Challenge Report

A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set

Yue Zhang [a,b,1], Jiong Wu [a,c,1], Yilong Liu [b], Yifan Chen [d], Wei Chen [e], Ed. X. Wu [b], Chunming Li [f], Xiaoying Tang [a,*]

[a] Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China
[b] Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China
[c] School of Computer and Electrical Engineering, Hunan University of Arts and Science, Hunan, China
[d] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China
[e] Department of Radiology, Third Military Medical University Southwest Hospital, Chongqing, China
[f] Department of Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China

ARTICLE INFO

ABSTRACT

In this paper, we propose and validate a deep learning framework that incorporates both multi-atlas registration and level-set for segmenting pancreas from CT volume images. The proposed segmentation pipeline consists of three stages, namely coarse, fine, and refine stages. Firstly, a coarse segmentation is obtained through multi-atlas based 3D diffeomorphic registration and fusion. After that, to learn the connection feature, a 3D patch-based convolutional neural network (CNN) and three 2D slice-based CNNs are jointly used to predict a fine segmentation based on a bounding box determined from the coarse segmentation. Finally, a 3D level-set method is used, with the fine segmentation being one of its constraints, to integrate information of the original image and the CNN-derived probability map to achieve a refine segmentation. In other words, we jointly utilize global 3D location information (registration), contextual information (patch-based 3D CNN), shape information (slice-based 2.5D CNN) and edge information (3D level-set) in the proposed framework. These components form our cascaded coarse-fine-refine segmentation framework. We test the proposed framework on three different datasets with varying intensity ranges obtained from different resources, respectively containing 36, 82 and 281 CT volume images. In each dataset, we achieve an average Dice score over 82%, being superior or comparable to other existing state-of-the-art pancreas segmentation algorithms.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Pancreatic cancer is a deterious disease that has been significantly affecting human health. In the United States alone, it is estimated that annually there are about 40,000 people dying of pancreatic cancer and 50,000 new cases (Siegel et al., 2017). It is the 4th-5th leading cause of cancer-related mortality (Rahib et al., 2014). The prognosis for patients with pancreatic cancer is dismal, with a 5-year survival rate being less than 5% (Schima et al., 2007; Ryan et al., 2014).

Computed tomography (CT) provides detailed imaging of the pancreas and is one of the most common examinations for diagnosing pancreatic cancer. A direct sign of pancreatic adenocarcinoma is the presence of hypoattenuating/hyperintense mass on contrast-enhanced CT images (Yang et al., 2013). Pancreas segmentation from CT images is also useful for detecting unusual volume changes and monitoring abnormal growths, playing important roles in diagnosis, prognosis, surgical planning as well as intra-operative guidance (Shimizu et al., 2010).

Usually, the recognized gold standard pancreas segmentation is obtained from experienced radiologists via manual delineation (Dou et al., 2017). However, manually tracing volumetric CT images in a slice-by-slice manner is tedious, time-consuming and labor expensive. In addition, manual labeling is highly prone to inter- and intra-variability. As such, a fully-automated pancreas segmentation pipeline is desired.

The pancreas is one of the most challenging organs in terms of automated segmentation because it occupies a relatively small region in the input CT image, e.g., less than 1.5% in a 2D image and less than 0.5% in a 3D image (Roth et al., 2015; Zhou et al., 2017). Another fundamental challenge is that pancreas's appearance and

---

* Corresponding author.
 E-mail address: tangxy@sustech.edu.cn (X. Tang).
[1] Equal contribution.

shape often vary significantly from subject image to subject image (Farag et al., 2017). Furthermore, the boundary between the pancreas and its neighboring tissues is usually ambiguous with limited contrast, because of their similar imaging-related physical properties, e.g., the attenuation coefficient in CT imaging (Kronman and Joskowicz, 2016).

To solve these challenges, various segmentation algorithms have been designed in the past decades, mainly employing statistical shape modeling, level-set, multi-atlas and graphical models with hand-crafted features (Shimizu et al., 2010; Saito et al., 2016; Tsai et al., 2003; Wolz et al., 2012). The hand-crafted features may nevertheless have limited representation capabilities in dealing with the large variations of pancreas's appearance and shape.

Recently, great progress has been made in natural image processing using convolutional neural network (CNN), given that multi-layer convolutions can hierachically learn highly representative features (Krizhevsky et al., 2012; He et al., 2016; Huang et al., 2017). It has also inspired some successful applications in the medical image analysis domain (Ronneberger et al., 2015; Chen et al., 2018; Setio et al., 2016; Li et al., 2018; Milletari et al., 2016; Yu et al., 2017; Shen et al., 2017; Roth et al., 2018). CNN based volumetric medical image segmentation methods can be roughly categorized into two groups, i.e., 2D based and 3D based. In 2D CNN segmentation methods, the volumetric data was typically sliced along one of the three image directions (axial, sagittal, and coronal), and then those 2D slices were used as the input images of specifically designed networks (Li et al., 2018). A representative work is the U-net (Ronneberger et al., 2015). To make further use of the information encoded in the volumetric data, 2.5D image patches composed of 2D slices along all three directions were extracted by sliding window through all voxels (Roth et al., 2016b; Ciompi et al., 2015). Another type is slice-based 2.5D methods, wherein three 2D CNNs were constructed to separately segment the input image at three directions and then the segmentation results were fused via voting. These kinds of methods have also been applied to pancreas segmentation (Zhou et al., 2017; Yu et al., 2018). However, 2D CNN based segmentation results usually suffer from topological errors (missing or extra parts). Later on, 3D CNN based methods have been proposed and adapted to pancreas segmentation (Dou et al., 2017; Zhu et al., 2018). For example, a 3D fully convolutional network (FCN) equipped with a 3D deep supervision mechanism was proposed (Dou et al., 2017). The input images of a 3D CNN are mainly local patches sampled from the whole image, considering the limited dataset with ground truth and computation memory. Patch-based models nevertheless could not sufficiently learn the global features of the volumetric data.

To alleviate this issue, we present a novel and efficient coarse-fine-refine deep learning-based fully-automated segmentation method for the pancreas. The fact that a smaller input region may lead to a more accurate segmentation motivates researchers to employ coarse-to-fine methods (Zhou et al., 2017; Zhu et al., 2018), wherein the coarse stage provides a rough localization and the fine stage polishes the segmentation to make it more accurate. Specifically, we first employ multi-atlas based registration to determine a relatively small input region (**coarse step**), then jointly use a patch-based 3D CNN and three slice-based 2D CNNs to fully utilize the volumetric information embedded in the CT image (**fine step**), and finally fine-tune the segmentation result based on a 3D level-set method (**refine step**).

Overall, our contributions in this work are three-fold:

(1) We come up with a coarse approach for pancreas segmentation in the framework of multi-atlas based fast diffeomorphic image registration (Wu and Tang, 2019). This step aims to eliminate redundant computation in 3D patch-based CNN and help sub-sequent neural networks achieve more accurate segmentation by decreasing the false positive probability.

(2) We propose a fine approach by jointly using a 3D patch-based CNN and a 2.5D slice-based CNN to fully extract local connected image features and global shape features. The probability maps obtained from the 3D CNN model are used to determine pancreas's bounding box which is then concatenated with the original CT image to form two-channel inputs to three subsequent 2D U-nets. Such a mechanism can simultaneously speed up the optimization process and improve segmentation accuracy.

(3) We propose a 3D level-set method to refine the probability maps predicted from the previous 3D + 2.5D CNNs, with the segmentation results from the fine step serving as a constraint in the level-set method. We validate the proposed segmentation pipeline on three pancreas CT datasets.

The remainder of this paper is organized as follows. Section 2 provides a detailed description of our proposed coarse-fine-refine framework. The experimental setup and results are presented in Section 3, followed by a discussion of the results in Section 4.

2. Method

The flow chart of the entire testing procedure is illustrated in Fig. 1. At the coarse segmentation stage, we roughly locate pancreas via multi-atlas based image registration, from which a bounding box of the pancreas is determined and then 3D image patches are extracted and fed into a 3D CNN. At the fine segmentation stage, we segment pancreas by jointly considering the local connection feature (3D patch-based CNN) and global shape feature (2D slice-based CNN). At the refine segmentation stage, we effectively maintain the topology of the pancreas and fully explore the information embedded in the 3D segmentation results obtained at the fine stage.

