IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 15, 2021, accepted March 29, 2021, date of publication April 1, 2021, date of current version June 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3070391

MCMC Guided CNN Training and Segmentation for Pancreas Extraction

MU TIAN[1], JINCHAN HE[1], XIAXIA YU[1], CHUDONG CAI[4], AND YI GAO[1,2,3]

[1]School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China
[2]Pengcheng Laboratory, Shenzhen 518060, China
[3]Marshall Laboratory of Biomedical Engineering, Shenzhen 518060, China
[4]Department of General Surgery, Shantou Central Hospital, The Affiliated Shantou Hospital of Sun Yat-sen University, Shantou 515031, China

Corresponding author: Yi Gao (gaoyi@szu.edu.cn)

This work was supported in part by the Department of Education of Guangdong Province Funding under Grant 2017KZDXM072, in part by the National Natural Science Foundation of China under Grant 61601302, in part by the Shenzhen Peacock Plan under Grant KQTD2016053112051497, and in part by the Faculty Development Grant of Shenzhen University under Grant 2018009.

ABSTRACT Efficient organ segmentation is the precondition of various quantitative analysis. Segmenting the pancreas from abdominal CT images is a challenging task because of its high anatomical variability in shape, size and location. What's more, the pancreas only occupies a small portion in abdomen, and the organ border is very fuzzy. All these factors make the segmentation methods of other organs less suitable for pancreas. In this work, we propose a Markov Chain Monte Carlo (MCMC) guided convolutional neural network (CNN) approach, in order to handle such difficulties in morphological and photometric variabilities. Specifically, the proposed method mainly consists of three steps: First, registration is carried out to mitigate the body weight and location variability. Then, an MCMC scheme is designed to guide the adaptive selection of 3D patches, which are fed to the CNN for training. At the same time, the pancreas distribution is also learned for subsequent segmentation. Eventually, the same MCMC process guides the segmentation process with patch-wise predictions fused using a Bayesian voting scheme. This method is evaluated on the NIH pancreatic dataset including 82 abdominal contrast-enhanced CT volumes. We have achieved a competitive result of 78.13% Dice Similarity Coefficient value and 82.65% Recall value in testing data.

INDEX TERMS Pancreas segmentation, image registration, MCMC, 3D convolutional neural network.

I. INTRODUCTION

For many applications of computer aided analysis, obtaining accurate segmentation for organs is a critical prerequisite step. In recent years, with the rapid development of deep neural networks, automatic segmentation of many organs and tissues have achieved good results, such as for cortical and sub-cortical structures, lung, liver, heart, etc.. [1]–[6]. However, segmenting the pancreas accurately from CT images remains a challenging task. Though there were many advances since the pre-deep learning era, the accuracy of pancreas segmentation is still relatively low comparing to other organs [7]–[12]. Comparing to other organs, the shape, size and position of the pancreas vary greatly in abdomen among different patients. In CT images, the contrast between the pancreas and its surrounding tissues is weaker. Also, the pancreas is relatively soft and easy to be pushed by surrounding organs, leading to larger shape deformation. Last but not least, the pancreas occupies only a small portion of the entire CT image. Figure 1 shows the CT images with the pancreas annotated from 4 sampled patients in the NIH dataset. We can see that the pancreas has large variations in size, shape and location among different patients; also it occupies only a small part of the whole volume.

Formally, segmentation is a learning task where we need to derive a mapping from an image to a binary mask. Comparing to conventional machine learning approaches heavily relying on the design of hand crafted features, deep neural networks substantially improved the model capacity and generalization ability with automatic learning of feature representations optimized for particular tasks. In the field of semantic segmentation, deep learning plays the fundamental part for almost all state-of-the-art advances [13]–[16]. However, comparing to natural images, there are fundamental challenges in medical imaging field, especially for segmenting small organs such as the pancreas. First, obtaining high quality

The associate editor coordinating the review of this manuscript and approving it for publication was Essam A. Rashed.

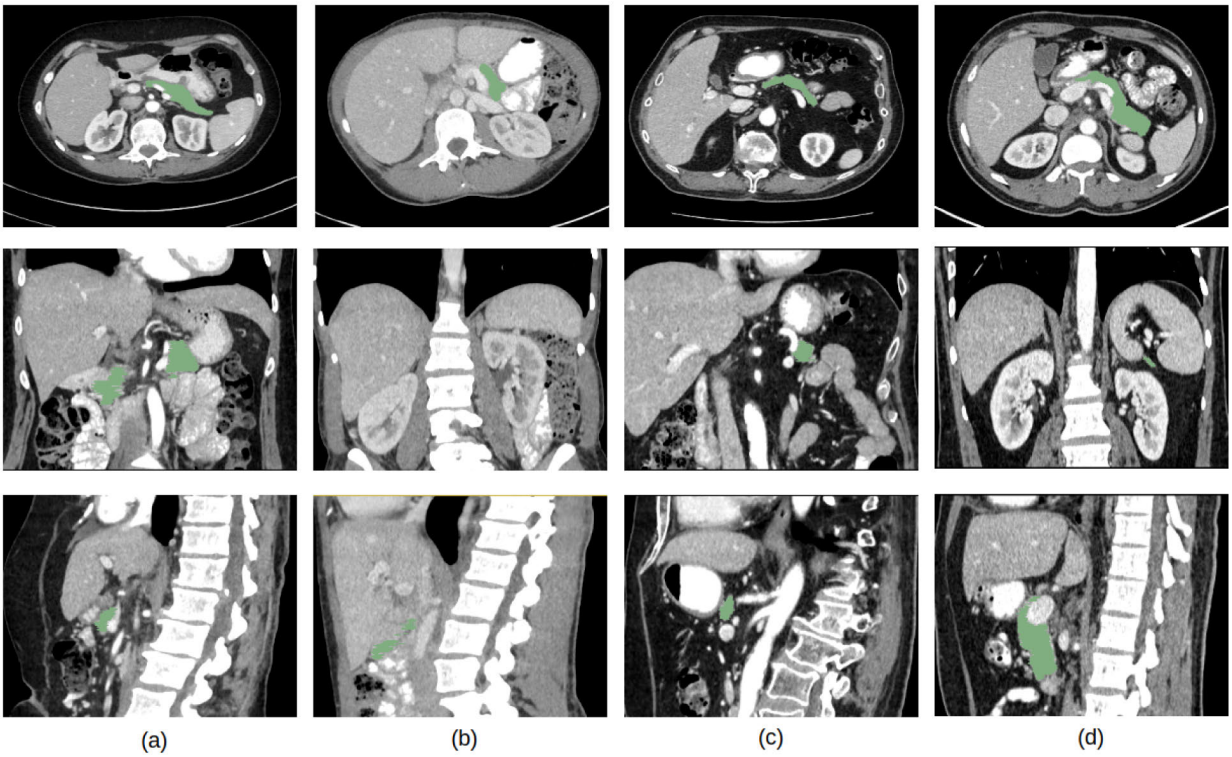

FIGURE 1. An illustration of the challenges for pancreas segmentation from 4 selected patients from the NIH pancreatic dataset. Columns (a) - (d) gives the axial (top row), coronal (middle row) and sagittal (bottom row) views of the 3D CT image, with the annotated pancreas highlighted with green mask. Note that we select the 2D slices at consistent relative positions in the CT volume of each patient so they are visually comparable. We can see that the pancreas has large variations in size, shape and location among these patients; also it occupies only a small part of the whole volume.

annotated data is generally more time consuming, leading to insufficient data for training deep networks. Second, the pancreas occupies only a small portion in the entire image, which inherently causes a class imbalance problem and makes efficient training more difficult. In addition, due to large shape deformation and variations in size and location, it could be tricky to learn robust feature representations which could generalize well to unseen cases. Moreover, due to GPU memory limits, it is generally not realistic to directly use the whole 3D CT image as the network input. So it takes more efforts to design the model that properly captures contextual features of the target and background.

There are recent work trying to solve the above challenges from different ways. The first type of efforts improved robustness in feature learning though elegantly designed network architectures, and thus helped to deal with shape deformation and appearance variations. For example, U-Net [17], [18] and its extensions [19]–[22] proved to be very successful in simultaneously learning and fusing low and high level visual representations. Multi-scale convolution [13], [23], [24] is also a common technique in modeling both local details and global context.

We also found several common techniques that could deal with the data scarcity problem. For 3D image segmentation, one popular approach was to slice the 3D image into 2D slices, train the model and generate predicted masks in 2D, and finally aggregate them back to 3D [10]–[12], [25]–[27]. In the case of learning on 2D slices, we don't need to worry about memory overheads, also the model is easier to train with equivalently more training data. Yet it is not straightforward how to incorporate 3D information among slices.

Localization of the pancreas is another fundamental challenge due to size and position variations. When learning on 2D slices, multi-stage coarse-to-fine based approaches [9], [26] perform localization on the coarse stage, and the fine stage concentrates on the segmentation task. In contrast, a single-stage approach could also be built with attention modules [28] to automatically highlight the target region during end-to-end training.

However, in 3D cases, both types of methods become less flexible due to memory constraints. The work in [28] directly perform 3D learning with spatial attention, but they had to down-sample the input CT images first and use smaller batches for training. It is also difficult to derive a consistent bounding box based on prior knowledge due to large position variations [29].

In 3D cases, we normally perform patch based training and testing. Each patch is a 3D sub-volume with a proper size that fits into GPU memory while covering sufficient contextual information. A fundamental component here is

how patches are generated. A simple approach is to sample uniformly from the entire image domain. Obviously, this could cause a severe class imbalance problem when segmenting small organs. In this case the background class will dominate the training process which tends to generate a naive model that maps everywhere to zero with misleadingly high ''accuracy''. Another way is to sample within a bounding box from approximate localization. However, in addition to the difficulty mentioned above [29], it is still unclear how to guarantee the consistency of patch distributions between training and testing stages.

In this work, we propose a unified Markov Chain Monte Carlo (MCMC) guided CNN learning framework for efficient and robust pancreas segmentation. The target localization is estimated from a prior 3D spatial distribution and an MCMC based scheme is employed to guide both training and testing consistently. We use CNN based architectures here in the segmentation module but any other types of network could also be easily integrated. During testing, the identical MCMC procedure is again used to guide the network to generate patch-wise predictions, which will eventually be fused to form the full segmentation mask. We will show that in this framework the network could learn features of the target and boarder regions sufficiently and won't fall into a trivial model space caused by highly imbalanced dataset. Also, remote regions with similar appearances to the target will not be falsely extracted during inference. We perform extensive experiments on the NIH pancreatic dataset, as well as theoretical discussions how MCMC guided learning helps in variance reductions.

