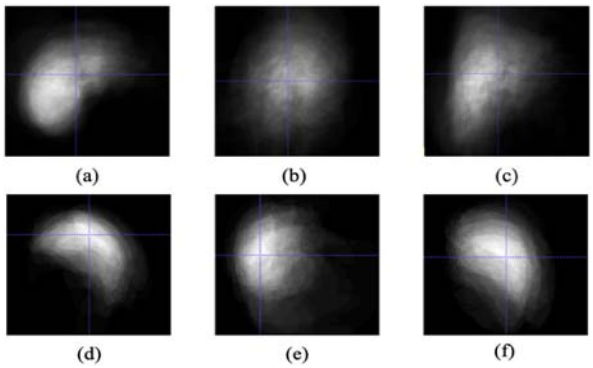

Fig. 2. The 3D Probabilistic atlas of liver and spleen; (a), (b), and (c) are the liver probabilistic atlas; (d), (e), and (f) are the spleen probabilistic atlas; (a) and (d) are in the superior–inferior direction; (b) and (e) are in the left–right direction; and (c) and (f) are in the anterior–posterior direction.

Instead of directly registering the medical image volumes, bones are extracted, and bone registration is performed on the input volumes and ground truths, which is a simpler but effective way as the first step. After the transform, we average the ground truth of all the training volumes to obtain the probabilistic atlas. The probabilistic atlas is the position information of the target organ, as shown in Fig. 2. Fig. 2(a), (b), and (c) are three views of the liver probabilistic atlas extracted from the dataset of the ISBI LiTS 2017 Challenge. Fig. 2(d), (e), and (f) are three different views of the spleen probabilistic atlas extracted from the hospital dataset.

C. Deep Atlas Prior

Inspired by the focal loss [36], we proposed DAP loss. Cross entropy loss is usually used in the general segmentation network. However, vast number of easy negatives can overwhelm training, which contribute no useful learning signal and lead to degenerate models. For the focal loss, it is

believed that the easy samples are useless as compared with the difficult samples; therefore, it decreases the weight of the easily classified samples and makes the network focus on the difficultly classified samples.

From the probabilistic atlas, we know the prior about the location of the target organ. Each value in probabilistic atlas represents the probability that the organ will appear at that point, and we hope to reduce the weight of the determined organ or non-organ points. Thus, we choose a Gaussian function to convert the probabilistic atlas to the DAP, which is a prior weight of our loss function:

$$W_{DAP} = exp\left(-\frac{(PA-0.5)^2}{2\sigma^2}\right) \tag{1}$$

where σ is a hyper-parameter and PA is the probabilistic atlas, which is 3D just like a medical CT image. When training the model, we pick the related 2D atlas for each slice and calculate the loss of weight. The value of PA was between 0 and 1. This means that the target organ has never appeared at this point if the value is 0. Conversely, if the value is 1, it means that the organ is always present at this point. In the training process, we hope to reduce the weight of the determined points. Therefore, if the PA is 0.5, which refers to the difficultly classified samples according to the probabilistic atlas, the weight will be maximal. However, if the PA is 0 or 1, which refers to the easily classified samples, the weight will be minimal. The smaller the σ value, the smaller the weight of the determined points, and the more we want to focus our training attention on the point where PA equals 0.5. The larger the σ value, the closer the sample weight is to 1, which indicates that our DAP loss is closer to the original cross entropy loss.

The Gaussian function in Eq. (1) is symmetrical and is determined based on the organ and non-organ having the same weight. To address the problem of imbalance between positive and negative samples, we further proposed the asymmetrical DAP weight to relatively increase the weight of the positive samples and decrease the weight of the negative samples:

$$W_{a-DAP} = \begin{cases} \exp\left(-\frac{(PA-0.5)^2}{2\sigma_1^2}\right) & 0 \le PA \le 0.5 \\ \exp\left(-\frac{(PA-0.5)^2}{2\sigma_2^2}\right) & 0.5 < PA \le 1 \end{cases} \tag{2}$$

We set $\sigma_1 < \sigma_2$ to ensure that the weight of the positive samples would be greater than the weight of the negative samples. The results of visualizing the DAP can be seen in Fig. 1. We can see that the position of the organ edge is brighter than the organ center and the background, which indicates that the training weight of the organ center and background is reduced by the DAP.

We propose a DAP loss function based on the α-balanced cross-entropy loss. Our DAP loss was as follows:

$$L_{DAP} = -\sum_{\mathrm{h,w}} W_{a-DAP} \sum_{c \epsilon C} \alpha_c Y^{(h,w,c)} log\left(S(X)^{(h,w,c)}\right) \tag{3}$$

where X is the input images and Y is the label; h and w are the height and weight of the X; C is the labels; the S means the segmentation network and α_c is the hyper-parameter of the α-balanced cross-entropy loss:

$$\alpha_c = \begin{cases} \alpha, & y = 0 \\ 1-\alpha, & y = 1 \end{cases} \tag{4}$$

D. Bayesian Loss

The main idea of the focal loss is to perform a deep training procedure to optimize the model parameters for correctly recognizing both the easy and difficult samples. The conventional focal loss employs the predicted probabilities of the training samples in the previous learning epoch for mining the sample recognition difficulty; this conventional loss can be called the likelihood focal loss. Similarly, the Dice loss is also a likelihood loss in a sense.

This study explores the atlas prior for difficult sample mining in deep medical image segmentation so that the deep training procedure can concentrate the uncertain samples with position priori to optimize the model parameters. Thus, we regard our DAP loss as a prior loss.

By treating the likelihood loss and the DAP loss as the likelihood and prior of the Bayesian loss, respectively, we can put them together as follows:

$$L_{Bayesian} = \delta L_{prior} + (1-\delta) L_{likelihood} \tag{5}$$

where L_{prior} is our DAP loss and $L_{likelihood}$ is focal loss or Dice loss. δ is a factor between 0 and 1. When $\delta = 0.00$, the model uses only the likelihood loss, and when $\delta = 1.00$, the model uses only the prior loss, which is our DAP loss. Although in the Bayesian model, the likelihood and prior are multiplied, we can convert the multiplication into addition by using a logarithmic function.

