PAPER

A two-stage data-model driven pancreas segmentation strategy embedding directional information of the boundary intensity gradient and deep adaptive pointwise parameters

To cite this article: Lu Tian *et al* 2023 *Phys. Med. Biol.* **68** 145005

IOP Publishing Phys. Med. Biol. 68 (2023) 145005 https://doi.org/10.1088/1361-6560/ace099

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
13 January 2023

REVISED
12 May 2023

ACCEPTED FOR PUBLICATION
21 June 2023

PUBLISHED
7 July 2023

PAPER

A two-stage data-model driven pancreas segmentation strategy embedding directional information of the boundary intensity gradient and deep adaptive pointwise parameters

Lu Tian[1,2], **Liwen Zou**[1] **and Xiaoping Yang**[1,*]

[1] Department of Mathematics, Nanjing University, Nanjing, 210093, People's Republic of China
[2] Department of Mathematics, Nanjing University of Science and Technology Zijin College, Nanjing, 210023, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: xpyang@nju.edu.cn

Keywords: adaptive pointwise parameter, pancreas segmentation, directional information, hybrid variational model

Abstract

In this paper, we propose a two-stage data-model driven pancreas segmentation method that combines a 3D convolution neural network with adaptive pointwise parametric hybrid variational model embedding the directional and magnitude information of the boundary intensity gradient. Firstly, nnU-net is used to segment the entire abdominal CT image with the aim of obtaining the region of the interest of pancreas. Secondly, an adaptive pointwise parametric variational model with a new edge term containing the directional and magnitude information of the boundary intensity gradient is used to refine the predicted results from CNN. Although CNN is good at extracting texture information, it does not capture weak boundary information very well. In order to well acquire more weak boundary information of the pancreas, we utilize not only the magnitude of the gradient, but also the directional information of the boundary intensity gradient to obtain more accurate results in the new edge term. In addition, the probability value for each pixel obtained by calculating the softmax function is exploited twice. Actually, it is applied firstly to generate the binary map as the initial contour of the variational model and then to design the adaptive pointwise weight parameters of internal and external area terms of the variational model rather than constants. It not only eliminates the trouble of manual parameter adjustment, but also, most importantly, provides a more accurate pointwise evolutionary trend of the level set contour, i.e. determine the tendency of the level set contour to pointwisely contract inward or expand outward. Our method is evaluated on three public datasets and outperformed the state-of-the-art pancreas segmentation methods. Accurate pancreatic segmentation allows for more reliable quantitative analysis of local morphological changes in the pancreas, which can assist in early diagnosis and treatment planning.

1. Introduction

According to the latest statistics from the National Cancer Center of China, pancreatic cancer ranks in the top 10 of the most prevalent cancers in China. What is even more frightening is that pancreatic cancer has the top 5 mortality rate which means that it has a very poor prognosis (Zheng *et al* 2022). Therefore, how to accurately segment the pancreas and the lesion in the abdominal CT images is a key step for doctors to treat and evaluate pancreatic cancer.

In this paper, we focus on segmenting pancreas from CT scan, which is one of the most difficult tasks among organ segmentation problems. As shown in figure 1, the main challenges are as follows: firstly, the pancreas often suffers from high variability in shape, size and location. Secondly, it occupies only a very small fraction of the entire abdominal CT image (e.g. less than 1.5% in a 2D image and less than 0.5% in a 3D image) (Zhou *et al* 2017) and is obscured by surrounding organs and tissues. Thirdly, the pancreas in a CT image has inhomogeneous

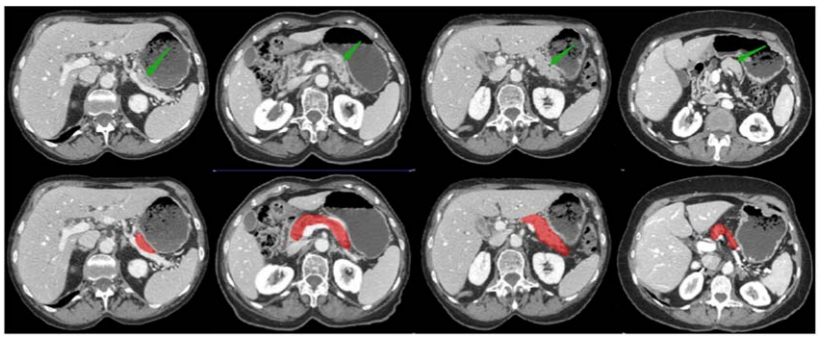

Figure 1. The entire abdominal CT image. The first row is the original abdominal CT images, and the second row is the corresponding CT images with the pancreas marked. It is evident that the size, shape and location of pancreas is variable.