In summary, this work mainly have four-fold contributions as follows:

1) We proposed a unified localization and segmentation framework based on MCMC guided learning. Though we used a 3D U-Net architecture as the segmentation module, this framework can be generalized to using any other networks.
2) We implemented a novel MCMC guided patch selection. It solved the common problem of memory limit, class imbalance and data scarcity in 3D segmentation. It also ensures the consistency of patch distributions between training and testing stages.
3) The framework effectively learns contextual features around the target while preventing false discovery on irrelevant locations. It provided competitive accuracy, efficiency and robustness in pancreas segmentation on the NIH dataset.
4) In addition to extensive experiment, we also provided additional theoretical explanations about why the framework can lead to effective variance reduction.

The remaining the manuscript is organized as follows. Section II reviews related work, and section III describes our proposed framework. We provide experiments with detailed implementation and results in Section IV and finally discuss conclusion and future directions in Section V.

II. RELATED WORKS

We already provided a description of overall context in Section I; now we provide a more detailed review on pancreas segmentation, along with two other related tasks sharing similar challenges: small organ/target segmentation and brain tumor segmentation. Note that our proposed MCMC guided learning framework could be easily generalized to these two tasks as well, regardless of specific network architectures, as far as patch based training and testing is needed.

A. PANCREAS SEGMENTATION

In recent years, more attention has been paid to pancreas segmentation and many algorithms have been proposed. A multi-atlas framework was proposed in [30]. In this work, the region of the pancreas is firstly extracted by the relative position and structure of the pancreas and liver. Then, using the vessel structure around the pancreas, images from training dataset are registered to the image to be segmented. Then the best registration is chosen according to the vessel structure similarity. This work reported a dice similarity coefficient (DSC) of 78.5±14.0%.

To address the issue of low contrast, more advanced learning based approaches are used. Unlike the top-down approach based on multi-atlas [30], researchers in [10], [27] proposed a bottom-up strategy that decomposes all 2D slices of a patient into boundary-preserving superpixels through over-segmentation. Then, it extracts superpixel-level features and built a cascaded random forest classifier to classify superpixels as pancreas and non-pancreas regions. Comparing with [30], these methods have less data requirements, but the results have a slightly lower DSC of 68.8±25.6% in [27] and 70.7±13.0% in [10].

Similar to [10], the methods proposed by [9], [11], [12] combine random forest and deep CNN. In [9], authors presented a coarse-to-fine approach in which multi-level CNN is employed on both image patches and regions. In this approach, an initial set of superpixel regions are generated from the input CT images by a coarse cascade process of random forests based on [27]. Serving as regional candidates, these superpixel regions possess high sensitivity but low precision. Next, the trained CNN are used to classify superpixel regions as pancreas and non-pancreas. 3D Gaussian smoothing and 2D conditional random fields are used for post-processing finally. Different from [9], researchers in [11] proposed using random forest to classify superpixels similar to [27] and [10]. But the superpixels and features are generated via Holistically-Nested Networks, which extract the pancreas' interior and boundary mid-level cues. Before this step, this model gets the region of interest (ROI) by the method in [27]. Based on [11], authors in [12] made further improvements where the ROI is estimated by a new deep network. In [12], the algorithm learns mid-level cues via Holistically-Nested Networks firstly. Then, it obtains the ROI by a multi-view aggregated Holistically-Nested Networks and the largest connected component analysis. Finally,

the random forest classification is used again to classify superpixels.

There have also been some purely CNN based methods. For example, a fixed-point model that shrinks the input region by the predicted segmentation mask was proposed in [7]. While the parameters of network remain unchanged, the regions are optimized by an iterative process. In contrast to [7], researchers in [8] took more efforts in architecture optimization; it proposed an extension to the Richer Feature Convolutional Network which adopted multi-layer upsampling to capture multi-scale contextual information.

In contrast to building separate steps for localization, several more recent deep learning based approaches tend to integrate localization as a sub-module of unified optimization frameworks. A lightweight DCNN based approach was proposed in [31] for accurate segmentation with low computational cost. Lightweight network blocks could greatly reduce the number of parameters and mitigate over-fitting problems. This approach integrated both localization and segmentation sub-networks, where spatial priors and deep feature maps were used together to guide segmentation predictions. Also, Attention U-Net was introduced in [28] where the spatial attention module could automatically focus on salient features related to target regions while suppressing irrelevant locations. Recently, a deep reinforcement learning based approach was also proposed [32] to drive dynamic pancreas localization with its contextual interaction mechanism. It also effectively dealt with shape deformation by building a deformable version of deep U-Net.

Comparing to the majority of existing work that built their model either from 2D slices or on 3D volumes, a new multi-stage coarse-to-fine framework proposed in [33] incorporated features learned by both 3D and 2D CNNs. This method first generated a coarse segmentation through multi-atlas, followed by a fusion of 3D and 2D CNNs to produce a finer segmentation, and finally adopted 3D level set algorithm for further refinement. This method combined location, contextual, shape and edge features all together to achieve a new state of the art result for pancreas segmentation.

Another type of recent efforts focused on solving data scarcity and class imbalance issues through more generic learning techniques such as weighted loss function and data augmentation. For example, a weighted combination of DICE and Binary Cross Entropy loss functions was explored in [29] that improved overall learning capability and segmentation results. Multiple data augmentation strategies, including mixup [34] and RICAP [35] were studied in [36], leading to consistent improvement on pancreas segmentation for various U-Net architectures.

B. SEGMENTING SMALL/Texture-LESS TARGETS

Since the small size and shape variations are major causes for difficulties in pancreas segmentation, we also review some important work for detecting and segmenting small and texture-less targets in general.

First, the most relevant line of research is small organ segmentation. A two-stage scheme for small organ segmentation in whole-body MRI was proposed in [37]. Originated from coarse-to-fine frameworks, this approach adopted weighting and auto-context with spatial priors to solve class imbalance problems for small organ localization. Similarly, recurrent saliency transformation network [26] was also built upon conventional coarse-to-fine frameworks. But it enabled joint optimization and feature fusion over both coarse and fine stages through its saliency transformation module and iterative learning of spatial information. This approach showed improvements for general small organ segmentation including the pancreas. More recently, researchers in [38] proposed FocusNetV2 that could simultaneously segment large and small organs in head and neck CT images. This is also a multi-stage framework including separate steps for segmenting large and small organs and a localization step for small organs. Adversarial autoencoder is also firstly used in this work as shape constraints to improve localization and segmentation. Another framework, CentroidNetV2 [39] offered an elegant solution for small object segmentation and counting. Though the paper [39] did not study organ segmentation tasks, it could potentially be helpful when we need to extract multiple small organs from the same image.

Second, there are also recent advances in small target and texture-less segmentation in infrared and natural images. Researchers in [40] proposed a novel optimization method that formulates the infrared small target detection problem into sparse matrix reconstruction. They adopted overlapping edge information to enhance detection accuracy, and used self-regularization to mine background information. To address the issue of low contrast, small size and texture-less nature, the authors in [41] proposed a new deep neural network based approach, with the novel asymmetric contextual modulation mechanism that could effectively exchange high-level semantic and low-level fine details for small and texture-less target detection.

Though sharing common challenges, the context for our work has important distinctions to small/texture-less target segmentation. First, unlike infrared small target, the pancreas in CT images still carries important texture and shape information, and thus conventional CNN encoding layers are still important for feature extraction. Second, a popular assumption for the case of infrared small target [40], [42], [43] is that the background is low rank in nature which contains a large number of repeated elements, but this is not true for CT images. For pancreas segmentation, anywhere located outside of the pancreatic region is treated as ''background'', which could contain rich contextual information and large variations. Therefore, carefully designed training scheme is needed to make sure the model learns sufficiently about the texture and shape information of the target while not being overwhelmed by the background.

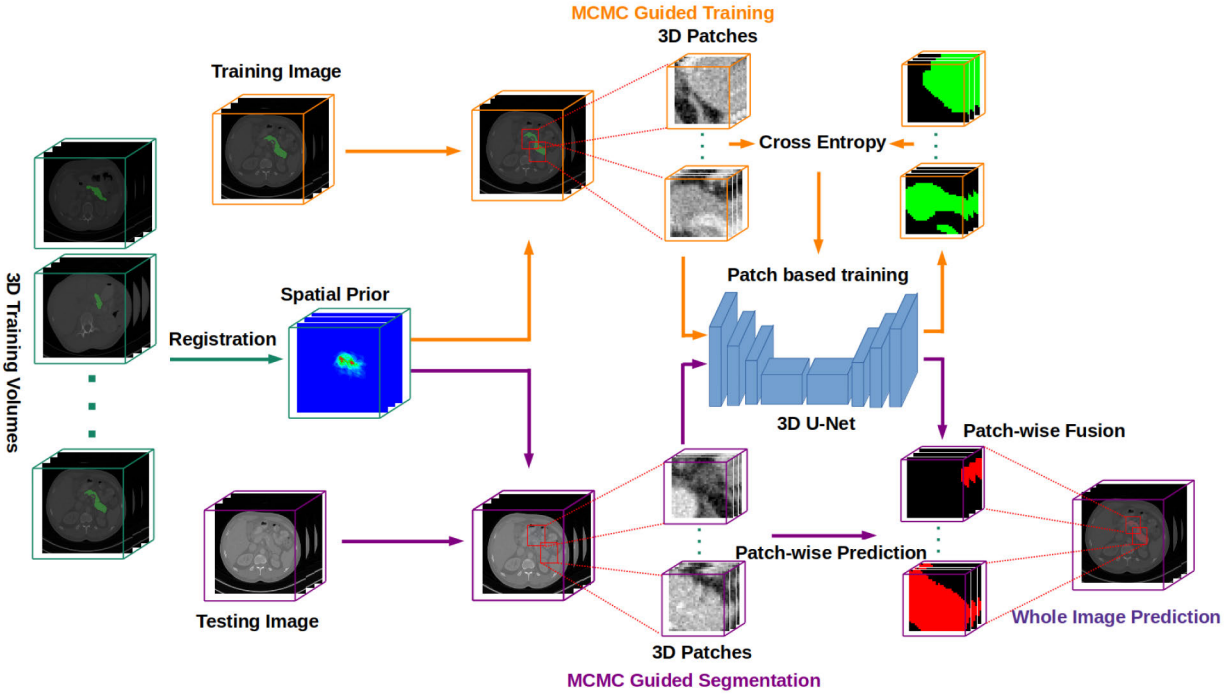

FIGURE 2. The MCMC guided learning framework. We first obtain a spatial prior through multi-atlas based registration step, then an MCMC process is constructed to guide the patch based CNN training to adaptively focus more on the target regions. During testing, the same MCMC scheme is used to guide segmentation with trained CNN, to generate patch-wise predictions which are then fused to provide the whole image mask.