E. Adaptive Bayesian Loss

In our initially proposed Bayesian loss (Eq. (5)), the δ is a hyper-parameter that controls the ratio of the two losses via manual adjustment; however, in order to make a more accurate adjustment of loss, we set it as an adaptive parameter. The following sigmoid function is used to control the range of δ:

$$\delta = sigmoid(\varepsilon) = \frac{1}{1+e^{-\varepsilon}} \tag{6}$$

where ε can be any real number, and δ takes a value between 0 and 1. Here, it is important to consider that, during the training process, ε is a loss parameter rather than a network parameter. Thus, its optimization goal differs from that of the network parameters, which leads to differences in their gradient descent.

The optimization goal of the network parameters is to reduce the loss. However, the adaptive Bayesian loss controls the ratio of different losses using ε. If ε is updated like a network parameter, it will prefer smaller loss, which weakens the learning ability of the network. Thus, the optimization direction of ε and the network parameters should be opposite, which means the ε attempts to increase loss. Here, ε is used to select a loss that is more worth learning. We determine

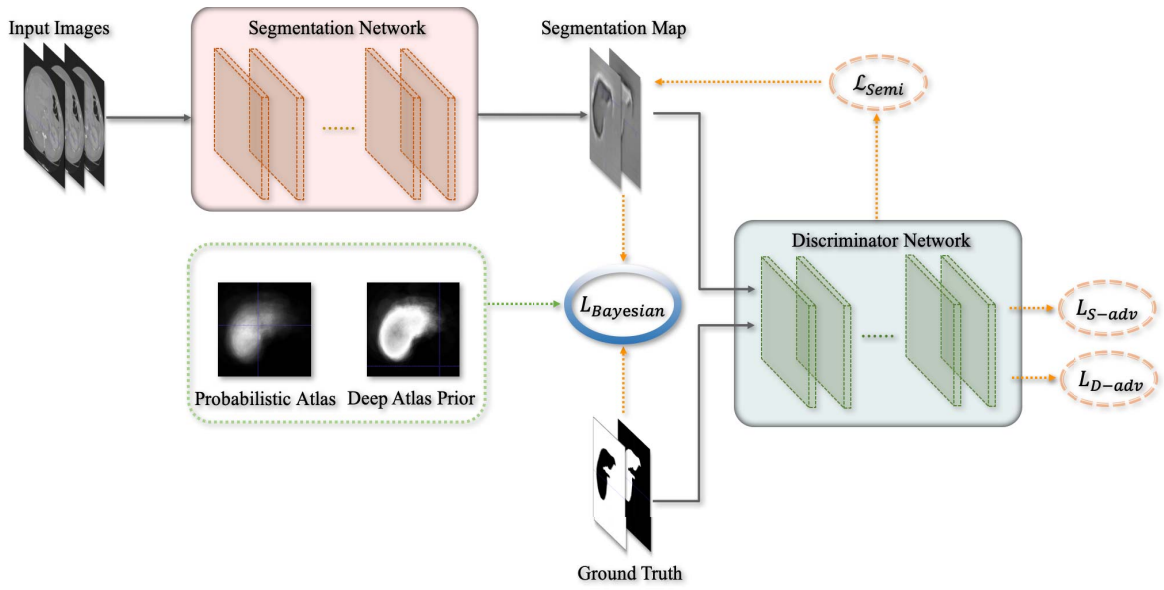

Fig. 3. Overview of semi-supervised segmentation framework. The input of discriminator network includes the ground truth and the segmentation map and the output of discriminator assists to pick up reliable pixels form unannotated samples.

whether the loss is worth learning according to its influence on the network, i.e., the gradient changes of network parameters after each loss backward. We use the average gradient changes of the feature map to represent that of the network parameters because the backpropagation is passed into the parameter through the feature map. The gradient calculation of ε is given as follows:

$$\varepsilon.grad = -log\left(\frac{|priorLoss.grad|}{|likelihoodLoss.grad|}\right) \tag{7}$$

where the *priorLoss.grad* and *likelihoodLoss.grad* are the average gradient changes of the feature map after DAP loss and likelihood loss backward, respectively. According to Eq.(7), if the gradient changes of the prior loss are greater than that of the likelihood loss, the gradient of ε below 0 and the value of ε and δ will increase in the next training step, which means that the weight of the prior loss increases. This is in line with the optimization goal ofε.

In this equation, division is applied rather than subtraction because the distribution of the training data may differ for each batch, and subtraction is easily affected by the data distribution.

Here, the *log* function is used to make the two gradients symmetric relative to 0.

$$\begin{aligned}\frac{\partial\delta}{\partial\varepsilon} &= \frac{\partial\left(\frac{1}{1+e^{-\varepsilon}}\right)}{\partial\varepsilon} = \frac{e^{-\varepsilon}}{\left(1+e^{-\varepsilon}\right)^2} \\ &= \frac{1}{1+e^{-\varepsilon}} * \left(1-\frac{1}{1+e^{-\varepsilon}}\right) \\ &= \delta * (1-\delta)\end{aligned} \tag{8}$$

According to the derivation of the sigmoid function in Eq. (8), the gradient of δ can be obtained as follows:

$$\delta.grad = \varepsilon.grad * \delta * (1-\delta) \tag{9}$$

F. Application to Semi-Supervised Segmentation

The proposed loss can also be applied to GAN-based semi-supervised models [27] as shown in Fig. 3. Compare to the fully-supervised models shown in Fig. 1, the framework of the semi-supervised segmentation included a discriminator network that distinguishes the ground truth and segmentation results on the annotated samples and pick up reliable pixels on unannotated samples. Note that the discriminator is a simple CNN with four convolutional layers and one up-sample layer.