intensity and blurred boundaries as well as the disturbance by noise or adjacent structures with intensity similar to pancreas (Farag *et al* 2016, Zhang *et al* 2021a).

Before the era of deep learning, the model-based methods play an important role in handling many segmentation tasks, which should still be paid more attention nowadays (Fortun *et al* 2018, Cai *et al* 2019, Zhang *et al* 2019). Among these methods, level set methods have been widely used to implement active contours for image segmentation applications (Kass *et al* 1988, Li *et al* 2005, 2010). Khadidos *et al* (2017) proposed a novel method based on a level set active contour model that provides more accurate boundary detection results around weak edges. The method combines edge intensity information with edge directional information collected from the adjacent region located inside and outside of the evolving contour, and is able to accurately drive the contour to the desired boundary even around weak edges. Min *et al* (2018) proposed a novel level-set method named local approximation of Taylor expansion (LATE) to segment images with intensity inhomogeneity that is based on first-order Taylor expansion. Since the first-order differential is represented by the variation degree of intensity inhomogeneity, LATE can improve the approximation quality and enhance the local intensity contrast of images with severe intensity inhomogeneity.

In recent years, due to the rapid development and impressive performance of deep learning, various deep convolution neural network (CNN) models have been successfully applied in medical image analysis. Zhou *et al* (2017) proposed a 2D fixed-point model which uses a predicted pancreas mask to shrink the input region, yields a dice score of 83.18% $\pm$ 4.81%. This was motivated by the fact that a smaller input region often leads to more accurate segmentation. Zhu *et al* (2018) proposed a novel 3D network called ResDSN integrated with a coarse-to-fine framework to simultaneously achieve high segmentation accuracy and low time cost, yields a dice score of 84.59% $\pm$ 4.86%. Zhang *et al* (2021b) proposed a coarse-fine-refine framework for automated pancreas segmentation which uses data-based method for fine segmentation and traditional model-based method for post-processing. Although these methods have been proven to be effective and improved the Dice score to some extent (Roth *et al* 2018, Yu *et al* 2018b, Man *et al* 2019, Xue *et al* 2019), there is still room for improvement about pancreas segmentation.

It is well known that edges and textures contribute different information to image understanding and analysis. Edges encode shape information, while textures determine region information. CNNs have proven to be effective at representing and classifying texture information (Hatamizadeh *et al* 2019). However, Geirhos *et al* (2018) empirically demonstrated that common CNN architectures are biased towards recognizing textures in the image, not object shape representations. Therefore, CNN predictions, in particular the segmentation results of the pancreas, often need to be refined to compensate for the edge details that the networks fail to learn during training.

To address above challenges, we propose a novel two-stage segmentation approach for the pancreas that is both data-driven and model-driven. Specifically, in the first stage, we first employ a 3D U-net to segment pancreas on the entire abdominal CT image to obtain an initial localization of pancreas (i.e. region of the interest (ROI) of pancreas). For one whole abdominal CT image, considering that most of the useless background regions outside of the pancreas and computational efficiency as well as hardware memory burden, the binary maps obtained from the 3D CNN are cropped into relatively small bounding boxes of pancreas which are then used as shape priors for the next stage. Then, in the second stage, we propose a hybrid variational model with adaptive pointwise parameters and the directional information of the boundary intensity gradient to refine the predicted results from the first stage, which leverages each binary map as the initial contour to guide the evolution of the level set curve to pointwisely contract inward or expand outward. Our proposed edge energy term takes into account not only the magnitude of the gradient, but also the directional information of the

boundary intensity gradient, for more details see section 3.2. We expect that our approach can meet the clinical needs of physicians by providing more accurate boundaries of the pancreas and more reliable quantitative analysis of local morphological changes in the pancreas, which is necessary for observing lesions, analyzing anatomical structures of the pancreas, and predicting patient prognosis (Lim *et al* 2022).