C. BRAIN TUMOR SEGMENTATION

Similar to the pancreas, brain tumors could also be located anywhere in the brain with highly variable shape, size and contrast [24]. But different from our work for CT images, effective multi-modal learning is a core consideration due to the nature of MRI data. Researches in [24] proposed a deep neural network based framework, with comprehensive study on optimized network architecture. They exploited both local and global features simultaneously thorough a two-pathway architecture and adopted a two-phase training scheme to deal with class imbalance effectively. Another recent work [44] instead focused on the optimizing cross-modality knowledge transfer and feature fusion and reached a new state-of-the-art for brain tumor segmentation in MRI images.

III. METHODOLOGY

In this section, we describe the proposed MCMC guided CNN learning framework for pancreas segmentation in greater detail. We provide the overall context in Section III-A and then discuss how our MCMC scheme guides both training (Section III-B) and inference and explain theoretically why this framework works in Section III-C

Figure 2 gives an overview of our entire framework. We first obtain a prior spatial distribution that serves as a reasonable initial estimation of the target location, then an MCMC process could be constructed to guide the CNN training to focus more on the key regions, as well as the border area around the target. The same process again works together with the trained CNN to segment new images through patch-wise fusing. This iterative learning process adaptively concentrates on regions with higher likelihood of finding the target while effectively avoiding sampling from irrelevant regions, which eventually leads to an accurate and stable segmentation.

A. JOINT LEARNING OF APPEARANCE AND LOCATION

Denote the gray-scale 3D image to be segmented as $I : \Omega \rightarrow \mathbb{R}$ where $\Omega \subset \mathbb{R}^3$ is the domain where the image is defined. The segmentation of I is seeking for an indicator function $J : \Omega \rightarrow \{0, 1\}$ whose 0 valued pixels indicate the background and the 1s indicate the pancreas. Such a characteristic function J can be viewed (up to a constant multiple) as a special case of a probability density function (pdf) $p : \Omega \rightarrow [0, 1]$ where the value $p(\mathbf{x})$ is a likelihood that the pixel $\mathbf{x}$ being inside the target.

Specifically, we are given a set of training images $I_i : \Omega \rightarrow \mathbb{R}, i = 1, \ldots, M$ with their pancreas segmentation masks as $J_i : \Omega \rightarrow \{0, 1\}, i = 1, \ldots, M$. Our task is to learn a mapping F such that it gives an estimate of the mask J given any image I: $F(I) = \hat{P}[J|I] = \hat{P}[J(\mathbf{y}) = 1|I]|_{\mathbf{y}\in\Omega} \approx P[J|I] = J$. The training set provides us the spatial and contextual information of the pancreas. As explained in Sections I and II, a joint learning of location and appearance features is fundamental for segmentation and we use patch based learning in this work.

Formally, for an image $I : \Omega \to \mathbb{R}$ to be segmented, for any $\mathbf{x} \in \Omega$, we define its local neighborhood of radius h centered around $\mathbf{x}$ as $N(\mathbf{x}) := \{\mathbf{y} \in \Omega : \|\mathbf{x} - \mathbf{y}\|_\infty \leq h\}$; we also restrict $I(\cdot)$ on $N(\mathbf{x})$ to define the "patch" centered at $\mathbf{x}$ as $L_\mathbf{x}(\cdot) : N(\mathbf{x}) \to \mathbb{R}$ with $L_\mathbf{x}(\mathbf{y}) := I(\mathbf{y})|_{N(\mathbf{x})}$. To make the following discussions more straightforward, we also define a "conjugate neighborhood" for any $\mathbf{y} \in \Omega$: $\bar{N}(\mathbf{y}) := \{\mathbf{x}|\mathbf{y} \in N(\mathbf{x})\}$, which is simply a set of all centers whose neighborhood covers $\mathbf{y}$.

Then, the joint distribution of $(L_\mathbf{x}, \mathbf{x})$ ultimately provides the information to infer the segmentation map for I, as in equation 1 below:

$$P[J(\mathbf{y})=1|I]=\int_{\mathbf{x}\in\bar{N}(\mathbf{y})} P[J(\mathbf{y})=1|\mathbf{y}\in N(\mathbf{x})]P[\mathbf{y}\in N(\mathbf{x})]d\mathbf{x} \tag{1}$$

In the right hand side equation 1 and throughout all following discussions, we omit the conditional term $P[\cdot|I]$ in probability notations for simplicity. The first term inside the integral part of equation 1 depicts the segmentation map on the selected patch on $N(\mathbf{x})$ which could be estimated by a deep network applied on $L_\mathbf{x}$, as in equation 2 below:

$$P[J(\mathbf{y}) = 1|\mathbf{y} \in N(\mathbf{x})] \approx \hat{P}[J(\mathbf{y}) = 1|L_\mathbf{x}] \tag{2}$$

The second term $P[\mathbf{y} \in N(\mathbf{x})]$ indicates the likelihood of "selecting the neighborhood" $N(\mathbf{x})$ covering $\mathbf{y}$. If we have some global spatial distribution $\tilde{p}(\mathbf{x})$ determining the likelihood of selecting neighborhood centering around $\mathbf{x}$, then simply

$$P[\mathbf{y} \in N(\mathbf{x})] = \tilde{p}(\mathbf{x})|_{\bar{N}(\mathbf{y})} := \frac{\tilde{p}(\mathbf{x})}{\int_{\bar{N}(\mathbf{y})} \tilde{p}(\mathbf{x})d\mathbf{x}} \tag{3}$$

Thus we now have an estimate for the segmentation mask J based on equations 1, 2 and 3 expressed eventually as a conditional expectation in equation 4:

$$\begin{aligned}\hat{P}[J(\mathbf{y}) = 1] &= \int_{\mathbf{x}\in\bar{N}(\mathbf{y})} \hat{P}[J[\mathbf{y}] = 1|L_\mathbf{x}]\tilde{p}(\mathbf{x})|_{\bar{N}(\mathbf{y})}d\mathbf{x} \\ &= E_{X\sim\tilde{p}(\mathbf{x})|_{\bar{N}(\mathbf{y})}}[\hat{P}[J[\mathbf{y}] = 1|L_X]]\end{aligned} \tag{4}$$

In equation 4, the conditional probability part $\hat{P}[J[\mathbf{y}] = 1|L_X]$ can be learned through our MCMC guided CNN framework detailed in III-B and III-C. In addition, a prior estimated of the distribution $\tilde{p}$, denoted as p^0, can be derived from the registration process as we will describe below.

It is realized that the training set has a large variance on the shape and size of the pancreas. Though ideally all such variations could be automatically captured through deep neural networks, normalization registration still proved to be helpful in reducing the variance and thus leaves less burden on the training steps.

Indeed, before the emerging of the convolutional neural networks, multi-atlas is a robust algorithm that addresses the problem of medical image segmentation by (multiple) registrations. It achieved very high segmentation accuracy, especially for the brain structures [45]–[50]. Unfortunately, comparing to brain segmentation, the performance on other sites is much worse.

Such discrepancy is understandable since the shape, size, and appearance in abdominal images carry much more variations than those in the brain image. Also, as a result, non-linear registration performs worse on abdominal images.

Based on this rationale, in this study we only use affine registration among the images to mitigate the training variance, leaving the rest to the machine learning framework.

We first randomly pick one training image $I_i : i = 1, \ldots, M$ as the "anchor" image $\tilde{I}$, with its mask $\tilde{J}$ for registration. Then for each $i \in \{1, \ldots, M\}$, we have the registration computed in equation 5 below:

$$T_i^* = \text{argmin}_{T:\Omega\to\Omega} D(\tilde{I}, I_i \circ T) \tag{5}$$

where $T : \Omega \to \Omega$ is a family of affine transformations: $T(\mathbf{x}) = A\mathbf{x} + b$ with $A \in \mathbb{R}^{3\times3}, b \in \mathbb{R}^3$. $D(I_i, I_j)$ denotes a suitable dis-similarity functional between the two images I_i and I_j. The optimal registration transformations T_i^* can be computed through regular gradient or Newton based procedures. As a result, $\tilde{I} \approx I_i \circ T_i^* =: \tilde{I}_i$ and $\tilde{J} \approx J_i \circ T_i^* =: \tilde{J}_i$, $\forall i$.

Once registered to a common space, the collection of $\tilde{J}_i$ represent the spatial distribution of the pancreas in the training data and therefore we can compute the prior $p^0 : \Omega \to [0, 1]$ as:

$$p^0(\mathbf{y}) = \sum_{i=1}^{M} \tilde{J}_i(\mathbf{y}) \Big/ \int_\mathbf{x} \sum_{i=1}^{M} \tilde{J}_i(\mathbf{x})d\mathbf{x} \tag{6}$$

In the following, we will describe in detail the MCMC guided training and testing framework.

B. MCMC GUIDED TRAINING WITH 3D CNN

As explained in Section I, one important challenge for patch based learning is how patches are selected from the entire image domain. If taken uniformly, then a large portion of the patches are likely to be just "background", with their labels as all-zero masks. In practice, learning from such empty masks does more harms than wasting time. Indeed, it encourages the model to learn towards a trivial global optimal solution that maps every input to zero, especially when the area of the object takes a small portion of the entire image domain. Many researchers introduced extra mechanisms, such as using a bounding box to limit the volume from where patches could be generated. However, how to determine such a bounding box during testing again poses new problems.

Therefore, on one hand, we want the training patches to contain a certain amount of background patches to learn negative appearances. On the other hand, we don't want too many of them to steer the learning towards the trivial global optimal.