In addition to obtaining better segmentation results, there is another optimization objective in the GAN, i.e., adversarial learning. For adversarial learning, the segmentation and discriminator networks have different losses:

$$L_{S-adv} = -\sum_{h,w} \log\left(D\left(S\left(X_n\right)\right)^{(h,w)}\right) \tag{10}$$

$$\begin{aligned}L_{D-adv} = -\sum_{h,w} &\left((1-y)\log\left(1-D\left(D_{in}\right)^{(h,w)}\right)\right. \\ &\left.+ y log\left(D\left(D_{in}\right)^{(h,w)}\right)\right)\end{aligned} \tag{11}$$

where D is the discriminator network and D_{in} is the input of D, which can be the prediction from the SN or one-hot coded ground truths. If the D_{in} is the prediction, $y = 0$; otherwise $y = 1$. Note that the adversarial learning losses in Eq. (10) and Eq. (11) do not need to be applied to the proposed Bayesian loss.

In addition to the two adversarial losses, the semi-supervised segmentation model shown in Fig. 3 also has two segmentation losses that correspond to annotated and unannotated training samples. Note that the segmentation loss of the annotated samples is consistent with the loss in Eq. (5); another loss of the unannotated samples is different:

$$L_{Semi} = -\sum_{h,w} I\left(D\left(S\left(X_n\right)\right)^{(h,w)} > T_{semi}\right) f_{bce} \tag{12}$$

where

$$f_{bce} = \sum_{c \epsilon C} \alpha_c \hat{y}_n log\left(S\left(X\right)^{(h,w,c)}\right) \tag{13}$$

where $D\left(S\left(X_n\right)\right)$ is the output of the discriminator, i.e., the confidence score map of the segmentation results. If the values of the map are upper to a threshold T_{semi} (set to 0.3), we select the related pixels to calculate the loss using an indicator function. Here, f_{bce} is a α-balanced binary cross-entropy loss and $\hat{y}_n$ represents the prediction labels from the SN.

We replace f_{bce} in Eq. (13) with $L_{Bayesian}$ and multiply the $D\left(S\left(X_n\right)\right)$ as confidence map:

$$L_{Semi} = -\sum_{h,w} I\left(D\left(S\left(X_n\right)\right)^{(h,w)} > T_{semi}\right) * D\left(S\left(X_n\right)\right)^{(h,w)} L_{Bayesian} \tag{14}$$

where we use the prediction labels as the ground truths in $L_{Bayesian}$. We believe that the proposed loss is also applicable to semi-supervised networks because both annotated and unannotated medical images have the anatomical prior. Noted that, in the semi-supervised mode, we construct the probability atlas and DAP loss without unannotated training samples.

IV. Experiments

A. Datasets

We conducted experiments to verify our proposed method on two organs: the liver and spleen. (**i**) **LiTS dataset** [53] for liver organ segmentation: ISBI LiTS 2017 Challenge dataset contained 131 contrast-enhanced 3D abdominal CT scans with different sizes, which were divided into 80 samples for training, 23 samples for validation and 28 samples for testing. This dataset was acquired by different scanners and protocols from six different clinical sites, with a largely varying in-plane resolution from 0.55 mm to 1.0 mm and slice spacing from 0.45 mm to 6.0 mm. The image resolution was a relatively high 512×512. (**ii**) **Spleen dataset** [13]: The private spleen dataset consisted of 49 low-contrast volumes, of which 32, 8 and 9 volumes were used for training, validation and testing, respectively. In this dataset, each case had a slice-plane resolution of 512×512 pixels and was resampled with the same spacing of $1.0 \times 1.0 \times 1.0$ mm. It is worth to note that only the training dataset was calculated to develop the probabilistic atlas (PA), excluding the validation and testing datasets. After the bone registration, the size of the volumes was unified to $320 \times 320 \times 400$ on liver and $320 \times 320 \times 206$ on spleen. To utilize 3D context information, the input image consisted of three slices: the slice to be segmented and the upper and lower slices, that is, $X \in \mathbb{R}^{320 \times 320 \times 3}$.

B. Experimental Setting

1) Implementation: The SN in our framework could be any semantic SN, and we performed the experiments with U-Net, DeepLabV2 (Resnet101) and GAN. The generator of GAN was DeepLabV2 (Resnet101), and the discriminator was a simple CNN with four convolutional layers and an upsampled layer; the activation function was Leaky ReLU. U-Net, DeepLabV2 and GAN were the three basic networks of semantic segmentation. Many other studies have improved on these basic models. Here we used only these three basic networks to verify the effectiveness of our DAP loss. Our experimental results may not improve upon the results of more complex network models with multi-scale or multi-models. We used stochastic gradient descent to optimize our network. The learning rates were initialized to 1e-4 for the U-Net and to 1e-3 for the DeepLabV2 and GAN. The momentum was initialized to 0.9, and the weight decay was initialized to 5e-4 in all the models. The batch size was 5.

2) Evaluation: The evaluation relied on a train-validation-test strategy. The actual training of the models was accomplished on the training set. The hyper-parameters optimization and network development were conducted on the validation set. The overall performance of every method was evaluated on the test set.

To evaluate our models, three metrics were used to measure the accuracy of segmentation results, including the volumetric overlap error (VOE), 95% Hausdorff distance (95HD) and Dice coefficient per case (DSC). For the first two evaluation metrics, the smaller the value is, the better the segmentation result.

C. Hyper-Parameters

1) α of Balanced Cross-Entropy: Our DAP loss was designed based on α-balanced cross-entropy; therefore, we first determined the hyper-parameter α first. α is the weight of the negative samples in the training, and the weight of positive samples to $1-\alpha$, correspondingly. Since the number of positive samples in the training data was much smaller than the number of negative samples; α is usually less than $1-\alpha$, which means $\alpha < 0.5$. The results of the liver and spleen segmentation with different balanced cross-entropy were shown in Table I. We can see that the result is better when $\alpha = 0.2$ in the liver dataset, and $\alpha = 0.3$ in our spleen dataset. We set the corresponding values of α for all subsequent experiments.