2.1. Multi-atlas based label fusion

In the coarse segmentation step, affine followed by a fast large deformation diffeomorphic metric mapping for image (LDDMM-image) (Wu and Tang, 2019) is used to determine a bounding box of the pancreas in a testing image of interest, the effectiveness of which has already been validated in various existing works, in terms of localizing anatomical structures (Beg et al., 2005; Ceritoglu et al., 2013; Tang et al., 2013).

Given a 3D grayscale template image I_0 and a 3D grayscale target image I_1, both I_0 and I_1 are represented as real-valued functions defined on the background space $\Omega \in R^3$. LDDMM-image seeks a diffeomorphic mapping $\varphi : \Omega \to \Omega$ such that $I_0 \circ \varphi^{-1}$ is well aligned to I_1. The diffeomorphism $\varphi = \psi_1$ is defined as the end point of a curve $\psi_t(t \in [0, 1])$ computed as

$$\psi_t = id + \int_0^t \nu_\tau(\psi_\tau) d\tau, \tag{1}$$

where $id : \Omega \to \Omega$ is the identity mapping such that $id(x) = x, x \in \Omega, \psi_0 = id$ and $\nu_t : \Omega \times t \to R^3$ is a smooth velocity vector field.

To obtain a bounding box of the pancreas in a testing image and to increase the robustness and computational efficiency of the coarse segmentation procedure, five samples are pre-selected from the training dataset to serve as the multiple atlases. Affine followed by LDDMM-image between each atlas image and the testing image is conducted to obtain the corresponding transformation, and then the transformation is applied to the atlas label to obtain the transformed atlas label. After a union operation of those transformed atlas labels followed by a dilation operation, a coarse segmentation of pancreas of the testing image is obtained. The smallest cuboid
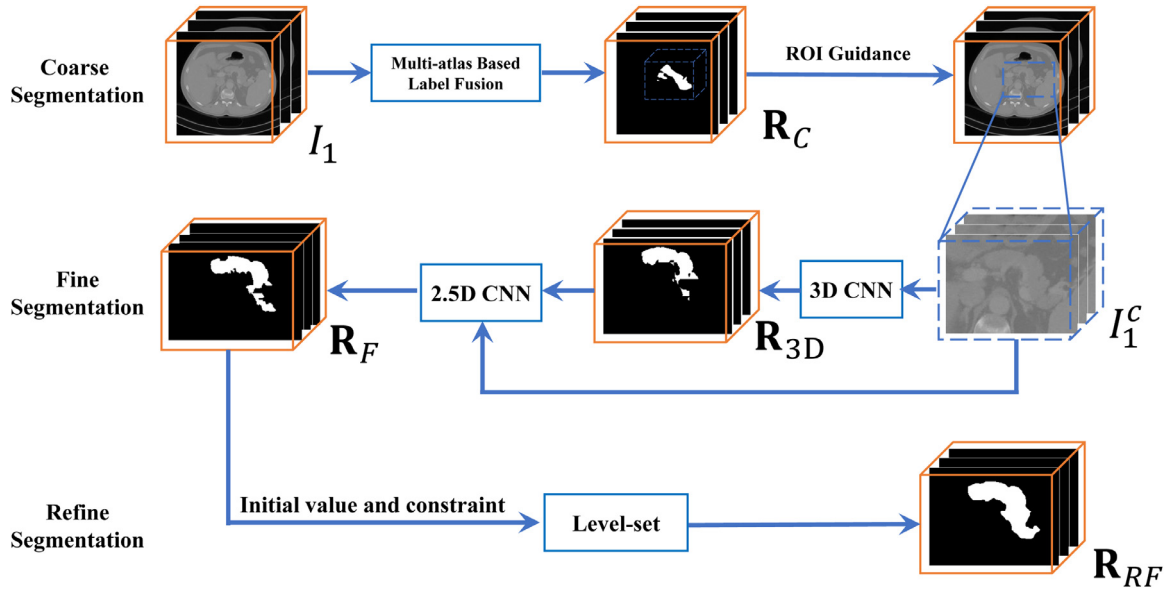
Fig. 1. Flow chart of the entire procedure. Keys: ROI - region of interest.
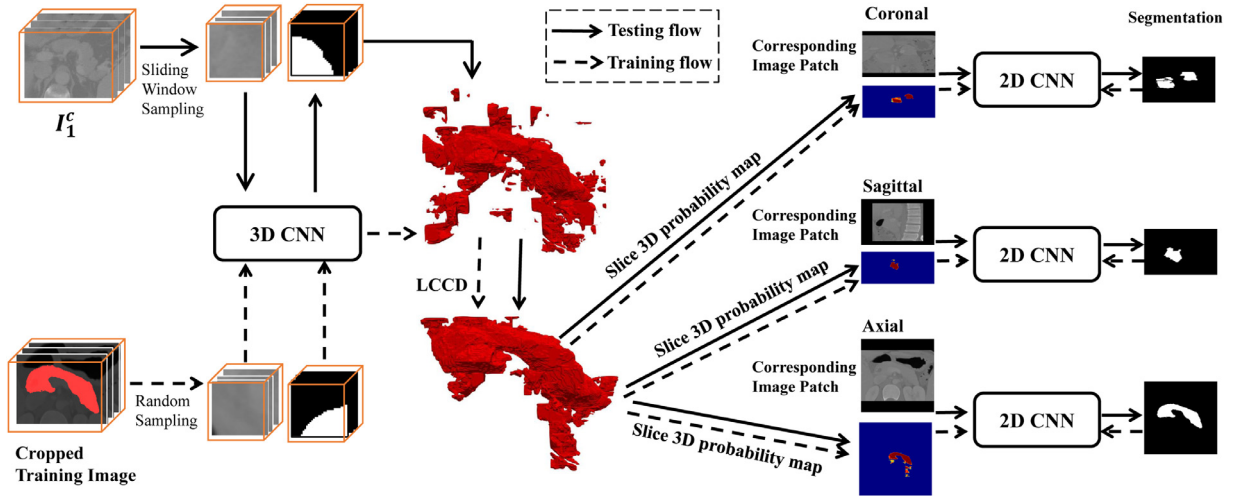

Fig. 2. Details of how 2D and 3D CNNs are used in this work. Keys: LCCD - largest connected component detection.

that contains the coarse segmentation then serves as pancreas's bounding box.

2.2. 3D and 2.5D convolutional neural networks

Training and testing utilizing 3D patches or 2D slices is adopted in most FCN architectures. Typically, 3D patch-based FCNs perform better than 2D slice-based ones since they provide additional context information in the direction orthogonal to 2D slices. It is nevertheless difficult to identify the most suitable patch size for 3D FCNs to ensure sufficient learning of all features, especially the boundary features. To address this issue, in this work, we propose a novel method in a joint 3D + 2D fully convolutional fashion for segmenting pancreas. The overall flow is shown in Fig. 2. Both 2D and 3D CNNs are built based on the U-net (Ronneberger et al., 2015), which consists of a contracting path and an expansive path. The contracting path consists of repeated applications of two convolutions, each followed by a rectified linear unit (ReLU) and a max-pooling operation for downsampling. At each downsampling step, we double the number of feature channels. The expansive path consists of an upsampling of the feature map followed by a

Table 1
Parameters of the 3D and 2D CNNs.

Parameter	3D CNN	2D CNN
Convolution kernel size	$3 \times 3 \times 3$	3×3
Max pooling size	$2 \times 2 \times 2$	2×2
Downsampling times	4	5
Upsampling times	4	5
Number of feature channels in the 1st layer	16	32
Loss function	Cross-entropy	Dice

convolution that halves the number of feature channels, a concatenation with the corresponding feature map from the contracting path, and two convolutions each followed by a ReLU. The key network parameters are tabulated in Table 1. The cross-entropy loss is used in 3D CNN and the Dice-coefficient loss is used in 2D CNNs.

2.2.1. 3D CNN

The 3D CNN in our pipeline is designed to capture connected features from a large number of local patches. The performance of CNN is significantly influenced by the strategy used for extract-

ing training samples. At the training stage, we select 3D patches in each epoch via the sampling method used in Dense training (Kamnitsas et al., 2017). Given that pancreas is much smaller compared to the whole CT image, directly applying dense training to the whole CT image may impair the segmentation performance. To abandon the large-scale useless background patches, we set the sampling region to be a cubic bounding box containing pancreas. To reduce class imbalance, the patches are uniformly sampled from pancreas and background. The size of the training patches is set to be $49 \times 49 \times 49$. The training strategy is the same as that employed in LiviaNET (Dolz et al., 2018).