To address this challenge, we could use the prior distribution $p^0(\mathbf{x})$ defined in Eq 6 to guide patch generation during training. Since the higher values in $p^0(\mathbf{x})$ indicate being more

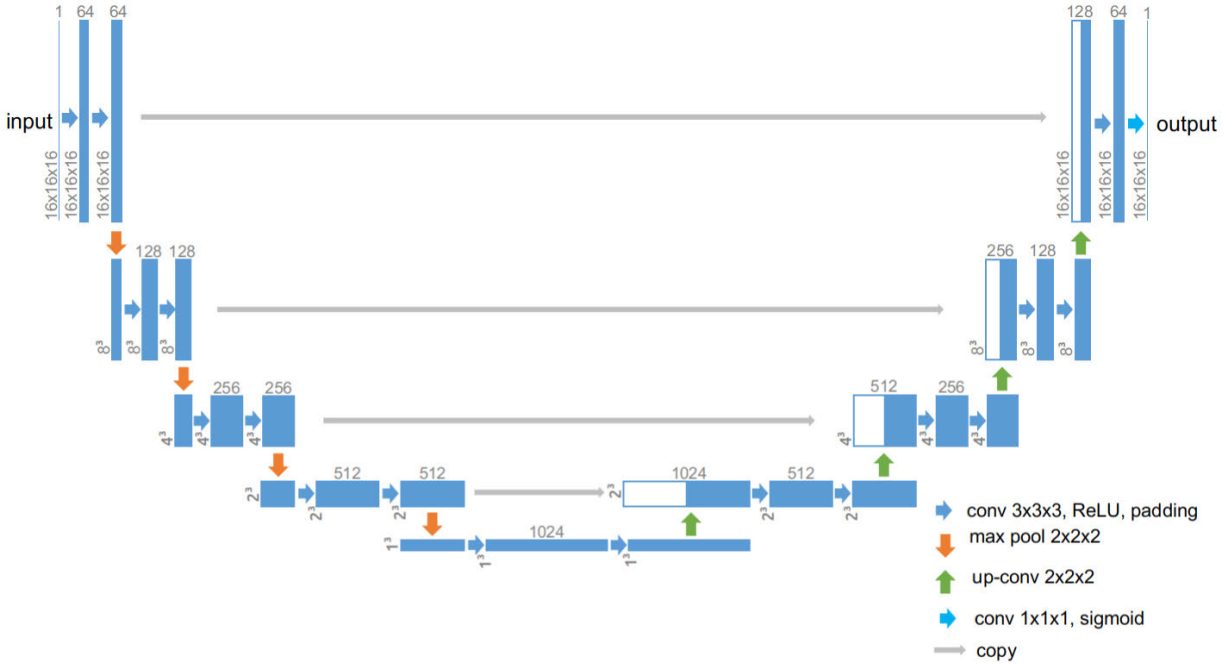

FIGURE 3. The architecture of 3D U-Net used in our work based on [17].

likely to be inside the pancreas, the image appearances in those regions are generally more typical for pancreas. Therefore, during training, more emphasis should be put on those regions. Once we can draw a set of sample points $S = \{\mathbf{s}_i \in \Omega | \mathbf{s}_i \sim p^0(\mathbf{x})\}$ according to p^0, then we can feed generated patches $\{L_{\mathbf{s}_i}\}$ CNN training.

The arbitrariness of p^0 poses a new challenge since regular sampling schemes (such as for a Gaussian distribution) could not be applied. Therefore, in this work we integrate the Markov chain Monte Carlo (MCMC) algorithm with CNN training.

MCMC is an class of algorithms that can iteratively generate samples to approximate some (usually multi-dimensional) theoretical distribution [51]. First, we construct a markov chain based on the theoretical distribution treated as the stationary probability distribution; then we iteratively draw samples from this markov chain and eventually we can get the target samples when the markov chain approaches its stationary distribution. Specifically, the Metropolis-Hastings (MH) algorithm [52] is used in this work to obtain random samples from p^0 which is hard to sample directly. The main steps of MH guided patch generation are detailed in Algorithm 1. Accordingly, given arbitrary image I and spatial distribution p, we set a candidate kernel q for the markov chain along with $n_1 > 0$ and $n_2 > 0$ as the state transition steps and sampling steps respectively in MCMC. Then we initialize empty queues S and $\mathcal{L}$ to store generated seeds and their corresponding patches from the algorithm. Then, starting from an initial state $\mathbf{x}_0$, we proceed with the customised MH iterations to sequentially enqueue the candidate patches.

We apply Algorithm 1 consistently for each of the training image $I_j, j = 1, \ldots, M$, and obtain set of patch seeds $S^j := \{\mathbf{s}_i^j | i = 1, \ldots, |S^j|\}$ along with the 3D patches around each of the seeds $\mathbf{s}_i^j$ as shown in equation 7 below:

$$L^j_{\mathbf{s}_i^j} := I_j(\mathbf{x})|_{\mathbf{x} \in N(\mathbf{s}_i^j)}, j \in \{1, \ldots, M\}, i \in \{1, \ldots, |S^j|\} \quad (7)$$

These 3D patches are then set as the input for the 3D CNN to learn.

Various convolutional neural network architectures have been proposed in the recent years [17], [18], [53]–[56]. Among them, the U-Net is a CNN framework for segmentation [18] which have achieved substantial success particularly for biomedical images. The success of U-Net is built upon its encoder-decoder structure that captures both low-level and high-level information in an efficient way.

In our work, we build our learning framework based on 3D U-Net in [17], which demonstrated its advantage comparing to 2D version, in capturing joint information from adjacent slices and generating more coherent segmentation predictions. The detailed architecture of our 3D U-Net is shown in Figure 3

Moreover, it is worth mentioning that since the focus of this work is to demonstrate the benefit of MCMC guided learning framework, though several other variants were proposed in recent years, our network architecture is based on the vanilla U-Net design. As mentioned in Section I, it is certainly possible to synergize our framework with extended versions of U-Net or other popular architectures for segmentation.

Algorithm 1 Metropolis-Hastings Guided Patch Generation

Input: $I, p, q, n1, n2$
Initialize: $S = \{\}$, $\mathcal{L} = \{\}$, initial state $\mathbf{x}_0 \in \Omega$
for $t = 0, 1, 2, \ldots, n_1 + n_2 - 1$ **do**
  Generate a proposal state $\mathbf{x}^*$ from $q(\mathbf{x}|\mathbf{x}_t)$
  Draw a random number $u \sim$ Uniform(0, 1)
  Calculate the acceptance probability

$$\alpha(\mathbf{x}_t, \mathbf{x}^*) \leftarrow min\{\frac{p(\mathbf{x}^*)q(\mathbf{x}^*|\mathbf{x}_t)}{p(\mathbf{x}_t)q(\mathbf{x}_t|\mathbf{x}^*)}, 1\}$$

  if $u < \alpha(\mathbf{x}_t, \mathbf{x}^*)$ **then**
    Set $\mathbf{x}_{t+1} \leftarrow \mathbf{x}^*$
    if $t > n_1 - 1$ **then**
      $S \leftarrow S \cup \{\mathbf{x}_{t+1}\}$
      $\mathcal{L} \leftarrow \mathcal{L} \cup \{L_{\mathbf{x}_{t+1}}\}$
    else
      Continue
    end if
  else
    $\mathbf{x}_{t+1} = \mathbf{x}_t$
  end if
end for
return $S, \mathcal{L}$

Once we have a trained network, we will use the same MCMC process to guide segmentation during inference, as described in the next subsection.

C. MCMC GUIDED SEGMENTATION

Given a new image I_t to be segmented, it is first registered to the same common space defined during training. In particular, we retrieve $\tilde{I}$ used in equation 5 from training stage and solve

$$T_t^* = \text{argmin}_{T:\Omega\to\Omega} D(\tilde{I}, I_t \circ T) \tag{8}$$

Likewise, in equation 8, $T : \Omega \to \Omega$ is a family of affine transformations and T_t^* is the optimal transformation for I_t. Then we have the registered image $\tilde{I}_t := I_t \circ T_t^*$. Also, let $\tilde{J}_t$ represent the transformed "ground-truth mask" J_t for I_t which is unknown at the moment.: $\tilde{J}_t := J_t \circ T_t^*$.

Next, we apply the same MCMC scheme from algorithm 1 on the registered image $\tilde{I}_t$ and then obtain a set of generated patches $S_t := \{L_{\mathbf{s}_i^t} | i = 1, \ldots, |S^t|\}$. It is obvious that the same spatial distribution p^0 from III-B is again used here. Therefore, according to equation 4, The estimation for the global mask $\tilde{J}_t$ is then obtained as in equation 9 below:

$$\hat{P}[\tilde{J}_t(\mathbf{y}) = 1] = E_{X\sim p^0(\mathbf{x})|_{\bar{N}(\mathbf{y})}}[\hat{P}[\tilde{J}_t[\mathbf{y}] = 1|L_X]] \tag{9}$$

Now let NET denote the trained network from Section III-B, then we can have the "predicted mask" on an arbitrary patch L_X, expressed as the estimated conditional probability as: $\hat{P}[\tilde{J}_t(\mathbf{y}) = 1|L_X] = \text{NET}(L_X)$. Therefore, we have equation 10 below which serves as the principle for patch-wise prediction fusion, that aggregates model predictions on each patch into the predicted mask on whole image.

$$\hat{P}[\tilde{J}_t(\mathbf{y}) = 1] = E_{X\sim p^0(\mathbf{x})|_{\bar{N}(\mathbf{y})}}[\text{NET}(L_X)] \tag{10}$$

It is obvious that due to MCMC the points in S_t are sampled from p^0, therefore inside any local domain $\bar{N}(\mathbf{y})$, the samples $S_t(\mathbf{y}) := \{\mathbf{s} \in S_t | \mathbf{s} \in \bar{N}(\mathbf{y})\}$ will follow the local distribution $p^0(\mathbf{x})|_{\bar{N}(\mathbf{y})}$. Now we have an estimate for the expectation term in equation 10:

$$\begin{aligned}\hat{P}[\tilde{J}_t(\mathbf{y}) = 1] &\approx \frac{1}{|S_t(\mathbf{y})|}\sum_{\mathbf{s}\in S_t(\mathbf{y})} \hat{P}[\tilde{J}_t(\mathbf{y}) = 1|L_\mathbf{s}] \\ &= \frac{1}{|S_t(\mathbf{y})|}\sum_{\mathbf{s}\in S_t(\mathbf{y})} \text{NET}(L_\mathbf{s})\end{aligned} \tag{11}$$

Note that equation 11 is generally true not just for the particular case in equation 10, but for any arbitrary distributions for $\mathbf{s}$.

We can therefore summarize the steps of MCMC guided segmentation in algorithm 2 below. Accordingly, given I_t, a new image to be segmented, we retrieve the "anchor" image $\tilde{I}$ from training stage and search and perform transformation from equation 8. After that, we apply algorithm 1 to generate seeds S_t and patches $\mathcal{L}_t$, and then apply the trained network NET on each of the patches followed by aggregation according to equation 11. Eventually, we obtain the whole image prediction by inverting the registration transform.

Algorithm 2 MCMC Guided Segmentation

Input: $I_t, \tilde{I}$, NET
Solve: $T_t^* \leftarrow \text{argmin}_{T:\Omega\to\Omega} D(\tilde{I}, I_t \circ T)$
Set: $\tilde{I}_t \leftarrow I_t \circ T_t^*$
Execute: Algorithm 1 on $\tilde{I}_t$ and obtain $(S_t, \mathcal{L}_t)$
Initialize: $W(\mathbf{x}) = 0$ and $K(\mathbf{x}) = 0\ \forall \mathbf{x} \in \Omega$
for $\mathbf{s}$ in S_t **do**
  for $\mathbf{x}$ in $N(\mathbf{s})$ **do**
    Set $W(\mathbf{x}) \leftarrow W(\mathbf{x}) + \text{NET}(L_\mathbf{x})$
    Set $K(\mathbf{x}) \leftarrow K(\mathbf{x}) + 1$
  end for
end for
Set $W(\mathbf{x}) \leftarrow W(\mathbf{x})/K(\mathbf{x}),\ \forall \mathbf{x} \in \bigcup_{\mathbf{s}\in S_t} N(\mathbf{s})$
return $W \circ [T_t^*]^{-1}$

We now provide further explanations about the rationale behind MCMC guided learning and segmentation. First, as mentioned before, higher values in the spatial prior roughly correspond to locations near or inside target region. Since the pancreas, in particular, has higher variations in appearance, size and location comparing to other organs located in the background region. We can naturally assume that the pixel values inside or around the target region have larger variance. Second, the learned model itself would be more sensitive near the target region; extracting the interior and boundary of the target is a much more difficult than predicting the background.