2) σ_1 and σ_2 of W_{a-DAP}: In the process of adjusting α, we initially solved the problem of sample imbalance, but this was too simple and not enough; therefore, we hope to go further in our DAP loss. To differentiate between the weights of the positive and negative samples, we propose a new asymmetrical DAP loss using two Gaussian functions with different σ values. The weight of the positive samples is always larger than that of the negative samples. Thus, σ_1 is always lower than σ_2 in the asymmetrical loss function for increasing the weight of the positive samples. Table II showed the results with the different σ_1 and σ_2 values of W_{a-DAP}. The lower left corner of the table showed that $\sigma_1 > \sigma_2$, and the weight of the negative samples was larger than that of the positive samples. We did not perform this part of the experiment because we wanted to focus our training more on the positive samples. The results in the Table II diagonal have been obtained with different symmetrical DAP losses ($\sigma_1 = \sigma_2$), and the other results were obtained with asymmetrical DAP losses ($\sigma_1 < \sigma_2$).

TABLE I
SEGMENTATION RESULTS WITH DIFFERENT BALANCED CROSS-ENTROPY

α	0.4	0.3	0.2	0.1
Liver	91.25%	91.43%	**93.20%**	91.25%
Spleen	88.37%	**91.12%**	90.80%	89.47%

TABLE II
SYMMETRICAL AND ASYMMETRICAL DAP LOSS

	σ_1 \ σ_2	0.25	0.50	1.00	2.00
Liver	**0.25**	91.80%	92.41%	91.84%	91.96%
	0.50	-	93.37%	**93.90%**	93.52%
	1.00	-	-	93.47%	93.85%
	2.00	-	-	-	93.76%
Spleen	**0.25**	91.48%	90.57%	90.84%	90.24%
	0.50	-	90.91%	**91.74%**	91.70%
	1.00	-	-	90.24%	91.22%
	2.00	-	-	-	90.18%

TABLE III
δ IN FIXED BAYESIAN LOSS WITH FOCAL LOSS

	δ	0.00	0.25	0.50	0.75	1.00
Liver	**DeepLabV2**	92.59%	93.39%	**93.64%**	93.57%	92.66%
	U-Net	93.85%	93.12%	**94.10%**	93.96%	93.90%
	GAN	93.24%	94.08%	**94.57%**	94.30%	93.97%
Spleen	**DeepLabV2**	91.31%	**93.18%**	92.58%	91.04%	92.02%
	U-Net	88.64%	87.15%	92.64%	**92.88%**	91.74%
	GAN	91.92%	91.53%	**92.79%**	92.77%	92.74%

TABLE IV
δ IN FIXED BAYESIAN LOSS WITH DICE LOSS

	δ	0.00	0.25	0.50	0.75	1.00
Liver	**DeepLabV2**	92.29%	92.20%	93.02%	**94.04%**	92.66%
	U-Net	91.73%	92.75%	92.93%	93.19%	**93.90%**
	GAN	91.30%	93.47%	94.02%	**94.74%**	93.97%
Spleen	**DeepLabV2**	93.41%	**94.39%**	94.12%	93.70%	92.02%
	U-Net	91.18%	91.69%	**93.10%**	92.24%	91.74%
	GAN	87.50%	89.92%	91.12%	92.08%	**92.74%**

For comparing the symmetrical and asymmetrical DAP losses and finding the optimal values, we tried different σ_1 and σ_2 values. From the experimental results, it is clear that the asymmetrical DAP losses were always better than the symmetrical DAP losses. We found that the segmentation results for the liver (93.90%) and spleen (91.74%) had the same σ_1 (0.50) and σ_2 (1.00) values.

3) δ of Fixed Bayesian Loss: For fixed Bayesian loss, the δ is a hyper-parameter, which is the weight of DAP loss. We chose the focal loss and Dice loss as our likelihood losses, adjusted the value of δ and performed experiments using the DeepLabV2, U-Net and GAN models. When $\delta = 0.00$, the model used only the likelihood loss, and when $\delta = 1.00$, the model used only the DAP loss.

The results that regard the focal loss as the likelihood loss were shown in Table III. A comparison of the results in the leftmost and rightmost columns showed that our DAP loss was always better than the focal loss, which illustrated that prior knowledge had a huge impact on medical image segmentation. Comparing the results in the middle and edge columns showed that the middle results were always better. This proved that combining the prior loss and the likelihood loss into a Bayesian loss in the training models can further improve the segmentation performance. Different models and datasets sometimes gave optimal results with different δ values, which showed that the importance of the prior and likelihood varies in different situations.

The results of the Dice loss as the likelihood loss were shown in Table IV. We can also see that the DAP loss performed better than the Dice loss, and their Bayesian loss performed even better. The experimental results in Table III and Table IV verified that our Bayesian loss, which assembled a likelihood loss and a prior loss, was optimal under different δ on different models and datasets. Since adjusting δ for each models and datasets was time consuming, we proposed an

TABLE V
SIGNIFICANCE OF PROPOSED LOSS FUNCTIONS IN SUPERVISED MODE ON BOTH LIVER AND SPLEEN DATASETS (P-VALUE < 0.05)