At the testing stage, we first set the sampling region according to the bounding box obtained from the coarse segmentation step. Then a sliding-window strategy is applied to predict the class probabilities of each pixel, using patch centered at the pixel of interest as input to the network. The window size is set to be $15 \times 15 \times 15$. Overall, the 3D CNN can be formulated as a procedure to segment a testing image I_1 within the bounding box determined by the coarse segmentation $\mathbf{R}_C$,

$$\mathbf{R}_{3D} = \mathbb{M}^{3D}(I_1, \mathbf{R}_C). \tag{2}$$

where $\mathbb{M}^{3D}$ denotes the 3D CNN model with parameters shown in Table 1.

2.2.2. Two-channel 2.5D CNN

The main drawback of a patch-based 3D CNN is that its label prediction is based on very localized information. The segmentation results from our 3D CNN are typically interfered by some noisy isolated components, most of which can be removed by the *largest connected component detection* (*LCCD*). Specifically, we first identify all isolated components of our 3D segmentation result $\mathbf{R}_{3D}$ using connected component analysis (Van der Walt et al., 2014). We then select the component with the largest area and treat it as an updated 3D segmentation result $\mathbf{R}_{3DL}$

$$\mathbf{R}_{3DL} = LCCD(\mathbf{R}_{3D}). \tag{3}$$

Given that a more accurate segmentation can be obtained by using a smaller input region around the region of interest (ROI) and that in a typical 2D CNN the anatomic context in the direction orthogonal to the 2D plane is completely discarded, we design a two-channel 2.5D CNN by using three 2D CNNs at different planes. We not only make use of the predicted segmentation $\mathbf{R}_{3DL}$ to constrain the input region but also slice $\mathbf{R}_{3DL}$ and use them as the second input channel in our 2D CNN.

To feed a 3D volume $\mathbf{X}$ into a 2D network $\mathbb{M}^{2D}$, we use its 2D slices along each of the three planes (axial, sagittal, and coronal). We train three 2D models $\mathbb{M}_C^{2D}, \mathbb{M}_S^{2D}$ and $\mathbb{M}_A^{2D}$ to respectively perform segmentation using the set of 2D slices at each direction, where the subscripts C, S and A respectively stand for "coronal", "sagittal" and "axial".

Cropping or padding is adaptively conducted along each plane based on a 3D bounding box of $\mathbf{R}_{3DL}$. We set the size of the bounding box as $256 \times 256 \times 160$, which means the input sizes for $\mathbb{M}_C^{2D}, \mathbb{M}_S^{2D}$ and $\mathbb{M}_A^{2D}$ are respective $256 \times 160 \times 2, 256 \times 160 \times 2$ and $256 \times 256 \times 2$. Then the final result obtained from the fine segmentation stage $\mathbf{R}_F$ comes from a majority voting of the three 2D CNN results $\mathbb{M}_C^{2D}(I_1, \mathbf{R}_{3DL}), \mathbb{M}_S^{2D}(I_1, \mathbf{R}_{3DL}), \mathbb{M}_A^{2D}(I_1, \mathbf{R}_{3DL})$.

2.3. Level-set

An implicit contour representation of pancreas can be formed with the zero level-set of an Euclidean signed distance function ϕ : $\Omega \to \mathbb{R}$, represented by $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi(\mathbf{x}) = 0\}$, where points inside the contour have $\phi(\mathbf{x}) < 0$ and outside have $\phi(\mathbf{x}) > 0$. The level-set evolution is sensitive to the initial position of the zero level-set

contour (Li et al., 2010). As such, the binary step function ϕ_0 is defined as

$$\phi_0(\mathbf{x}) = \begin{cases} -c_0, & \text{if } \mathbf{R}_F > 0.5 \\ c_0, & \text{otherwise}. \end{cases} \tag{4}$$

Given an initial segmentation $\phi_0(\mathbf{x}) < 0$ as defined in Eq. (4), we run an optimization algorithm to estimate the final pancreas boundary using the distance regularized level-set (DRLSE) formulation (Li et al., 2010). The energy functional is represented by

$$\varepsilon(\phi) = \mu \mathcal{R}_p(\phi) + \varepsilon_{ext}(\phi), \tag{5}$$

where $\mathcal{R}_p(\phi) \triangleq \int_\Omega p(|\nabla\phi|)d\mathbf{x}$ (with $p(s) = \frac{1}{2}(s-1)^2$) is a regularization term that guarantees $|\nabla\phi| \approx 1$ and $\varepsilon_{ext}(\phi)$ is define as Ngo et al. (2017)

$$\varepsilon_{ext}(\phi) = \lambda \mathcal{L}(\phi) + \alpha \mathcal{A}(\phi). \tag{6}$$

The length term $\mathcal{L}(\phi) = \int_\Omega g\delta(\phi)|\nabla\phi|d\mathbf{x}$ computes the line intergral of the edge indicator g along the zero level contour and δ is the Dirac delta function. The area term is $\mathcal{A}(\phi) = \int_\Omega gH(-\phi)d\mathbf{x}$ (H is the Heaviside function).

The gradient flow of the energy $\varepsilon(\phi)$ is then defined as

$$\frac{\partial\phi}{\partial t} = \mu\text{div}(d_p(|\nabla\phi|\nabla\phi)) + \lambda\delta(\phi)\text{div}\left(g\frac{\nabla\phi}{|\nabla\phi|}\right) + \alpha g\delta(\phi) \tag{7}$$

where $\text{div}(\cdot)$ denotes the divergence operator, $\phi(\mathbf{x})$ denotes the current level-set function, ϕ_0 is also used as a constraint in the level-set process and $d_p(\cdot)$ denotes the derivative of the function $p(\cdot)$ used in Eq. (5).

The performance of edge-based level-set methods largely relies on a prerequisite that there are clear and differentiating boundaries boundaries surrounding the object of interest (Salah et al., 2009). Unfortunately, the pancreas on a CT image often has very ambiguous boundaries, which may significantly impair the performance of level-set. To solve this challenge, we generate an edge indicator g based on the flowchart shown in Fig. 3. Given the the CNN-derived pancreas segmentation ($\mathbf{R}_F$), we obtain histogram statistics for the CT image intensity within the pancreas and then use the 99$^{\text{th}}$ percentile maximum and minimum (T_{low} and T_{high}) to reduce noisy segmentation. A linear mapping is used to normalize image intensity via

$$y = \begin{cases} 0, & \text{x} < T_{\text{low}} \\ \frac{x-T_{\text{low}}}{T_{\text{high}}-T_{\text{low}}}, & T_{\text{low}} < \text{x} < T_{\text{high}}. \\ 1, & \text{x} > T_{\text{high}}. \end{cases} \tag{8}$$

Such truncation and normalization can enhance the image contrast surrounding pancreas and remove residual tissues. The edge indicator g used in Eq. (7) is typically obtained by

$$g = \frac{1}{1 + \nabla G_\sigma * I^2}, \tag{9}$$

where G_σ is a Gaussian kernel with a standard deviation σ. The convolution $*$ is used to smooth the image to reduce noise. This function usually takes smaller values at object boundaries than at other locations.

There are nevertheless some tissues having higher image intensity than the pancreas, such as the aorta, edges of which have smaller values in g and make edges of the pancreas fuzzy (see the green arrows in Fig. 3(a) and (b)). As such, we employ fuzzy c-means (FCM) to estimate the probability that a pixel belongs to a certain image class and mask out these high-intensity tissues. It can be observed that the proposed method can obtain clear edges around the pancreas (see the green arrows on the bottom right corner of Fig. 3). To fully use 3D information, these image processing steps and level-set evolution are conducted over the whole volume.
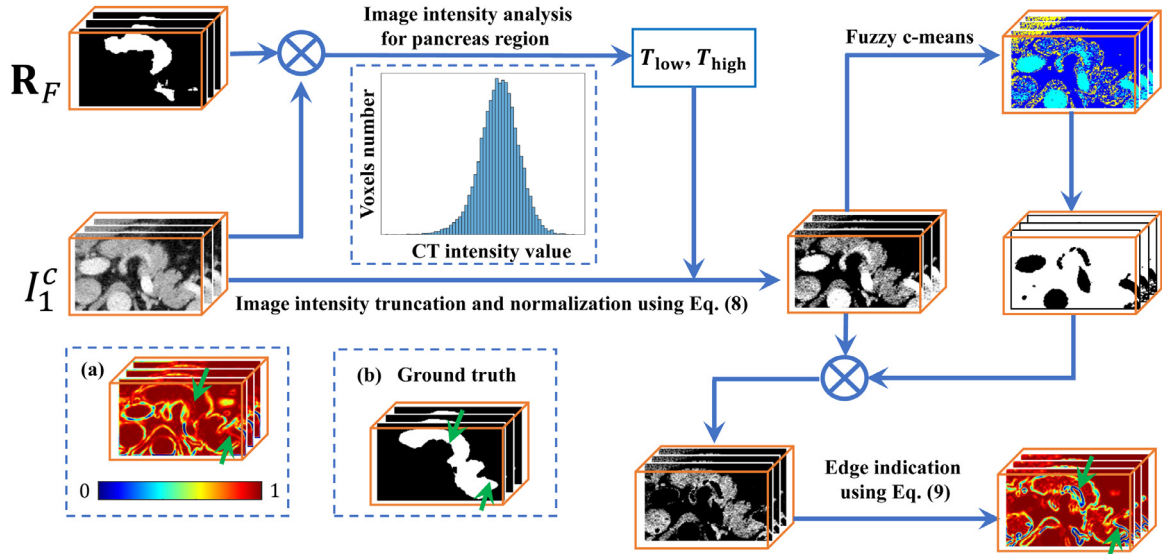
Fig. 3. Flowchart of edge indicator enhancement. Firstly, the CT image intensity is truncated and normalized based on histogram analysis for the pancreas region. After that, fuzzy c-means is used to mask out tissues with high image intensity, resulting in clearer edges.