Let $\mathbf{y} \in \Omega$ be an arbitrary point in the image domain and $X \in \Omega$ be a random variable uniformly distributed inside the local conjugate neighborhood of $\mathbf{y} \in \Omega$: $X \sim \mathrm{U}(\bar{N}(\mathbf{y}))$, then the above argument can be expressed as equation 12 below.

$$\mathrm{VAR}_{X\sim \mathrm{U}|_{\bar{N}(\mathbf{y})}}[\hat{P}[\tilde{J}_t(\mathbf{y}) = 1|L_X]] \propto f(p(\mathbf{y})) \tag{12}$$

where $f : [0, 1] \to \mathbb{R}^+$ is just some existing non-decreasing functional to preserve "positive correlation" for simplicity; and p is a spatial prior representing the likelihood to be inside the target region.

In fact, since our patch size, and thus $\bar{N}(\mathbf{y})$ is small comparing to the entire image domain, we can assume locally i.i.d sampling and use uniform distribution $U|_{\bar{N}(\mathbf{y})}$ to approximate $p|_{\bar{N}(\mathbf{y})}$, so the estimator in equation 11 has the variance:

$$\begin{aligned}\mathrm{VAR}[\hat{P}[\tilde{J}_t(\mathbf{y}) = 1]] &\approx \frac{\mathrm{VAR}_{X\sim \mathrm{U}|_{\bar{N}(\mathbf{y})}}[\hat{P}[\tilde{J}_t(\mathbf{y}) = 1|L_X]]}{|S(\mathbf{y})|} \\ &\propto \frac{f(p(\mathbf{y}))}{|S(\mathbf{y})|}\end{aligned} \tag{13}$$

It is obvious from equation 13 that if $p(\mathbf{y})$ is large, i.e. $\mathbf{y}$ is located near the target region, then we need more sample points around this point to generate an accurate and robust prediction. On the other hand, in background regions where $p(\mathbf{y})$ is small, we don't need that many samples and uniform sampling would cause a waste of resources. Therefore, MCMC guided learning and segmentation provides effective variance reduction while saving computational resources and inference time.

IV. IMPLEMENTATION, EXPERIMENTS AND RESULTS

In this section, we provide detailed algorithm implementation, experiment settings and results.

A. IMPLEMENTATION

In this work, we use the DeedsBCV library [57] to perform image registration. As discussed in III-A, we apply the affine registration that takes about one minutes for one task. The registered training mask images thus forms the spatial prior p^0 according to Eq 6.

The shape of pancreas varies greatly among different people. So that the segmentation might not be sufficiently accurate by using single moving image to register. In order to improve the accuracy, we try to use multiple moving images to register to the fixed image from the multi-atlas idea. The obtained prior image p^0 can then be threshold-ed to form a binary image, as a rough segmentation of pancreas, with a threshold d:

$$R(\mathbf{x}) = \begin{cases} 0, & p^0(\mathbf{x}) < d/100 \\ 1, & p^0(\mathbf{x}) \geq d/100 \end{cases} \tag{14}$$

The positive and negative regions should be identified by the 3D U-Net. But pancreas is surrounded by many other organs so that the peripancreatic morphology might be variable, making the classification task more difficult. What's more, pancreas is small in abdomen, so true negatives are far more than the others, and the most of these true negatives carry no information about the features of pancreas and its border, so the input of network might not adequately represent the morphological features of the pancreas. The proposed MCMC guided learning (algorithm 1) directly addresses this issue by focusing on the target region according to p^0.

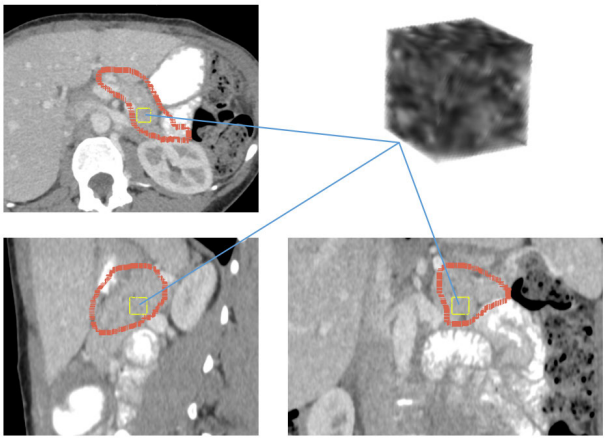

FIGURE 4. 3D patches drawn from the image.

In the experiment, we found out that one additional pre-processing step could further increase the algorithm efficiency while covering the entire pancreatic region. Once d is determined, we expand the pancreatic region in the image R by k pixels to form a new image:

$$E = \{\mathbf{x} \in \Omega | \|\mathbf{x} - \mathbf{y}\|_\infty <= k, \forall \mathbf{y} \text{ s.t. } R(\mathbf{y}) = 1\} \tag{15}$$

The MCMC scheme is then applied on $E \subset \Omega$ which in practice led to faster convergence and more effective learning on the morphological and edge features of the pancreas. We use the 3D U-Net architecture given in Figure 3 based on the original design in [17]. The 3D patches extracted from the image have a size of $16 \times 16 \times 16$, depicted in Figure 4.

In the experiments, we evaluate the results by Precision, Recall, dice similarity coefficient (DSC), and Jaccard similarity coefficient. Let tp, fn, fp, and tn represent the number of true positives, false negatives, false positives, and true negatives, respectively. Some commonly used values are listed below:

Precision is the proportion of positives correctly predicted among all positives predicted in prediction image.

$$Precision = \frac{tp}{tp + fp} \tag{16}$$

Recall is the proportion of positives correctly predicted among all positives in ground truth.

$$Recall = \frac{tp}{tp + fn} \tag{17}$$

Dice similarity coefficient(DSC) is a statistic used to measure the similarity of prediction image and ground truth.

$$DSC = \frac{2 \times tp}{(tp + fp) + (tp + fn)} \tag{18}$$

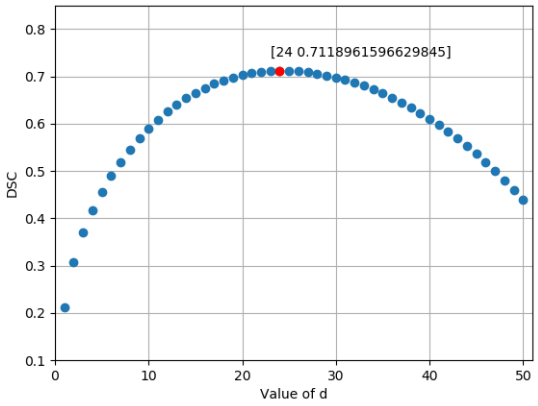

FIGURE 5. the average DSC changes with different values of d.

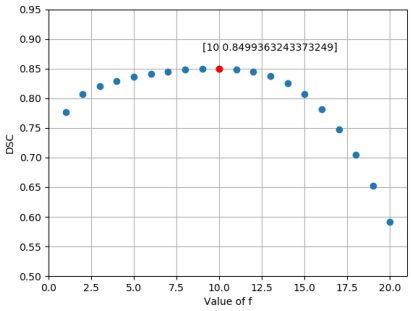

FIGURE 6. The average DSC changes with different values of f.

Jaccard similarity coefficient is a statistic used to measure the similarity and diversity of prediction image and ground truth.

$$Jaccard = \frac{tp}{(tp + fp) + (tp + fn) - tp} \tag{19}$$

B. EXPERIMENTS CONFIGURATION

1) DATASET

To facilitate the comparison of results across different publications, we use the dataset provided by NIH [9], [58], [59]. The dataset contains 82 abdominal contrast enhanced 3D CT images and has been manually labeled the segmentations of pancreas as ground-truth slice-by-slice. Among them, 72 are picked for training and the remaining 10 are used testing.

2) PARAMETER OPTIMIZATION

After registration, we get the prior distribution p^0. In order to get the binary image R, we set a threshold d to classify pixel value as 0 and 1. To determine the value of d, an image is randomly selected and its pixels are classified into 0 and 1 by $d \in [0, 72]$. We set d's value in $\{1, \ldots, 50\}$, then compute the average DSC of different values of d. The result is shown in Figure 5. It is found that when the value of d is 24, the DSC between R and ground truth is maximum, so we set threshold d as 24.

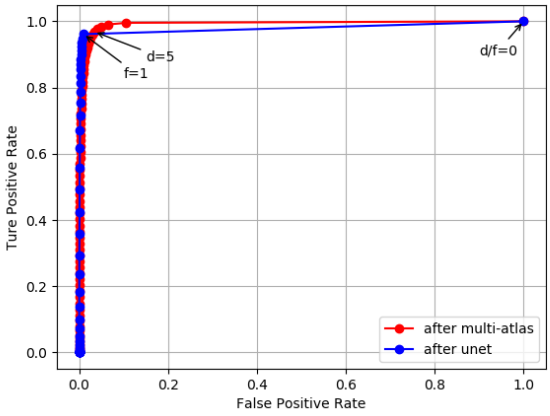

FIGURE 7. ROC curves with varying values of d and f. Red curve is from the prediction of multi-atlas step, blue curve is from the prediction of our (MCMC-)U-Net method.).

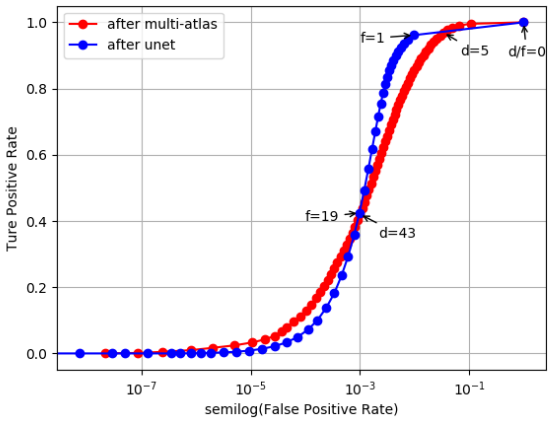

FIGURE 8. ROC curves with semilog false positive rate with varying values of d and f. Red curve is from the prediction of multi-atlas step, blue curve is from the prediction of our (MCMC-)U-Net method.).