Losses (Liver Dataset)	DeepLabV2			U-Net			GAN		
	DSC (%)	VOE (%)	95HD (voxel)	DSC (%)	VOE (%)	95HD (voxel)	DSC (%)	VOE (%)	95HD (voxel)
Cross-Entropy [15]	90.21	17.70	17.79	92.86	13.13	14.71	92.28	14.26	15.15
Focal Loss [36]	92.59	13.78	14.92	93.85	11.52	13.74	93.24	12.68	14.32
Dice Loss [34]	92.29	14.21	15.20	91.73	15.03	15.95	91.30	15.86	16.38
DAP Loss	92.66	13.54	14.81	93.90	11.44	13.69	93.97	11.29	13.55
Bayesian (focal)	93.64	11.86	13.83	94.10	11.07	13.06	94.57	10.33	12.40
Bayesian (Dice)	94.04	11.00	12.98	93.90	11.58	13.67	**94.74**	**9.87**	**11.66**
Adaptive Bayesian (focal)	**94.78**	**9.85**	**11.64**	**94.17**	**10.96**	**12.83**	94.64	10.14	12.39
Adaptive Bayesian (Dice)	94.63	10.17	12.05	94.03	11.17	13.51	94.54	10.27	11.73
Losses (Spleen Dataset)	**DeepLabV2**			**U-Net**			**GAN**		
	DSC (%)	VOE (%)	95HD (voxel)	DSC (%)	VOE (%)	95HD (voxel)	DSC (%)	VOE (%)	95HD (voxel)
Cross-Entropy [15]	90.97	16.16	18.26	88.53	19.78	23.73	89.19	19.13	22.26
Focal Loss [36]	91.31	15.81	17.76	88.64	19.55	23.49	91.92	14.42	16.44
Dice Loss [34]	93.41	12.21	10.49	91.18	16.17	17.89	87.50	22.01	24.91
DAP Loss	92.02	14.50	14.44	91.74	15.07	16.84	92.74	13.51	13.07
Bayesian (focal)	93.18	12.72	10.94	92.88	13.25	11.59	92.79	13.17	12.57
Bayesian (Dice)	94.39	10.65	9.03	93.10	12.83	11.12	92.74	13.20	12.92
Adaptive Bayesian (focal)	**95.25**	**9.13**	**8.13**	**93.25**	**12.63**	**10.69**	93.18	12.48	10.81
Adaptive Bayesian (Dice)	94.97	9.53	8.47	92.83	13.20	11.64	**93.74**	**11.65**	**10.12**

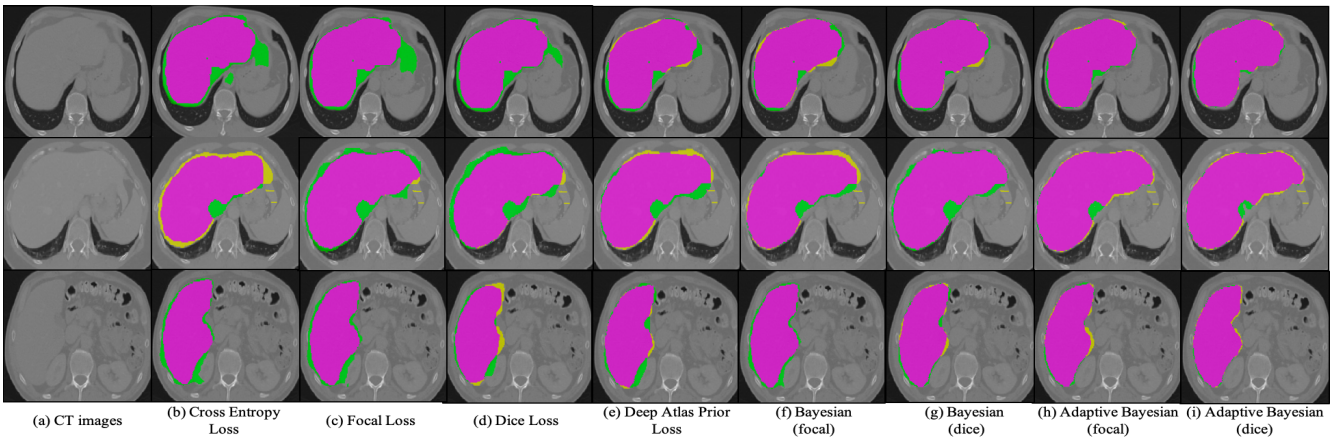

Fig. 4. Visualization of three different liver segmentation results with different loss functions. The **purple areas** are the true positive (TP) samples; the **yellow areas** are the false negative (FN) samples; and the**green areas** are the false positive (FP) samples.

adaptive Bayesian loss which automatically optimized itself during the training process.

D. Significance of the Proposed Loss Functions

We validated the significance of the proposed loss function on different models, e.g., U-Net, DeepLabV2 (Resnet101), and GAN. We also applied the proposed losses to a previously proposed semi-supervised segmentation model [27].

The proposed loss functions include DAP loss, fixed Bayesian loss with optimal δ and adaptive Bayesian loss. The Bayesian loss comprised the DAP loss and a conventional likelihood loss, which we tried two different likelihood losses (focal loss and Dice loss). Hence, we considered four different Bayesian losses in our experiments, i.e., fixed Bayesian loss with focal loss, fixed Bayesian loss with Dice loss, adaptive Bayesian loss with focal loss, and adaptive Bayesian loss with Dice loss. We compared the fixed Bayesian and adaptive Bayesian losses to verify that the adaptation of δ is significant.

Table V compared the results obtained with the different losses. The performance of the proposed method surpassed the cross-entropy, focal, and Dice losses on all models. In addition, the adaptive Bayesian loss outperformed the fixed Bayesian loss on most models, and the adaptive Bayesian loss with focal loss outperformed better.

We further visualized the experimental results obtained for the contrast-enhanced liver (in Fig.4) and low-contrast spleen (in Fig.5), where (a) was a CT image, and (b), (c), (d) were the segmentation results obtained with cross-entropy, focal, and Dice losses, respectively. (e) showed the result for DAP loss, and (f) showed the result for the fixed Bayesian loss consisting of the DAP loss and the focal loss. (g) showed the result obtained for fixed Bayesian loss with DAP and Dice losses; and (h) showed the result for adaptive Bayesian loss with DAP and focal losses. Finally, (i) showed the result for adaptive Bayesian loss with the DAP loss and Dice losses.