3. Experimental results

3.1. Dataset and evaluation measures

The accuracy of the pancreas segmentation results produced by our methodology is assessed using three datasets. The first dataset is obtained from a pancreas segmentation challenge on the 2nd International Symposium on Image Computing and Digital Medicine (ISICDM 2018)[2]. During the ISICDM 2018 Pancreas Segmentation Challenge, the organizers release 20 training and 16 testing images. The training set has manual annotations, but the ground truth segmentations for the testing data are hidden during the challenge. The challenge organizers release the ground truth segmentation results for the testing data after the challenge. Currently, all datasets with expert annotations are available. For this dataset, we use the training set to estimate all parameters and the testing set to evaluate the proposed method.

The second one is an NIH dataset which is publicly available and contains 82 abdominal contrast-enhanced 3D CT scans.[3] This NIH dataset has been widely used in previous studies (Roth et al., 2015; Zhou et al., 2017; Cai et al., 2017). For the NIH dataset, we conduct 4-fold cross-validation.

The third one comes from the Medical Segmentation Decathlon (MSD) challenge and contains 281 CT scans with labelled pancreas masks (Fang et al., 2019), with pancreas tumor annotations as well. We combine pancreas and tumor to belong to a single category at both training and testing stages in our experiments given that our proposed method is designed for binary segmentation. For this dataset, we also conduct 4-fold cross-validation.

The characteristics of the three datasets are tabulated in Table 2. The intensity ranges of those three datasets vary significantly. The CT image sizes are $512 \times 512 \times D_i, i \in \{1, 2, 3\}$ and the spatial resolutions are $w \times h \times d$ mm. For the ISICDM dataset, $D_1 \in [205, 376], w = h$ ranging from 0.57 mm to 0.83 mm and d ranging from 1 to 1.1 mm. For the NIH dataset, $D_2 \in [181, 466], w = h$ ranging from 0.66 mm to 0.98 mm, and d ranging from 0.5 mm to 1 mm. For the MSD dataset, $D_3 \in [37, 751], w = h$ ranging from 0.61 mm to 0.98 mm, and d ranging from 0.7 mm to 7.5 mm.

Table 2
Characteristics of the three datasets used in this work.

Dataset	Number	Image size	Intensity range
ISICDM	36	$512 \times 512 \times D_1$	(1024, 3071)
NIH	82	$512 \times 512 \times D_2$	(2048, 3071)
MSD	281	$512 \times 512 \times D_3$	(2048, 4009)

There are three widely used evaluation criteria when measuring the volumetric overlap ratio between a segmentation **R** and its corresponding ground truth **G**, including the Dice score (DSC $= \frac{2|\mathbf{R} \cap \mathbf{G}|}{|\mathbf{R}| + |\mathbf{G}|}$), Jaccard index (JI $= \frac{|\mathbf{R} \cap \mathbf{G}|}{|\mathbf{R} \cup \mathbf{G}|}$), and volume difference (VD $= \frac{|V_R - V_G|}{|V_G|}$), where V_R and V_G denote the volumes of **R** and **G**, and $|\cdot|$ denotes the volume size in Dice score and JI and the absolute value in volume difference.

3.2. Experimental setup

For data pre-processing, we normalized each CT image to have zero mean and unit variance to decrease the data variance induced by the imaging process (Gravel et al., 2004). The parameters in the three segmentation stages were determined based on cross-validation using the training dataset. The atlas number was set to be 5 for multi-atlas based label fusion and the images were down-sampled so that the image resolution was a quarter of that of the original image in the registration process. The epochs number of our 3D model $\mathbb{M}^{3D}$ was set to be 30, with each one again composed of 20 subepochs. At each subepoch, a total of 500 samples were randomly selected from the training images. The epoch numbers of our 2D models $\mathbb{M}_C^{2D}, \mathbb{M}_S^{2D}, \mathbb{M}_A^{2D}$ were all set to be 10.

3.3. Results of each stage of the proposed pipeline

Representative 3D segmentation results obtained from each stage of the proposed pipeline and the ground truth are shown in Fig. 4. Clearly, the automated segmentation results become better and better as the segmentation stage evolves from coarse to fine and finally refine. Overall, a combination of the proposed three-
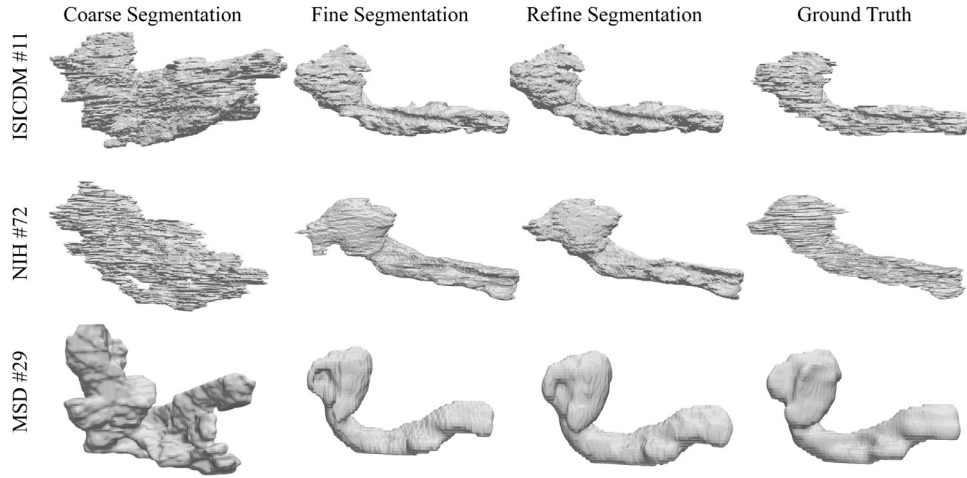

Fig. 4. Three segmentation examples from the three datasets (ISICDM #11, NIH #72, and MSD #29). From left to right, segmentation results obtained from the coarse, fine and refine stages as well as the corresponding ground-truth segmentation.
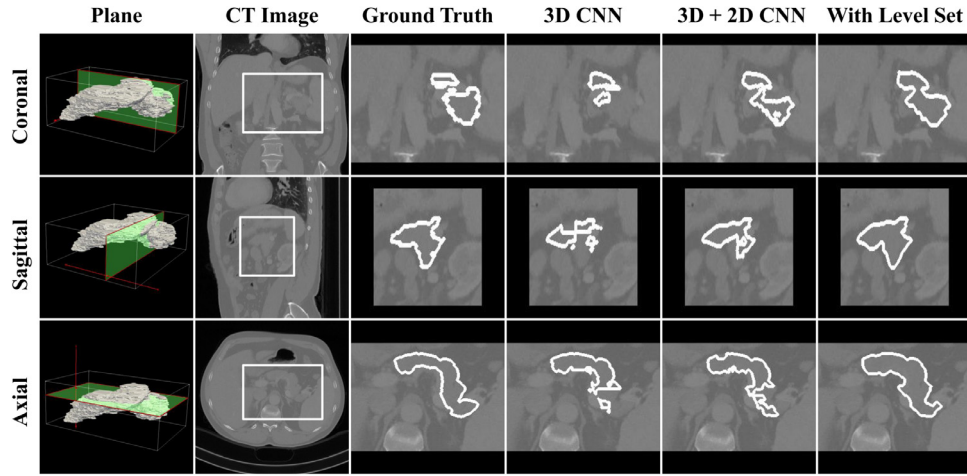

Fig. 5. Representative segmentations results at each of the three planes (coronal, sagittal, and axial). The sliced points is (250,330,148) in ISICDM #6. The green planes in the first column show the location of the selected slice in the 3D pancreas volume. The white frames in the second column show the bounding box determined from the multi-atlas label fusion (coarse segmentation). The white contours in the third column and the last three columns show the ground truth and the automated segmentation results obtained from the three stages.

stages components builds our state-of-the-art pancreas segmentation system. The refine segmentation step is particularly useful for polishing segmentation results with missing parts. In Fig. 5, to clearly show the difference between fine segmentation and refine segmentation, we slice a CT image and its corresponding segmentation mask at a fixed point along three planes. Given that our 3D CNN is based on local patches, there are usually missing voxels, as shown in Fig. 5. In our framework, the 2D CNN is designed to capture global features. It can be seen that a portion of the missing region in the 3D CNN output has been successfully recovered after the 3D + 2.5D CNNs. But these fine segmentation results are still not smooth and complete. In the last column, the refine segmentation results are very close to the ground truth since the level-set method can capture the missing region based on the edge information of the CT image.