In the next step, we expand the pancreatic region in the image R by k pixels. It is found that when $k = 5$, the candidate region contains most of the pancreatic region and the non-pancreatic region is also in a suitable range. So we set $k = 5$.

As for training the 3D U-Net, we set the patch size as $16 \times 16 \times 16$, and set batch size as 100. In our experiment, we use the binary cross entropy as loss function. During segmentation, we search the threshold f in $\{1, \ldots, 20\}$ for the best average DSC. The result is shown in Figure 6. As can be seen, the average DSC of training data is the highest when $f = 10$.

C. EXPERIMENTAL RESULTS

1) PANCREAS SEGMENTATION

In this work, we proposed a pancreas segmentation method based on MCMC guided deep learning. This framework effectively reduces the burden of network training, as well

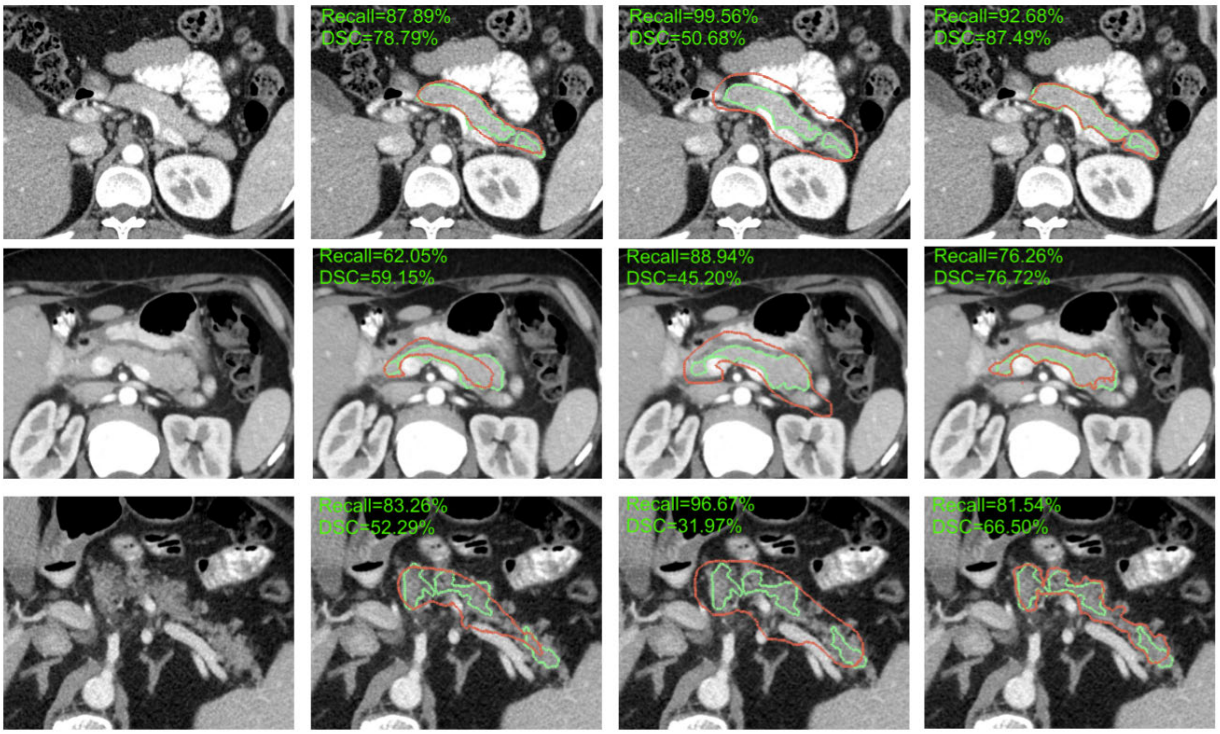

FIGURE 9. Displays predictions and their performances on three examples from test dataset. Ground truths are labeled as green curve. Predictions (red curve) are generated from 1) the multi-atlas step, 2) the selected candidate region and 3) our learning framework. From top to bottom, the first row displays the case where our framework gives the best performances; the second row shows the case with average performance while the third row shows worst performances. From left to right, column 1 shows one slice of original CT image of each case, column 2 shows the predictions from multi-atlas, column 3 shows the predictions from candidate region, and column 4 shows the predictions from our learning framework.

TABLE 1. Performance of the multi-atlas step in testing images.

	Precision(%)	Recall(%)	DSC(%)
Mean	61.64	74.48	66.54
Min	38.11	40.42	44.82
Max	80.58	88.45	80.25

TABLE 2. Our model's performance in training images.

	Precision(%)	Recall(%)	DSC(%)
Mean	80.20	91.13	84.99
Min	70.56	60.67	69.35
Max	89.07	98.57	91.47

as allows the network to locate the target more robustly. Benefited from this setting, our model finally get a competitive result with an average recall of 82.65% recall and an average DSC of 78.13%. The training of the 3D-UNet takes 8 hours for 50000 epochs on a GPU (Nvidia GTX Titan X).

Table 1 shows the performance after the multi-atlas step. The mean recall is 74.48% that indicates the position of pancreas is roughly located in abdomen.

Besides, in the extracted candidate regions, there are 90% of cases being above 88.94% recall with the mean recall reaching 93.04%. Therefore it can be seen that most of the pancreas region is covered in these candidate regions.

Table 2 and 3 provide the our model's performance in training and testing images respectively. Only one outlier case has a recall below 70%. In addition, 80% of cases have precisions above 71.56%. The high recall values could indicate that

TABLE 3. Our model's performance in testing images.

	Precision(%)	Recall(%)	DSC(%)
Mean	74.64	82.65	78.13
Min	56.15	65.99	66.50
Max	84.12	93.81	87.49

the pancreas area can be effectively preserved during model prediction.

The average, maximum and minimum Hausdorff distances of our predicted cases and ground truths are 11.90mm, 23.88mm and 4.49mm respectively.

Figure 7 shows the ROC curves from multi-atlas vs. from MCMC U-Net. But the left side of this picture is too crowded to see clearly, so we make false positive rate to be semilog. In Figure 8,we find that when $0 < f < 19$, the ROC curve of U-Net is above the ROC curve of multi-atlas distinctly. This reveals that when we extract the candidate regions by $f = 10$, our model successfully reject substantial amount of false-positive regions.

In Figure 9, we shows three examples from testing dataset with their predicted segmentation and ground truth. For each example, we show predictions generated from the multi-atlas step, the selected candidate region, and from our learning framework, respectively. The first row displays the case with best performances: 87.49% DSC, 92.68% Recall and 82.85% Precision. The second row shows the case whose performance

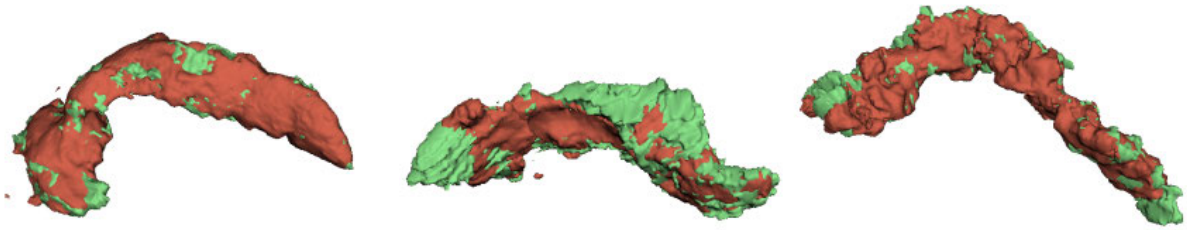

FIGURE 10. Displays three examples' results in 3D rendering, this three examples have been shown in Figure 9, the ground truth is marked as green volume, and the predictions from our framework is marked as red volume.

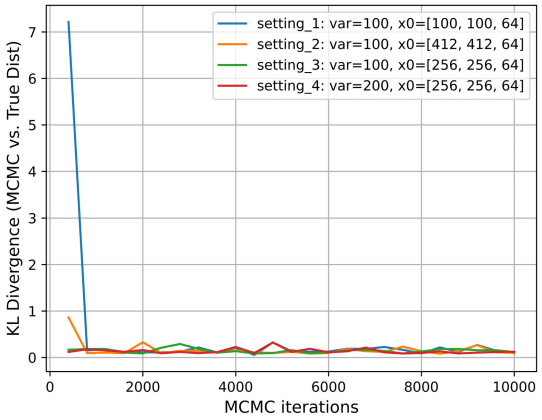

FIGURE 11. KL Divergence between MCMC samples and spatial prior distribution. We collect valid sample points accepted in Algorithm 1 with $n_1 = 0, 400, \ldots 9600$ and $n_2 = 400$. The horizontal axis represent $n_1 + n_2$, total number of MCMC iterations. The vertical axis represent the KL divergence between the estimated kernel density from the accepted samples and the theoretical distribution p. We apply 4 different candidate kernels and initial proposal point in four settings.

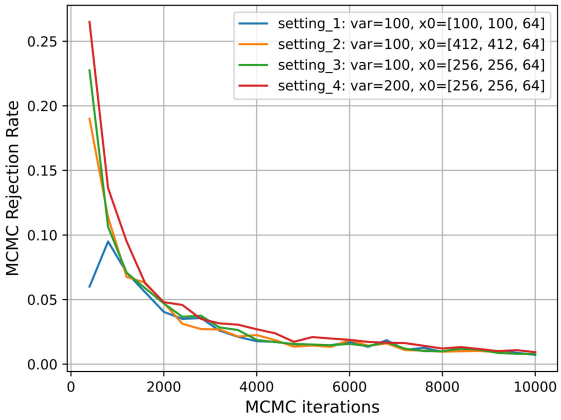

FIGURE 12. Rejection rate in Algorithm 1. We collect valid sample points accepted in Algorithm 1 with $n_1 = 0, 400, \ldots 9600$ and $n_2 = 400$. The horizontal axis represent $n_1 + n_2$, total number of MCMC iterations. The vertical axis represent the rejection rate measured by the ratio of rejected samples to total number of generated points. We apply 4 different candidate kernels and initial proposal point in four settings.

is close to average with 76.72% DSC, 76.26% Recall and 77.19% Precision. The third row is the image with worst performance of 66.50% DSC, 81.54% Recall and 56.15% Precision.

In Figure 10, we also show the 3d rendering of these three examples. We can approximately locate the pancreas' region in its vicinity with multi-atlas, but there are some regions missed. With the process of extracting candidate regions, more pancreas regions is covered correctly. In addition, the boarder area of the pancreas is included so that we could get the marginal information of pancreas.