The results demonstrated that, for the segmentation results obtained with the DAP (Fig. 4(e) and Fig.5(e)) segments on

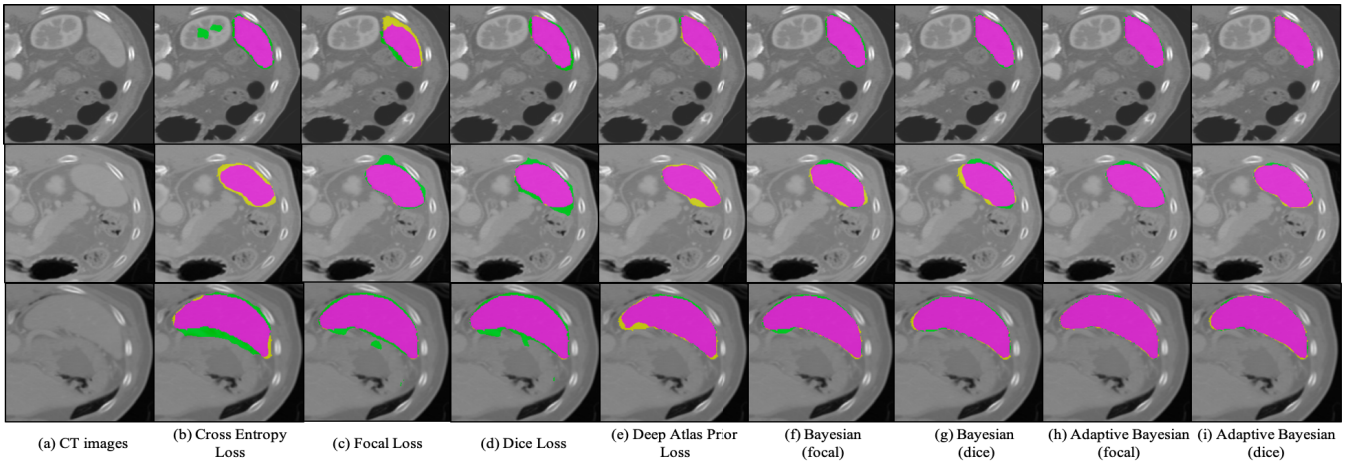

Fig. 5. Visualization of three different spleen segmentation results with different loss functions. The **purple areas** are the true positive (TP) samples; the **yellow areas** are the false negative (FN) samples; and the**green areas** are the false positive (FP) samples.

TABLE VI
COMPARE TO STATE-OF-THE-ART MODELS ON LIVER AND SPLEEN DATASETS (P-VALUE < 0.05)

Models (Liver Dataset)	Without our loss			With our loss		
	DSC(%)	VOE(%)	95HD(voxel)	DSC(%)	VOE(%)	95HD(voxel)
DeeplabV3**[20]**	93.25	12.38	14.34	**94.67**	**10.09**	**12.18**
PSANet**[1]**	94.97	9.56	11.19	**95.17**	**9.15**	**10.66**
DANet**[22]**	95.19	9.17	10.71	**95.29**	**9.66**	**10.49**
DUNet**[48]**	95.12	9.28	10.83	**95.31**	**8.98**	**10.44**
EMANet**[24]**	95.40	9.49	10.21	**96.05**	**8.50**	**9.53**
Models (Spleen Dataset)	**Without our loss**			**With our loss**		
	DSC(%)	VOE(%)	95HD(voxel)	DSC(%)	VOE(%)	95HD(voxel)
DeeplabV3**[20]**	95.02	9.41	8.34	**95.20**	**9.15**	**8.14**
PSANet**[1]**	94.97	9.72	8.52	**95.24**	**9.06**	**8.03**
DANet**[22]**	94.18	10.93	9.33	**95.41**	**8.97**	**7.82**
DUNet**[48]**	95.49	8.84	7.74	**95.56**	**8.72**	**7.65**
EMANet**[24]**	95.76	8.40	7.42	**95.77**	**8.22**	**7.24**

TABLE VII
APPLICATION TO SEMI-SUPERVISED MODELS ON LIVER AND SPLEEN DATASETS (P-VALUE < 0.05)

Losses (Liver Dataset)	0.3: 0.7			0.5: 0.5			0.7: 0.3		
	DSC (%)	VOE (%)	95HD (voxel)	DSC (%)	VOE (%)	95HD (voxel)	DSC (%)	VOE (%)	95HD (voxel)
Cross-Entropy [15]	92.16	14.63	15.42	93.60	11.98	13.90	93.97	11.46	13.67
Focal Loss [36]	90.64	16.99	17.38	92.04	14.59	15.48	93.34	12.60	14.08
Dice Loss [34]	92.82	13.32	14.76	94.13	11.10	13.12	94.03	11.21	13.53
DAP Loss	93.48	12.21	14.01	93.75	11.63	13.74	94.25	10.85	12.67
Bayesian (focal)	93.46	11.86	13.91	93.46	12.23	13.95	94.45	10.42	12.48
Bayesian (Dice)	93.54	12.08	13.88	93.40	12.39	14.07	94.09	11.03	13.30
Adaptive Bayesian (focal)	**93.97**	**11.42**	**13.52**	94.13	11.01	12.92	**94.50**	**10.38**	**12.42**
Adaptive Bayesian (Dice)	93.34	12.30	14.25	**94.15**	**10.81**	**12.66**	94.25	10.72	12.53
Losses (Spleen Dataset)	**0.3: 0.7**			**0.5: 0.5**			**0.7: 0.3**		
	DSC (%)	VOE (%)	95HD (voxel)	DSC (%)	VOE (%)	95HD (voxel)	DSC (%)	VOE (%)	95HD (voxel)
Cross-Entropy [15]	88.73	19.25	23.10	91.55	15.52	17.14	91.10	16.30	18.03
Focal Loss [36]	82.48	28.92	29.49	84.14	28.92	27.42	84.34	26.57	27.27
Dice Loss [34]	88.74	19.16	23.62	89.58	18.58	22.76	90.23	17.29	20.33
DAP Loss	85.55	23.73	26.81	89.88	18.16	21.35	90.83	16.47	19.40
Bayesian (focal)	86.66	22.46	25.72	91.94	14.37	15.25	91.65	15.30	17.00
Bayesian (Dice)	86.43	24.03	26.07	90.49	16.57	20.65	90.66	16.96	19.86
Adaptive Bayesian (focal)	**89.69**	**18.42**	**21.80**	91.56	15.09	17.23	92.42	13.71	14.27
Adaptive Bayesian (Dice)	89.23	18.85	22.81	**92.52**	**13.49**	**13.51**	**92.94**	**12.86**	**11.58**

edge were better than those of the cross-entropy, focal, and Dice losses. In addition, the Bayesian loss that combined the likelihood and prior losses yielded further segmentation results. More importantly, we find that the adaptive Bayesian