Table 3 summarizes the contribution of each component in our segmentation framework. For each metric, we conduct the Wilcoxon signed-rank test between successive stages to evaluate the statistical significance of each component's contribution. Obviously, the accuracy of the coarse segmentation is not high. This is because the label fusion result **C** is much larger than the ground truth **G**. However, for all testing images, pancreases are contained in the bounding box determined by **C** and detailed analysis will

be shown in Section 3.3.1. The average Dice score, Jaccard index, volume difference of the fine segmentation outputs are respective 84.4%, 73.4% and 15.0%, which have been boosted significantly compared to the coarse stage ($p < 0.01$). These improvements may have been contributed to an effective combination of 3D CNN, 2D CNN, and *LCCD*, detailed analysis of which will be shown in Section 3.3.2. After the level-set polishing, those three metrics become respective 85.8%, 75.4%, and 7.9%. Those improvements are significant, with *p*-values of respective 0.0120, 0.0097 and 0.0003. These improvements may have been due to a restoration of the missing parts and an example will be shown in Section 3.3.3.

3.3.1. Coarse segmentation

The main objective of multi-atlas based label fusion is to generate a relatively small region that contains the pancreas. To quantify the ratio between the cropped region to the whole image when using different numbers of atlases, we define $\mathbf{R}_C^{box}$ as the smallest cuboid containing the coarse segmentation $\mathbf{R}_C$, wherein 1 and 0 respectively represent the foreground and the background. The foreground volume fraction (FVF) in $\mathbf{R}_C^{box}$ is defined as

$$\text{FVF} = \frac{\text{Number of voxels in the foreground}}{\text{Number of voxels in the whole image}}. \tag{10}$$

Table 3

The contribution of each component in the proposed segmentation framework, shown as mean ±standard deviation and [worst, best] in %. ↓ indicates that a smaller value represents a better performance.

	Dice score	Jaccard index	Volume difference ↓
Coarse	**24.4 ± 6.4 [14.1, 36.9]**	14.1 ± 4.2 [7.6, 22.7]	**288.0 ± 96.1 [511.1, 156.4]**
Fine	84.4 ± 5.4 [74.1, 92.4]	73.4 ± 8.1 [58.9, 85.9]	15.0 ± 10.3 [38.2, 0.5]
Refine	**85.8 ± 4.8 [74.4, 92.6]**	**75.4 ± 7.2 [59.2, 86.2]**	**7.9 ± 7.0 [20.2, 0.3]**

Table 4

Definitions of TP, FP, TN, and FN.

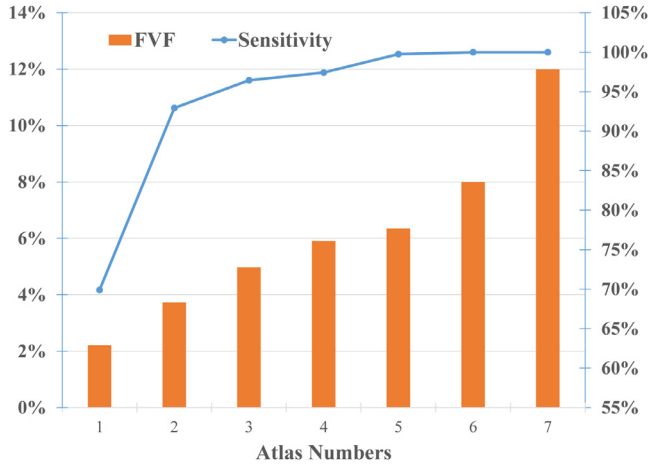|	$\mathbf{R}_C^{box}(i, j, k) = 1$	$\mathbf{R}_C^{box}(i, j, k) = 0$
$\mathbf{G}^{box}(i, j, k) = 1$	True Positive (TP)	False Negative (FN)
$\mathbf{G}^{box}(i, j, k) = 0$	False Positive (FP)	True Negative (TN)

Fig. 6. The average FVF (left vertical axis) and sensitivity (right vertical axis) with different atlas numbers.

Fig. 7. Visualization of feature maps in the first convolution layer and the last deconvolution layer as well as a comparison of segmentation results.

A similar smallest cuboid $\mathbf{G}^{box}$ can be defined for the ground truth $\mathbf{G}$. The definitions of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are listed in Table 4.

The true positive rate, also named sensitivity, is defined as

$$\text{Sensitivity} = \frac{\text{Number of TP voxels}}{\text{Number of (TP + FN) voxels}}. \tag{11}$$

We plot the average FVF and sensitivity of our coarse segmentation when using different numbers of atlases in Fig. 6. If we use one atlas, the FVF for coarse segmentation is about 2%, which is roughly the same as that for ground truth. However, the sensitivity of this case is only 70%, which means there are still 30% of the ground truth missed in the coarse segmentation. With an increase in the atlas number, the sensitivity increases dramatically but the FVF also increases. Our goal is to minimize the FVF but maximize the sensitivity. As such, we set the number of atlases in fast LDDMM-image to be 5, in which case FVF and sensitivity are well balanced.

3.3.2. Fine segmentation

To extract spatial and contextual information, we firstly employed 3D CNN at the fine segmentation stage. And then, the probability maps outputted from the 3D CNN were sliced in the same way as that of the original CT image and used as the second input channel to the subsequent 2D CNNs. To demonstrate the effectiveness of this strategy, we compare the first layer and the last layer feature maps of 2D CNN with a single channel (CT images) and 2D CNN with two channels (CT images and 3D probability maps) in Fig. 7.
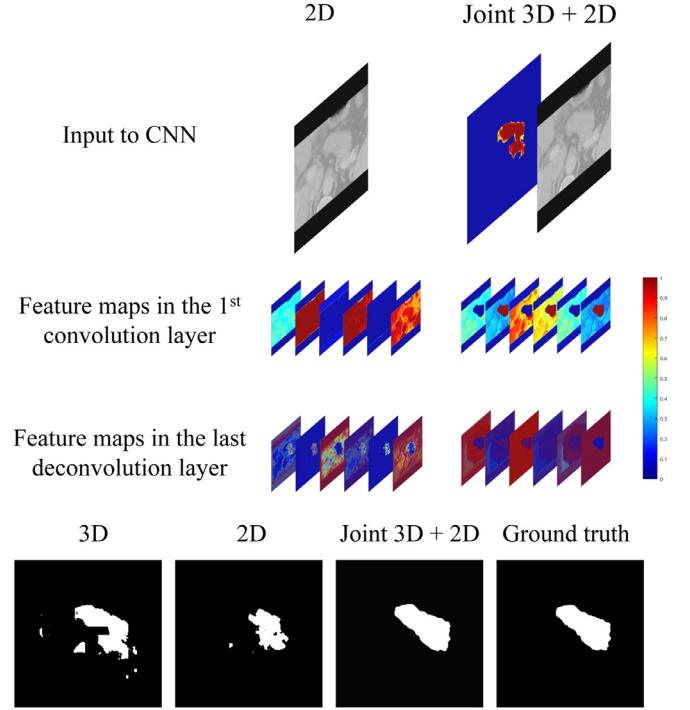
Evidently, the pancreas region is more actively activated in the first layer of the two-channel CNN and is also more concentrated in its last layer. Comparing their segmentation results in the last row of Fig. 7, there are some missing and extra parts in both results from the single 3D CNN and the single 2D CNN. The segmentation result of the proposed joint 3D + 2D is much closer to the ground truth. In conclusion, our two-channel strategy can fully use the 3D patch-based results by providing rough location information, which makes the activation region more concentrated on the pancreas region. As shown in Fig. 5, the pancreas looks quite different at the three 2D planes. To fully learn features at different planes, we trained three 2D models to perform individual segmentation. Finally, the three segmentation results were fused via majority voting. This strategy is named 2.5D.