In the third example (third row of Figure 9), despite high recall, our framework gives much lower precision, with observed false positives near the pancreatic region. Though our work already provided competitive results for the worst evaluation case (see Table 4) comparing to other popular methods, we will discuss the limitations revealed from this observation and possible improvements in Section V.

In addition, we have compared the accuracy, robustness and efficiency of our proposed framework with recent state-of-the-arts methods for pancreas segmentation. In Table 4, we listed precision, recall, the mean, minimum and maximum value of DSC on evaluation, as well as the time used to generate predictions on test images. We should note that the work we are comparing with might not have the full set of evaluation metrics available.

First, our average DSC on evaluation is higher than the work in [9]–[11], and also comparable to [30], but lower than the results in [7], [31], [33]. However, different from our work, in methods in [7], [31], [33], modeling on 2D slices of the 3D CT volume serves as an important part in the learning pipeline. The work in [7], [31] built localization and segmentation modules purely on 2D slices and then aggregate the results back into 3D, while [33] combines a 3D patch-based CNN and a 2.5D slice-based CNN for fine segmentation. Another difference is that all of them [7], [31], [33] essentially are hybrid systems combining localization (coarse) and segmentation (fine) steps. Researchers in [33] also included an extra fine-tuning step for segmentation refinement. In contrast, our framework is purely built on 3D patch-based modeling, and is single-stage without extra steps for localization. We also did not took extensive efforts in calibrating CNN architectures, feature fusion and fine-tuning as in [7], [31], [33] since the focus of this work is to demonstrate MCMC guided segmentation.

TABLE 4. Some recent state-of-the-art methods on pancreas segmentation are compared.

		Mean[min, max]			
Method	Cases	Precision(%)	Recall(%)	DSC(%)	Testing time
Roth et al., MICCAI'2015 [9]	82	-	-	71.8[25.0∼86.9]	6 - 8 min
Roth et al., MICCAI'2016 [11]	82	-	-	78.01[34.11∼88.65]	2 - 3 min
Zhou et al., MICCAI'2017 [7]	82	-	-	**82.37**[62.43∼90.85]	-
Karasawa et al.,SCI'2017 [30]	150	-	-	78.5	-
Farag et al.,IEEE'2017 [10]	80	71.6[34.8∼85.8]	74.4[15.0∼90.9]	70.7[24.4∼85.3]	-
Zhang et al, PR'2020 [31]	82	-	-	84.90[61.82∼91.46]	-
Zhang et al, MEDIA'2021 [33]	82	-	-	84.47[70.61∼91.54]	3 - 5 min
Ours	82	**74.64**[56.15∼84.12]	**82.65**[65.99∼93.81]	78.13[66.50∼87.49]	0.5 - 2 min

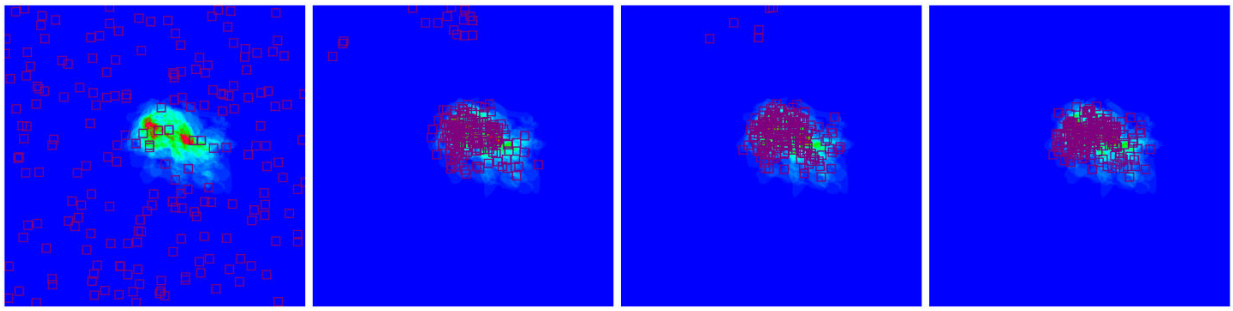

FIGURE 13. Visualizing generated patches from global uniform sampling and different iterations of MCMC Algorithm 1. The heatmap represent a selected slice of the 3D prior spatial probability density p. The red boxes represent those generated 3D patches intersecting with this slice. For figures from left to right, the patches are generated through 1) global uniform sampling of 400 patches 2) last 400 samples from MCMC iteration 400 3) last 400 samples from MCMC iteration 800 4) last 400 samples from MCMC iteration 1200.

Second, we realized that our lowest DSC is substantially higher than other methods [7], [9]–[11], [30], [31], except [33]. Also, our lowest recall and precision are also significantly higher than other available results [10]. This observation demonstrated that, benefited from MCMC guided learning, our framework is robust to variations and could still perform well when dealing with hard cases in the dataset.

In addition, our framework is also efficient in that it takes less time during inference. The MCMC scheme provided us efficient generation of patch-wise predictions during inference, as explained in III-C.

2) MCMC GUIDED LEARNING

To illustrate why MCMC is important in guiding training and segmentation, we have also trained the identical 3D-UNet in an conventional way. We sample patches of the same size uniformly inside the whole image domain. We also use the same batch size of 100 during training. We found out that it takes much more epochs for the network to converge, since in each batch most patches are simply selected from background area remote from the target region. We stopped the training at 50000 epochs, the same as in the MCMC case. The metrics on evaluation set is much worse than our proposed approach. The 3D-UNet training with uniform patch generation gives an average DSC of 68.25%, precision 66.98% and recall 68.76%. Moreover, the average testing time is 7.5 minutes as it takes approximately 10 times more samples to generate reasonable segmentation masks, comparing to the MCMC version.

We will also show that the MCMC scheme we constructed converges pretty fast, and brings almost no computational overheads. To evaluate how well MCMC generated samples could approximate the spatial distribution p, we compute the KL divergence as a measure of dissimilarity between two distributions. In particular, let $S = \{\mathbf{s_i}, i = 1, \ldots, |S|\}$ be the sample generated from certain MCMC steps according to p, then we obtain a kernel density estimation from S as $\hat{f}_h(\mathbf{x}), \mathbf{x} \in \Omega$. The KL divergence of $\hat{f}_h$ and p, $D_{KL}(\hat{f}_h \| p)$ is expected to be small if S could approximate p well.

Figure 11 shows the changes of KL divergence with MCMC iterations in Algorithm 1. We set $n_2 = 400$ and change the value of n_1 within $\{0, 400, \ldots, 9600\}$; equivalently, we perform 25 MCMC experiments with different number of iterations. We also tested different settings of candidate kernels q. To test the robustness of the algorithm, we also test four settings including different candidate kernels q and initial proposals $\mathbf{x}^*$. We use multi-variate Gaussian density to construct q, whose covariance matrix has the form $\Sigma = [var, 0, 0 \| 0, var, 0 \| 0, 0, var]$, where $var = \{100, 100, 100, 200\}$ from settings 1 to 4, respectively. Also, the initial proposal points we used in four settings are $x_1^* = [100, 100, 64]$, $x_2^* = [412, 412, 64]$, $x_3^* = [256, 256, 64]$, and $x_4^* = [256, 256, 64]$. It is shown in Figure 11 that regardless of different initial conditions and candidate kernels, the algorithm converges pretty fast. Under all settings, the KL divergence stabilizes at a small value after the 800th iteration. Figure 12 gives the rejection rate at different iterations under the same experimental settings as in Figure 11. We can see that the rejection rate approaches to zero as we have more iterations, and it dropped under 10% across all settings after the 1200th iteration. It takes approximately 0.08 seconds for 800 iterations of the algorithm, and due to fast convergence,

800 iterations is already sufficient for driving the training and segmentation process. Moreover, Figure 13 gives a more straightforward visualization on patch generation. We can see that at 1200th iteration, the MCMC generated patches totally follows the spatial prior distribution p, and even at the 400th iteration, the patches could cover the high-value region pretty well already, with a few exceptions caused by initialization.

V. DISCUSSION, CONCLUSION AND FUTURE WORK

In this work, we proposed a general purpose segmentation framework that uses the Monte Carlo Markov Chain (MCMC) to guide segmentation of the 3D images. Specifically, the prior spatial distribution is learned and an MCMC scheme is utilized to generate 3D patches from the prior, serving as input to the convolutional neural network. During segmentation, the MCMC is employed again to guide patch selection from the high probability regions in the target image. The selected regions are fed to the trained CNN, from which the final segmentation are constructed through patch-wise fusion.

The proposed framework is applied to the abdominal CT images to extract the pancreas and have achieved competitive results and demonstrated robustness and efficiency. Moreover, the framework is highly flexible such that it can integrate any other variations of the segmentation network.

The current version of our framework has room for further improvements. Though it provided a more balanced dataset and effective variance reduction, as shown in Sections III-C and IV, there's still uncertainties in practice, for example, due to the choices of candidate kernel and initial proposal for MCMC. As a part of future directions, we could introduce additional mechanisms in Algorithm 1 to sample meaningful negative patches adaptively, which would mitigate the low precision issue as shown in the third row of Figure 9. Additionally, the current patch generation procedure could work with spatial attention to iteratively produce more optimized localization, without extra memory overheads, within the end-to-end learning framework. We could also explore relations among generated patches to capture more global representations.

Other future directions include investigating the variances induced by the imaging parameters, such as the field of view, with/without contrast agent, slice/slab thickness, etc. Moreover, the proposed method will also be used in conjunction with the recognition of various pancreatic diseases.

ACKNOWLEDGMENT

(Mu Tian, Jinchan He, and Xiaxia Yu contributed equally to this work.)

REFERENCES

[1] A. Mittal, R. Hooda, and S. Sofat, "LF-SegNet: A fully convolutional encoder–decoder network for segmenting lung fields from chest radiographs," *Wireless Pers. Commun.*, vol. 101, no. 1, pp. 511–529, Jul. 2018.

[2] J. Park, J. Yun, N. Kim, B. Park, Y. Cho, H. Park, M. Song, M. Lee, and J. Seo, "Fully automated lung lobe segmentation in volumetric chest CT with 3D U-Net: Validation with intra- and extra-datasets," *J. Digit. Imag.*, vol. 33, pp. 1–10, May 2019.

[3] W. Qin, J. Wu, F. Han, Y. Yuan, W. Zhao, B. Ibragimov, J. Gu, and L. Xing, "Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation," *Phys. Med. Biol.*, vol. 63, no. 9, May 2018, Art. no. 095017.

[4] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[5] M. Avendi, A. Kheradvar, and H. Jafarkhani, "Fully automatic segmentation of heart chambers in cardiac MRI using deep learning," *J. Cardiovascular Magn. Reson.*, vol. 18, no. S1, pp. 1–3, Dec. 2016.