TABLE VIII
SEGMENTATION PERFORMANCE OF DIFFERENT ATLAS ACCURACY CALCULATED WITH DIFFERENT NUMBER OF TRAINING SAMPLES RANGING FROM 10% TO 100%. (DEEPLABV3 ARCHITECTURE)

Method	With our loss				w/o our loss
	100%	70%	50%	10%	
Dice(%)	94.67	94.55	94.38	94.12	93.25

loss produced more accurate boundaries than the other losses despite the existence of complex background with a challenging edge, which means that our δ optimization strategy is effective.

E. Combined With State-of-the-art Models

We combined the proposed adaptive Bayesian loss with some state-of-the-art models and compared them to their original results. The results were shown in Table VI, which showed the experimental results on the liver and spleen datasets.

The experiment results demonstrate that the proposed framework with prior is competitive, and the proposed loss can be combined with other state-of-the-art models to improve the results of organ segmentation results in most cases.

F. Semi-Supervised Segmentation Results

The results for the GAN-based semi-supervised model were shown in Table VII. Here, we adjusted the ratio of annotated to unannotated images in these experiments. The proposed loss improved the segmentation results of the semi-supervised model at different ratios. Note that only annotated training images were applied to obtain the probabilistic atlas in the semi-supervised model training phase. By applying our adaptive Bayesian loss on semi-supervised models, we have further verified the versatility of the proposed loss.

G. Simulations of Atlas Accuracy

In this section, we investigated how the accuracy and bias of the atlas affect the segmentation results. As known, probabilistic atlas is obtained by averaging the ground truth of training volumes. The more training volumes used, the higher the atlas accuracy. In view of this, we calculated the atlas by using different number of training volumes, ranging from 10% to 100%. As can be seen in TABLE VIII, the segmentation performance on the DeepLabV3 architecture slightly declined when decreased the accuracy of the atlas. However, our proposed adaptive Bayesian loss with focal loss still surpassed the original focal loss, which demonstrated the effectiveness of employing anatomical prior knowledge.

V. DISCUSSIONS AND CONCLUSION

In this paper, we have proposed a framework that incorporates the anatomical prior into the deep learning models. In the proposed framework, a DAP loss function based on a probabilistic anatomical atlas was proposed for medical image segmentation. The proposed DAP includes priori location and shape information of organs, which are important information for accurate organ segmentation. We incorporated the proposed DAP loss with conventional likelihood losses (i.e., Dice and focal losses) into a Bayesian loss, similar to a Bayesian framework with a prior and a likelihood. In addition, we proposed a parameter optimization strategy for the adaptive Bayesian loss, which obtained better segmentation results than the fixed parameter Bayesian loss. We conducted extensive experiments using the public ISBI LiTS 2017 Challenge dataset for liver segmentation and a private dataset for spleen segmentation. The results verified the significance of the proposed losses for fully-supervised and semi-supervised models. In addition, we combined the proposed methods to some state-of-the-art models. The experimental results demonstrate that the prior loss can be used to facilitate medical image segmentation and the proposed adaptive Bayesian loss outperformed the individual prior and likelihood losses in medical image segmentation tasks.

To test the statistical significance of the segmentation accuracy differences between the proposed methods with Bayesian or adaptive Bayesian loss functions and existing methods without Bayesian loss functions, we employed t-test for Table V, VI, and VII. The results of analysis confirmed the statistically significant (p-value < 0.05) superior performance of the proposed methods with Bayesian or adaptive Bayesian loss functions against all methods without Bayesian loss.

We also observed that the proposed loss did not behave similarly for the liver and spleen, as seen in Table V and Table VII. The dissimilar behave may be due to the following reason: the large liver dataset contained 131 CT scans, which was acquired by different scanners and protocols from six different clinical sites; while the relative small spleen dataset consisted of 49 CT scans, which was obtained from one hospital. Thus, the liver dataset has a wider variability compared to the spleen.

Limitation: Despite our method achieves superior performance, its limitation should also be noted. The probabilistic atlas is suitable for the organ with relatively fixed position, but lose efficacy for tumor with large positional and morphological change. As for future work, the proposed method could be incorporated with other meaningful priors such as those proposed by Cherukuri *et al*. [52].

ACKNOWLEDGMENT

The authors would like to thank Prof. Akira Furukawa of Tokyo Metropolitan University, Japan and Dr. Shuzo Kanasaki of Koseikai Takeda Hospital, Japan for providing us spleen dataset.

REFERENCES

[1] S. Zhou, J. Wang, S. Zhang, Y. Liang, and Y. Gong, "Active contour model based on local and global intensity information for medical image segmentation," *Neurocomputing*, vol. 186, pp. 107–118, Apr. 2016.
[2] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.
[3] A. H. Foruzan *et al.*, "Segmentation of liver in low-contrast images using K-means clustering and geodesic active contour algorithms," *IEICE Trans. Inf. Syst.*, vol. E96.D, no. 4, pp. 798–807, 2013.
[4] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 6, pp. 641–647, Jun. 1994.
[5] Z. Pan and J. Lu, "A Bayes-based region-growing algorithm for medical image segmentation," *Comput. Sci. Eng.*, vol. 9, no. 4, pp. 32–38, Jul. 2007.

[6] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N–D image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 2, pp. 109–131, Nov. 2006.

[7] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Sep. 2006.

[8] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann, "Random walks for interactive organ segmentation in two and three dimensions: Implementation and validation," in *Proc. MICCAI*, 2005, pp. 773–780.