In Table 5, we further tabulate detailed performance statistics for each component in the fine segmentation stage. *LCCD* can remove a majority of redundant regions. Wilcoxon signed-rank tests are again conducted to quantify the statistical significance. It reduces the average volume difference from 42.7% to 27.7% and increases the average Dice score from 75.0% to 79.3% and the average Jaccard index from 60.8% to 66.3%. The improvement is statistically significant on each metric ($p < 0.05$). Our proposed joint 3D + 2D strategy can further improve the average Dice score and Jaccard index by about 3% and reduce the average volume difference by about 10% for each of the three planes. The 2D segmentation results in the coronal plane are slightly worse than those in the

Table 5

The contribution of each component in the fine segmentation stage, shown as mean ± standard deviation and [worst, best] in %. ↓ indicates that a smaller value represents a better performance.

	Dice score	Jaccard index	Volume difference ↓
3D CNN	75.0 ± 8.5 [55.6, 87.9]	60.8 ± 10.7 [38.5, 78.4]	42.7 ± 32.2 [103.4, 7.7]
3D CNN + LCCD	79.3 ± 7.2 [65.7, 90.6]	66.3 ± 9.9 [48.9, 82.9]	27.7 ± 17.9 [57.9, 7.7]
3D CNN + LCCD + 2D(Coronal)	82.4 ± 5.9 [69.4, 91.6]	70.4 ± 8.5 [53.2, 84.4]	16.1 ± 11.9 [38.6, 2.2]
3D CNN + LCCD + 2D(Sagittal)	83.0 ± 5.3 [72.3, 91.4]	71.2 ± 7.8 [56.7, 84.1]	17.8 ± 11.6 [41.1, 0.4]
3D CNN + LCCD + 2D(Axial)	83.0 ± 5.8 [73.0, 91.7]	71.4 ± 8.5 [57.5, 84.8]	13.8 ± 11.7 [**36.3, 0.0**]
3D CNN + LCCD + 2.5D	**84.4 ± 5.4 [74.1, 92.4]**	**73.4 ± 8.1 [58.9, 85.9]**	15.0 ± 10.3 [38.2, 0.5]

other two planes in terms of Dice score and Jaccard index. It may be because there are much more disconnections for 2D sliced pancreas annotations in the coronal plane. The performance is the best in the axial plane in terms of Dice score, Jaccard index, and Volume difference, which may be because the ground truth is usually labeled in the axial plane. Adding a 2D CNN at each of the three planes significantly improves the segmentation accuracy, with p-value less than 0.01 on Dice score and Jaccard index and p-value less than 0.05 on volume difference. The fused segmentation results are superior to those obtained from each of the three planes, with p-values less than 0.001 on Dice score and Jaccard index.

3.3.3. Refine segmentation

We use the axial plane slice in Fig. 5 to show the effect of the level-set post-processing. As shown in Fig. 8, the edge indicator in (b) can be easily obtained from the CT image in (a) through $g \triangleq \frac{1}{1+|\nabla G_\sigma * I|}$. The red curve in Fig. 8(c) denotes the zero level of ϕ_0, as defined in Eq. (4). This curve also represents the fine segmentation results in the axial plane in Fig. 5. The edge indicator g, which has been used in the length term $\mathcal{L}(\phi)$ and the area term $\mathcal{A}(\phi)$ in Eq. (6), can provide the edge information and drive the zero level of ϕ towards the edge. The zero level of the optimal curve ϕ^* is shown in Fig. 8(d) using the red curve. The region pointed with green arrows in Fig. 8(c) evolves dramatically because the length term $\mathcal{L}(\phi)$ and the area term $\mathcal{A}(\phi)$ provide a large gradient and the red curve gradually matches the edge. The corresponding region pointed with green arrows in Fig. 8(d) is smooth
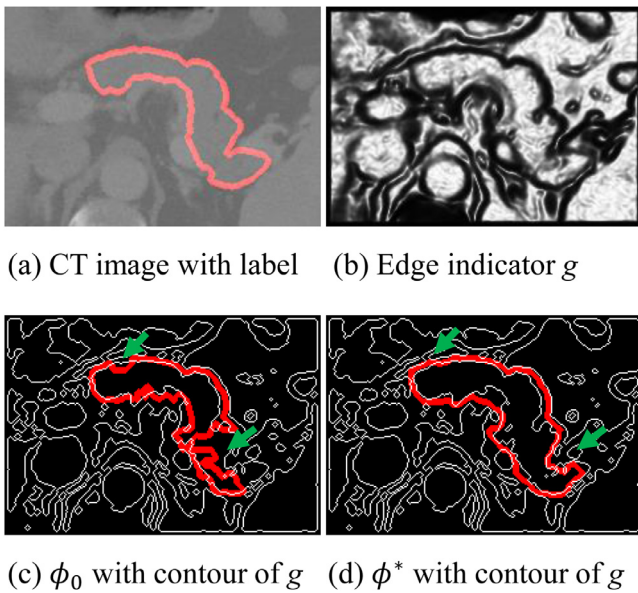

(a) CT image with label   (b) Edge indicator g

(c) ϕ_0 with contour of g (d) ϕ^* with contour of g

Fig. 8. A representative slice in level-set evolution. Red curve shows the contour $\mathcal{C}$ in the zero-level of ϕ_0 and ϕ^* and green arrows show the main difference between ϕ_0 and ϕ^*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6

Comparisons of different methods on the NIH dataset, in terms of Dice score [%] statistic the and computation time [minutes].

Methods	Mean ± STD	Max	Min	Testing time
Roth et al. (2015)	71.42 ± 10.11	86.29	23.99	6–8 min
Roth et al. (2016a)	78.01 ± 8.20	88.65	34.11	2–3 min
Zhou et al. (2017)	82.37 ± 5.68	90.85	62.43	–
Cai et al. (2017)	82.40 ± 6.7	90.1	60.0	–
The proposed	**84.47 ± 4.36**	**91.54**	**70.61**	3–5 min

and matches the edge and it also highly matches the ground truth (red curve in Fig. 8(a)).

In previous works (Li et al., 2010), level-set is typically applied to 2D segmentation by evolving curves. However, the contextual information in the 3D medical image is ignored in this slice-by-slice manner. To use the 3D information to constrain the search process, we conduct 3D level-set over the whole volume by evolving surfaces. As shown in Fig. 9, although 2D level-set and 3D level-set can obtain similar segmentation results in some slices (first row), 3D level-set is able to remove residual parts (second row) and restore missing parts (third row) in some situations when the CNN outputs do not localize the pancreas correctly.

3.4. Comparison with the state of the art

Comparisons between the proposed method and those previously-reported studies, in terms of Dice score over the NIH dataset, are tabulated in Table 6.

For data partitioning, we employ the same data partitioning approach as that in previous works (Roth et al., 2015; Cai et al., 2017), i.e., the 82 patients are randomly divided into four folds, respectively containing 21, 21, 20, and 20 samples. As for data augmentation, we apply random sampling in 3D CNN and random margin padding in 2D CNN as that in Zhou et al. (2017). To guarantee a fair comparison, the results reported in Table 6 are directly obtained from those previously published papers (Roth et al., 2015; 2016a; Zhou et al., 2017; Cai et al., 2017).

The proposed segmentation framework outperforms previous state-of-the-art results by about 2% (Zhou et al., 2017; Cai et al., 2017) in terms of the average Dice score. The relatively small standard deviation in Dice score indicates that our method is more stable and more robust. For challenging cases, the improvement of our method, in terms of the min Dice score, over the second-best (Zhou et al., 2017) is about 8% (from 62.53% to 70.61%). This clearly shows the effectiveness of our framework.