[6] T. A. Ngo, Z. Lu, and G. Carneiro, "Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance," *Med. Image Anal.*, vol. 35, pp. 159–171, Jan. 2017.

[7] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. Fishman, and A. Yuille, "A fixed-point model for pancreas segmentation in abdominal CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 10433, 2017, pp. 693–701.

[8] M. Fu, W. Wu, X. Hong, Q. Liu, J. Jiang, Y. Ou, Y. Zhao, and X. Gong, "Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images," *BMC Syst. Biol.*, vol. 12, no. S4, pp. 119–127, Apr. 2018.

[9] H. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. Turkbey, and R. Summers, "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9349, 2015, pp. 556–564.

[10] A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey, and R. M. Summers, "A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 386–399, Jan. 2017.

[11] H. Roth, L. Lu, A. Farag, A. Sohn, and R. Summers, "Spatial aggregation of holistically-nested networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9901, 2016, pp. 451–459.

[12] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Med. Image Anal.*, vol. 45, pp. 94–107, Apr. 2018.

[13] 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.

[14] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-SCNN: Gated shape CNNs for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5228–5237.

[15] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation," 2019, *arXiv:1903.11816*. [Online]. Available: http://arxiv.org/abs/1903.11816

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 386–397, Feb. 2020.

[17] O. Cicek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, 2016, vol. 9901, no. 7, pp. 424–432.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 9351, 2015, pp. 234–241.

[19] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "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.

[20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[21] B. Zhao, X. Chen, Z. Li, Z. Yu, S. Yao, L. Yan, Y. Wang, Z. Liu, C. Liang, and C. Han, "Triple U-Net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101786. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136184152030150X

[22] S. Zhou, D. Nie, E. Adeli, J. Yin, J. Lian, and D. Shen, "High-resolution encoder–decoder networks for low-contrast medical image segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 461–475, 2020.

[23] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2016, *arXiv:1511.07122*. [Online]. Available: https://arxiv.org/abs/1511.07122

[24] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[25] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. R. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Towards image-guided pancreas and biliary endoscopy: Automatic multi-organ segmentation on abdominal ct with dense dilated networks," in *Medical Image Computing and Computer Assisted Intervention*. Cham, Switzerland: Springer, 2017, pp. 728–736.

[26] Q. Yu, L. Xie, Y. Wang, Y. Zhou, E. K. Fishman, and A. L. Yuille, "Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8280–8289.

[27] A. Farag, L. Lu, E. Turkbey, J. Liu, and R. Summers, "A bottom-up approach for automatic pancreas segmentation in abdominal CT scans," in *Proc. Int. MICCAI Workshop Comput. Clin. Challenges Abdominal Imag.*, 2014, pp. 103–113.

[28] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: http://arxiv.org/abs/1804.03999

[29] W. Li, S. Qin, F. Li, and L. Wang, "MAD-UNet: A deep U-shaped network combined with an attention mechanism for pancreas segmentation in CT images," *Med. Phys.*, vol. 48, no. 1, pp. 329–341, Jan. 2021.

[30] K. Karasawa, M. Oda, T. Kitasaka, K. Misawa, M. Fujiwara, C. Chu, G. Zheng, D. Rueckert, and K. Mori, "Multi-atlas pancreas segmentation: Atlas selection based on vessel structure," *Med. Image Anal.*, vol. 39, pp. 18–28, Jul. 2017.

[31] D. Zhang, J. Zhang, Q. Zhang, J. Han, S. Zhang, and J. Han, "Automatic pancreas segmentation based on lightweight DCNN modules and spatial prior propagation," *Pattern Recognit.*, vol. 114, Jun. 2021, Art. no. 107762.

[32] Y. Man, Y. Huang, J. Feng, X. Li, and F. Wu, "Deep Q learning driven CT pancreas segmentation with geometry-aware U-Net," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1971–1980, Aug. 2019.

[33] Y. Zhang, J. Wu, Y. Liu, Y. Chen, W. Chen, E. X. Wu, C. Li, and X. Tang, "A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101884.

[34] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–13. [Online]. Available: https://openreview.net/forum?id=r1Ddp1-Rb

[35] R. Takahashi, T. Matsubara, and K. Uehara, "Data augmentation using random image cropping and patching for deep CNNs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 9, pp. 2917–2931, Sep. 2020.

[36] M. Nishio, S. Noguchi, and K. Fujimoto, "Automatic pancreas segmentation using coarse-scaled 2D model of deep learning: Usefulness of data augmentation and deep U-Net," *Appl. Sci.*, vol. 10, no. 10, p. 3360, May 2020.

[37] V. V. Valindria, I. Lavdas, J. Cerrolaza, E. O. Aboagye, A. G. Rockall, D. Rueckert, and B. Glocker, "Small organ segmentation in whole-body MRI using a two-stage FCN and weighting schemes," 2018, *arXiv:1807.11368*. [Online]. Available: http://arxiv.org/abs/1807.11368

[38] Y. Gao, R. Huang, Y. Yang, J. Zhang, K. Shao, C. Tao, Y. Chen, D. N. Metaxas, H. Li, and M. Chen, "FocusNetv2: Imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101831.

[39] K. Dijkstra, J. van de Loosdrecht, W. A. Atsma, L. R. B. Schomaker, and M. A. Wiering, "CentroidNetV2: A hybrid deep neural network for small-object segmentation and counting," *Neurocomputing*, vol. 423, pp. 490–505, Jan. 2021.

[40] T. Zhang, Z. Peng, H. Wu, Y. He, C. Li, and C. Yang, "Infrared small target detection via self-regularized weighted sparse model," *Neurocomputing*, vol. 420, pp. 124–148, Jan. 2021.

[41] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Asymmetric contextual modulation for infrared small target detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 950–959.

[42] Y. Dai, Y. Wu, Y. Song, and J. Guo, "Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values," *Infr. Phys. Technol.*, vol. 81, pp. 182–194, Mar. 2017.

[43] H. Zhu, S. Liu, L. Deng, Y. Li, and F. Xiao, "Infrared small target detection via low-rank tensor completion with top-hat regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1004–1016, Feb. 2020.

[44] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, and Y. Yu, "Cross-modality deep feature learning for brain tumor segmentation," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107562.

[45] Y. Gao, L. Zhu, J. Cates, R. S. MacLeod, S. Bouix, and A. Tannenbaum, "A Kalman filtering perspective for multiatlas segmentation," *SIAM J. Imag. Sci.*, vol. 8, no. 2, pp. 1007–1029, Jan. 2015.

[46] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 611–623, Mar. 2013.

[47] Y. Gao, B. Corn, D. Schifter, and A. Tannenbaum, "Multiscale 3D shape representation and segmentation with applications to hippocampal/caudate extraction from brain MRI," *Med. Image Anal.*, vol. 16, no. 2, pp. 374–385, Feb. 2012.

[48] Y. Huo, A. J. Plassard, A. Carass, S. M. Resnick, D. L. Pham, J. L. Prince, and B. A. Landman, "Consistent cortical reconstruction and multi-atlas brain segmentation," *NeuroImage*, vol. 138, pp. 197–210, Sep. 2016.

[49] G. Erus, J. Doshi, Y. An, D. Verganelakis, S. M. Resnick, and C. Davatzikos, "Longitudinally and inter-site consistent multi-atlas based parcellation of brain anatomy using harmonized atlases," *NeuroImage*, vol. 166, pp. 71–78, Feb. 2018.

[50] J. Huo, J. Wu, J. Cao, and G. Wang, "Supervoxel based method for multi-atlas segmentation of brain MR images," *NeuroImage*, vol. 175, pp. 201–214, Jul. 2018.

[51] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970.

[52] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *Amer. Statistician*, vol. 49, no. 4, pp. 327–335, Nov. 1995.

[53] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[54] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2015.

[55] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1–9.

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 770–778.

[57] H. P. Heinrich, M. Jenkinson, M. Brady, and J. A. Schnabel, "MRF-based deformable registration and ventilation estimation of lung CT," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1239–1248, Jul. 2013.

[58] H. R. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. M. Summers, "Data from pancreas-CT. The cancer imaging archive," 2016.

[59] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.

[60] C.-Y. Tsai and C.-C. Yu, "Real-time textureless object detection and recognition based on an edge-based hierarchical template matching algorithm," *J. Appl. Sci. Eng.*, vol. 21, pp. 229–240, Jan. 2018.

[61] M. H. Bagheri, H. Roth, W. Kovacs, J. Yao, F. Farhadi, X. Li, and R. M. Summers, "Technical and clinical factors affecting success rate of a deep learning method for pancreas segmentation on CT," *Acad. Radiol.*, vol. 27, no. 5, pp. 689–695, May 2020.

[62] J. Cai, L. Lu, Z. Zhang, F. Xing, L. Yang, and Q. Yin, "Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks," in *Proc. MICCAI*, 2016, pp. 442–450.
[63] J. Cai, L. Lu, F. Xing, and L. Yang, "Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning," 2018, *arXiv:1803.11303*. [Online]. Available: http://arxiv.org/abs/1803.11303
[64] M. Heinrich, O. Maier, and H. Handels, "Multi-modal multi-atlas segmentation using discrete optimisation and self-similarities," in *Proc. VISCERAL Challenge@ISBI*, 2015.

MU TIAN received the Ph.D. degree in applied mathematics and statistics from the State University of New York at Stony Brook, in 2017. He joined Facebook Inc., as a Research Scientist. He joined Shenzhen University, as an Associate Research Fellow, in 2020. His research interests include artificial intelligence, computer vision, and medical image analysis.

JINCHAN HE received the bachelor's degree from Jiangnan University, Wuxi, Jiangsu, China. She was a Graduate Student with the School of Biomedical Engineering, Health Science Center, Shenzhen University.

XIAXIA YU received the Ph.D. degree in computer science from Georgia State University, in 2015. She joined the School of Biomedical Engineering, Shenzhen University, as an Assistant Professor, in 2017. Her research interests include clinical informatics, designing machine learning algorithms to risk stratify patients from electronic health record, medical statistics, molecular bioinformatics, and drug resistant prediction.

CHUDONG CAI received the B.S. and M.S. degrees from Sun Yat-sen University, Guangzhou, China, in 2004 and 2011, respectively. He is currently an Associate Chief Physician with the Shantou Hospital Affiliated Sun Yat-sen University.

YI GAO received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 2003 and 2005, respectively, and the M.S. degree in mathematics and the Ph.D. degree in biomedical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2008 and 2011, respectively. He was a Postdoctoral Research Fellow with the Harvard Medical School and an Assistant Professor with the Department of Biomedical Informatics, State University of New York at Stony Brook. He is currently a Professor with the School of Biomedical Engineering, Shenzhen University.