[9] Y. Yuan, Y.-W. Chen, C. Dong, H. Yu, and Z. Zhu, "Hybrid method combining superpixel, random walk and active contour model for fast and accurate liver segmentation," *Computerized Med. Imag. Graph.*, vol. 70, pp. 119–134, Dec. 2018.

[10] H. Park, P. H. Bland, and C. R. Meyer, "Construction of an abdominal probabilistic atlas and its application in segmentation," *IEEE Trans. Med. Imag.*, vol. 22, no. 4, pp. 483–492, Apr. 2003.

[11] M. G. Linguraru, J. K. Sandberg, Z. Li, F. Shah, and R. M. Summers, "Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation," *Med. Phys.*, vol. 37, no. 2, pp. 771–783, Jan. 2010.

[12] T. Okada *et al.*, "Automated segmentation of the liver from 3D CT images using probabilistic atlas and multilevel statistical shape model," *Academic Radiol.*, vol. 15, no. 11, pp. 1390–1403, Nov. 2008.

[13] C. Dong *et al.*, "Segmentation of liver and spleen based on computational anatomy models," *Comput. Biol. Med.*, vol. 67, pp. 146–160, Dec. 2015.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[15] O. Ronneberger *et al.*, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Cham, Switzerland: Springer, 2015, pp. 234–241.

[16] Z. Zhou *et al.*, "Unet++: A nested U-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[17] H. Huang *et al.*, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.

[18] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*. [Online]. Available: http://arxiv.org/abs/1412.7062

[19] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[20] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[21] L. C. Chen *et al.*, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 801–818.

[22] J. Fu *et al.*, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3146–3154.

[23] H. Zhao *et al.*, "PsaNet: Point-wise spatial attention network for scene parsing," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 267–283.

[24] X. Li, Z. Zhong, J. Wu, Y. Yang, Z. Lin, and H. Liu, "Expectation-maximization attention networks for semantic segmentation," 2019, *arXiv:1907.13426*. [Online]. Available: http://arxiv.org/abs/1907.13426

[25] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1–7.

[26] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," 2016, *arXiv:1611.08408*. [Online]. Available: http://arxiv.org/abs/1611.08408

[27] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," 2018, *arXiv:1802.07934*. [Online]. Available: http://arxiv.org/abs/1802.07934

[28] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5688–5696.

[29] X. Liu *et al.*, "Semi-supervised automatic segmentation of layer and fluid region in retinal optical coherence tomography images using adversarial learning," *IEEE Access*, vol. 7, pp. 3046–3061, 2019.

[30] D. Nie *et al.*, "ASDNet: Attention based semi-supervised deep networks for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 370–378.

[31] M. Vakalopoulou *et al.*, "AtlasNet: Multi-atlas non-linear deep networks for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 658–666.

[32] Y. Huo *et al.*, "3D whole brain segmentation using spatially localized atlas network tiles," *NeuroImage*, vol. 194, pp. 105–119, Jul. 2019.

[33] Y. Tang, "Deep learning using linear support vector machines," 2013, *arXiv:1306.0239*. [Online]. Available: http://arxiv.org/abs/1306.0239

[34] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[35] C. H. Sudre *et al.*, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 240–248.

[36] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[37] X. Liu *et al.*, "Automatic organ segmentation for CT scans based on super-pixel and convolutional neural networks," *J. Digit. Imag.*, vol. 31, no. 5, pp. 748–760, Oct. 2018.

[38] H. R. Roth *et al.*, "Hierarchical 3D fully convolutional networks for multi-organ segmentation," 2017, *arXiv:1704.06382*. [Online]. Available: http://arxiv.org/abs/1704.06382

[39] K. He, X. Cao, Y. Shi, D. Nie, Y. Gao, and D. Shen, "Pelvic organ segmentation using distinctive curve guided fully convolutional networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 585–595, Feb. 2019.

[40] M. Larsson, Y. Zhang, and F. Kahl, "Robust abdominal organ segmentation using regional convolutional neural networks," *Appl. Soft Comput.*, vol. 70, pp. 465–471, Sep. 2018.

[41] T. Tong *et al.*, "Discriminative dictionary learning for abdominal multi-organ segmentation," *Med. Image Anal.*, vol. 23, no. 1, pp. 92–104, Jul. 2015.

[42] C.-P. Lee and C.-J. Lin, "A study on L_2-loss (squared hinge-loss) multiclass SVM," *Neural Comput.*, vol. 25, no. 5, pp. 1302–1323, May 2013.

[43] J. Zhang and J. Hu, "Image segmentation based on 2D Otsu method with histogram analysis," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, vol. 6, Dec. 2008, pp. 105–108.

[44] P. F. Christ *et al.*, "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Cham, Switzerland: Springer, 2016, pp. 415–423.

[45] Q. Zeng *et al.*, "Liver segmentation in magnetic resonance imaging via mean shape fitting with fully convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Cham, Switzerland: Springer, 2019, pp. 246–254.

[46] Z. Ding, X. Han, and M. Niethammer, "VoteNet +: An improved deep learning label fusion method for multi-atlas segmentation," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2020, pp. 363–367.

[47] H. Zheng *et al.*, "Semi-supervised segmentation of liver using adversarial learning with deep atlas prior," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2019, pp. 148–156.

[48] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019.

[49] R. Beare, B. Lowekamp, and Z. Yaniv, "Image segmentation, registration and characterization in R with SimpleITK," *J. Stat. Softw.*, vol. 86, no. 8, pp. 1–35, 2018.

[50] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, "SimpleITK image-analysis notebooks: A collaborative environment for education and reproducible research," *J. Digit. Imag.*, vol. 31, no. 3, pp. 290–303, Jun. 2018.

[51] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The design of SimpleITK," *Frontiers Neuroinform.*, vol. 7, p. 45, Dec. 2013.

[52] V. Cherukuri, T. Guo, S. J. Schiff, and V. Monga, "Deep MR brain image super-resolution using spatio-structural priors," *IEEE Trans. Image Process.*, vol. 29, pp. 1368–1383, 2020.

[53] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1–13.