3.5. Evaluation on the MSD dataset

We further evaluate the proposed pipeline on the MSD dataset. For data partitioning, we first index all 281 cases according to their names and then divide them into four folds according to their indices. More specifically, the data having indices {1, 5, 9, ..., 281}, {2, 6, 10, ..., 278}, {3, 7, 11, ..., 279}, {4, 8, 12, ..., 280} are respectively used for testing as fold 1, fold 2, fold 3 and fold 4, and the others
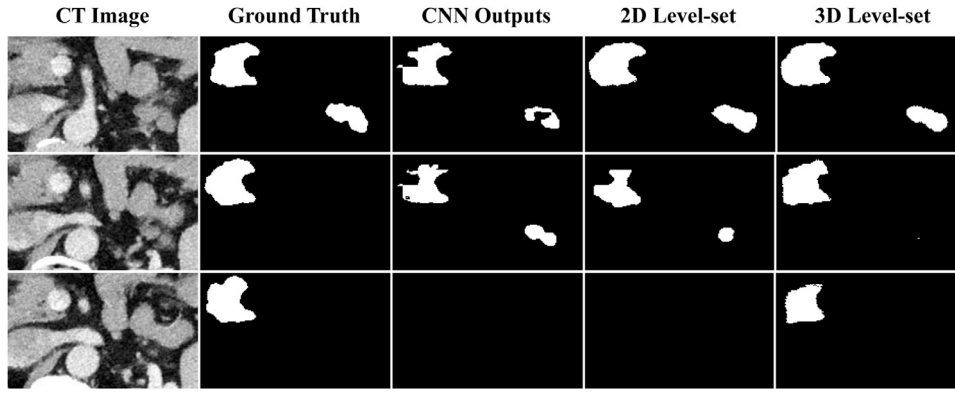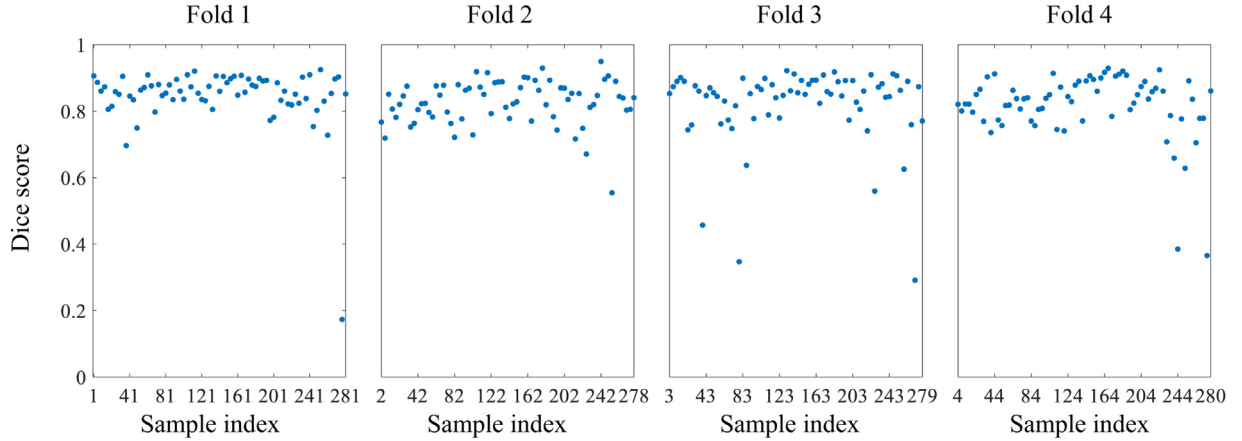
Fig. 9. Comparisons of 2D level-set and 3D level-set.

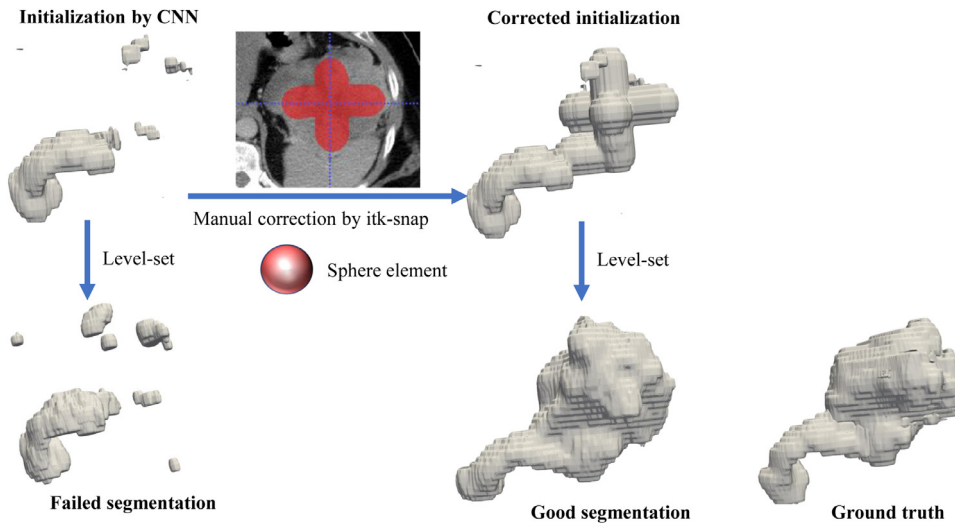**Fig. 10.** Dice score statistics on the MSD dataset through a four-fold cross-validation experiment.

Fig. 11. The worst result obtained from the proposed method on the MSD dataset and a potential correction stratery. There is a big hole in the CNN-derived segmentation with only a 16.31% Dice score (because of tumor). Directly applying 3D level-set evolution will not help. In this case, simply using a sphere element to manually repair the missing parts and then the proposed 3D level-set can significantly improve the semgentation result (a Dice score of 85%)

are used for training. The results on the MSD dataset are shown in Fig. 10. We observe that most cases have Dice scores above 70%, with the average Dice score being 82.74%.

We notice that there are a few cases (6 out of 281) having Dice scores lower than 50%. We would like to kindly point out that these failed cases can be easily fixed with very simple human intervention (Yushkevich et al., 2006), right before the refine segmentation step, such as the example shown in Fig. 11.

4. Discussion

The importance of each stage in our method has been rigorously identified based on the experimental results. They assembled effectively in the proposed pancreas segmentation framework. Firstly, the coarse segmentation is an important prior step for the subsequent CNN based segmentation. Multi-atlas based label fusion can effectively identify the rough location of the pancreas and

narrow the potential segmenting region of interest in the subsequent CNNs, especially when combined with the largest connected component analysis. Secondly, our joint 3D + 2D CNN strategy is novel and efficient. A single channel CNN may predict biased results when a majority of the training samples are the background (usually denoted as zeros) or when most of the voxels in one sample belong to the background. The last row in Fig. 7 shows that the segmentation results of purely 2D or purely 3D CNN are noisy. The sliced probability map obtained from the 3D CNN is used as the second channel in our 2D CNN and this makes the 2D CNN more concentrated on the pancreas region, as demonstrated in Fig. 7. If the missing part in the 3D CNN segmentation is large such as the example shown in Fig. 5, the 2D CNN can partially recover the missing part and level-set can further recover and smooth the segmentation results. Lastly, the DRLS method alone (Li et al., 2010) (i.e. without the CNN segmentation result serving as the initial value and the constraint) needs manually setting the initial value, which is subjective and may not perform well.

As shown in Table 6, the segmentation performance of a typical 3D CNN for pancreas is about 78% (Roth et al., 2015; 2016a). Later, a fixed-point coarse-to-fine strategy was proposed and the slice-based 2D CNNs with pre-trained initialized weights were used in both coarse stage and fine stage (Zhou et al., 2017). However, the CNN based coarse segmentation is not robust and some unsatisfying segmentation may be produced at the fine stage because the cropped region does not contain enough contextual information, whereas the coarse-scaled probability map carrying such information is discarded (Yu et al., 2018). Compared with these purely CNN based methods, the improvement of our framework can be contributed to three aspects: (1) diffeomorphic deformation based ROI detection, (2) joint 3D + 2D CNN, (3) level-set based post-processing. Please be noted that the original CT image has been used in each of all three stages, which can fully extract multidimensional information from the original input image. In other words, we jointly utilize global 3D location information (registration), contextual information (patch-based 3D CNN), shape information (slice-based 2.5D CNN) and edge information (level-set) in the proposed framework. These strategies collectively induce the effectiveness of the proposed pancreas segmentation pipeline.

5. Conclusion

In this study, we proposed and validated a coarse-to-fine-to-refine framework for automated pancreas segmentation. The proposed coarse segmentation can largely reduce the input size to a subsequent 3D CNN. The output of the 3D CNN provides a rough location for the subsequent 2-channel 2D CNN. The output of the 2.5D CNN provides an initial contour and constraint for level-set based segmentation used at the refine stage. These components constitute a cascaded framework for segmenting pancreas. In contrast to other coarse-to-fine methods (Zhou et al., 2017) relying mainly on slice-based CNN, we innovatively applied classical multi-atlas based fusion method to find a rough pancreas location and jointly used 3D patch- and 2D slice-based CNNs to ensure that our model can effectively learn both the local connection features and global shape features. The main results yielded from this study are summarized below.

1. The proposed framework yielded the highest Dice score, Jaccard index and lowest volume difference in the ISICDM challenge.
2. The proposed method is superior to state-of-the-art segmentation algorithms based on the same dataset (NIH) and four-fold cross-validation (Roth et al., 2015; Zhou et al., 2017; Cai et al., 2017).

3. The proposed registration based multi-atlas label fusion is an efficient way to narrow the input region to subsequent CNNs.
4. The proposed method can also work for data with tumor, such as the MSD dataset. The proposed pipeline is robust for most cases and a few failure cases can be easily fixed with very simple human intervention.

Further improvements to the proposed algorithm include the following.

(a) Despite that the proposed framework can segment a 3D CT image within 5 min. The multi-atlas label fusion costs 2 min even though we have used the fast LDDMM-image (Wu and Tang, 2019). Deep learning has been showing promising performance in registration and is very fast in the testing stage (Yang et al., 2017). Implementing LDDMM-image in a deep learning manner may largely improve the computational efficiency of the proposed pipeline.
(b) Currently, the 3D CNN and 2D CNN are trained independently. In the future, a strategy jointly optimizing 3D CNN and 2D CNN will be desired.
(c) A fusion of the segmentation results obtained from the three 2D CNNs at the three planes is used as the initial value and a constraint in our level-set postprocessing. The multi-constraint strategy may be used to adaptively combine the segmentations in the three planes. Shape prior can be used as another constraint when formulating the energy function in level-set, which is another research direction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Yue Zhang: Conceptualization, Methodology, Software, Writing - original draft. **Jiong Wu:** Conceptualization, Methodology, Software. **Yilong Liu:** Visualization, Investigation. **Yifan Chen:** Supervision, Funding acquisition. **Wei Chen:** Data curation. **Ed. X. Wu:** Supervision. **Chunming Li:** Software, Data curation. **Xiaoying Tang:** Supervision, Funding acquisition, Writing - review & editing.

Acknowledgments

References

Beg, M.F., Miller, M.I., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. 61 (2), 139–157.

Cai, J., Lu, L., Xie, Y., Xing, F., Yang, L., 2017. Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 442–450.

Ceritoglu, C., Tang, X., Chow, M., Hadjiabadi, D., Shah, D., Brown, T., Burhanullah, M.H., Trinh, H., Hsu, J., Ament, K.A., et al., 2013. Computational analysis of LDDMM for brain mapping. Front. Neurosci. 7, 151.

Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.-A., 2018. VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. NeuroImage 170, 446–455.

Ciompi, F., de Hoop, B., van Riel, S.J., Chung, K., Scholten, E.T., Oudkerk, M., de Jong, P.A., Prokop, M., van Ginneken, B., 2015. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. Med. Image Anal. 26 (1), 195–202.

Dolz, J., Desrosiers, C., Ayed, I.B., 2018. 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. NeuroImage 170, 456–470.

Dou, Q., Yu, L., Chen, H., Jin, Y., Yang, X., Qin, J., Heng, P.-A., 2017. 3D deeply supervised network for automated segmentation of volumetric medical images. Med. Image Anal. 41, 40–54.

Fang, C., Li, G., Pan, C., Li, Y., Yu, Y., 2019. Globally guided progressive fusion network for 3D pancreas segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 210–218.

Farag, A., Lu, L., Roth, H.R., Liu, J., Turkbey, E., Summers, R.M., 2017. A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling. IEEE Trans. Image Process. 26 (1), 386–399.

Gravel, P., Beaudoin, G., De Guise, J.A., 2004. A method for modeling noise in medical images. IEEE Trans. Med. Imaging 23 (10), 1221–1232.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.

Kronman, A., Joskowicz, L., 2016. A geometric method for the detection and correction of segmentation leaks of anatomical structures in volumetric medical images. Int. J. Comput. Assist. Radiol. Surg. 11 (3), 369–380.

Li, C., Xu, C., Gui, C., Fox, M.D., 2010. Distance regularized level set evolution and its application to image segmentation. IEEE Trans. Image Process. 19 (12), 3243–3254.

Li, H., Jiang, G., Zhang, J., Wang, R., Wang, Z., Zheng, W.-S., Menze, B., 2018. Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. NeuroImage 183, 650–665.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: International Conference on 3D Vision. IEEE, pp. 565–571.

Ngo, T.A., Lu, Z., Carneiro, G., 2017. 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. 35, 159–171.

Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M., Matrisian, L.M., 2014. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 74 (11), 2913–2921.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Roth, H.R., Lu, L., Farag, A., Shin, H.-C., Liu, J., Turkbey, E.B., Summers, R.M., 2015. Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 556–564.

Roth, H.R., Lu, L., Farag, A., Sohn, A., Summers, R.M., 2016. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 451–459.

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

Roth, H.R., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., Kim, L., Summers, R.M., 2016. Improving computer-aided detection using convolutional neural networks and random view aggregation. IEEE Trans. Med. Imaging 35 (5), 1170–1181.

Ryan, D.P., Hong, T.S., Bardeesy, N., 2014. Pancreatic adenocarcinoma. New Engl. J. Med. 371 (11), 1039–1049.

Saito, A., Nawano, S., Shimizu, A., 2016. Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs. Med. Image Anal. 28, 46–65.

Salah, M.B., Mitiche, A., Ayed, I.B., 2009. Effective level set image segmentation with a kernel induced data term. IEEE Trans. Image Process. 19 (1), 220–232.

Schima, W., Ba-Ssalamah, A., Kölblinger, C., Kulinna-Cosentini, C., Puespoek, A., Götzinger, P., 2007. Pancreatic adenocarcinoma. Eur. Radiol. 17 (3), 638–649.

Setio, A.A.A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., Van Riel, S.J., Wille, M.M.W., Naqibullah, M., Sánchez, C.I., van Ginneken, B., 2016. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans. Med. Imaging 35 (5), 1160–1169.

Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19, 221–248.

Shimizu, A., Kimoto, T., Kobatake, H., Nawano, S., Shinozaki, K., 2010. Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. Int. J. Comput. Assist. Radiol. Surg. 5 (1), 85.

Siegel, R.L., Miller, K.D., Jemal, A., 2017. Cancer statistics, 2017. CA 67 (1), 7–30.

Tang, X., Oishi, K., Faria, A.V., Hillis, A.E., Albert, M.S., Mori, S., Miller, M.I., 2013. Bayesian parameter estimation and segmentation in the multi-atlas random orbit model. PLoS One 8 (6), 1–14.

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W.E., Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Med. Imaging 22 (2), 137–154.

Van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., 2014. scikit-image: image processing in python. PeerJ 2, e453.

Wolz, R., Chu, C., Misawa, K., Mori, K., Rueckert, D., 2012. Multi-organ abdominal CT segmentation using hierarchically weighted subject-specific atlases. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 10–17.

Wu, J., Tang, X., 2019. A large deformation diffeomorphic framework for fast brain image registration via parallel computing and optimization. Neuroinformatics 28, 251–266.

Yang, M.-J., Li, S., Liu, Y.-G., Jiao, N., Gong, J.-S., 2013. Common and unusual CT and MRI manifestations of pancreatic adenocarcinoma: a pictorial review. Quant. Imaging Med. Surg. 3 (2), 113.

Yang, X., Kwitt, R., Styner, M., Niethammer, M., 2017. Quicksilver: fast predictive image registration–a deep learning approach. NeuroImage 158, 378–396.

Yu, L., Cheng, J.-Z., Dou, Q., Yang, X., Chen, H., Qin, J., Heng, P.-A., 2017. Automatic 3D cardiovascular mr segmentation with densely-connected volumetric convnets. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 287–295.

Yu, Q., Xie, L., Wang, Y., Zhou, Y., Fishman, E.K., Yuille, A.L., 2018. Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8280–8289.

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31 (3), 1116–1128.

Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E.K., Yuille, A.L., 2017. A fixed-point model for pancreas segmentation in abdominal CT scans. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 693–701.

Zhu, Z., Xia, Y., Shen, W., Fishman, E., Yuille, A., 2018. A 3D coarse-to-fine framework for volumetric medical image segmentation. In: International Conference on 3D Vision. IEEE, pp. 682–690.