Main contributions of our paper are summarized as follows.

(1) In the first stage, we propose a 3D CNN for coarse segmentation of pancreas to alleviate the issue of highly variable and irregular anatomy of the pancreas. We adopt the method of data pre-processing in nnU-net (Isensee *et al* 2021), which achieves higher accuracy compared to other fancy network architectures. Therefore, we can extract local, global and spatial information of pancreas from CT images. On one hand, the binary maps generated by calculating the softmax function are used to determine pancreas' (ROI), which is concatenated with the original CT image of the same size and subsequently fed into the following variational model to refine. In addition, the purpose of cropping the whole abdominal CT image into ROI is to eliminate redundant computation cost and achieve more accurate segmentation by decreasing the false positive probability. On the other hand, the probability values obtained by calculating the softmax function are used to design adaptive pointwise weight parameters about region terms of our variational model in the next stage.

(2) In the second stage, we propose a new hybrid variational model to capture the pancreas' weak boundary and provide the pointwise power for more accurate evolutionary trend of the level set contour. In the new edge term of the proposed model, we utilize not only the magnitude of the gradient, but also the directional information of the boundary intensity gradient to extract more accurate edge information. Additionally, we embed the probability value for each pixel obtained by calculating the softmax function to design adaptive pointwise weight parameters, which go beyond manual adjustment in the two region terms of our model and make it possible to determine the tendency of the level set curve to pointwisely expand outward or contract inward more precisely. It not only eliminates the trouble of manual parameter adjustment, but also solves the problem of overfitting and underfitting of the neural network results simultaneously.

The rest of this paper is organized as follows. In section 2, we briefly introduce some representative level set models which are related to our proposed model. In section 3, we present a detailed description of our proposed two-stage data-model driven pancreas segmentation strategy. In section 4, the experimental setup and designs are provided. In section 5, we demonstrate the performance of our proposed method through different experiments. Conclusions and discussions are given in section 6.

2. Preliminary

For an image in the domain Ω, $I(x)$ is its intensity, where $x \in \Omega$ is a point of the image. An implicit contour of pancreas can be represented as the zero level set ϕ: $\Omega \to R$, noted by $C = \{x \in \Omega | \phi(x) = 0\}$, assuming that the embedding level set function ϕ takes positive values inside the zero level set contour and negative values outside.

One of the most typical region-based level set models is proposed by Chan and Vese (2001), called the CV model. The energy functional can be expressed as

$$\begin{aligned} E_{\mathrm{CV}}(\phi) = \alpha_1 \int_\Omega (I(x) - c_1)^2 H(\phi)\, dx \\ + \alpha_2 \int_\Omega (I(x) - c_2)^2 (1 - H(\phi))\, dx, \end{aligned} \tag{1}$$

where c_1 and c_2 are two constants that approximate the image intensity inside and outside the zero level set contour, respectively. α_1 and α_2 are the weight parameters. H is the Heaviside function.

It is noted that the CV model is a piecewise constant model which is not so suitable for images with intensity inhomogeneity. To cope with this problem, Li *et al* (2008) proposed the region-scalable fitting (RSF) model. Its energy functional is expressed as

$$\begin{aligned} E_{\mathrm{RSF}}(\phi) = \beta_1 \int_\Omega \int_\Omega K_\sigma(x - y) |I(y) - f_1(x)|^2 H(\phi(y))\, dy dx \\ + \beta_2 \int_\Omega \int_\Omega K_\sigma(x - y) |I(y) - f_2(x)|^2 (1 - H(\phi(y)))\, dy dx, \end{aligned} \tag{2}$$

where β_1 and β_2 are the weight parameters. $K_\sigma(x) = \dfrac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{|x|^2}{2\sigma^2}}$ is the kernel function which is a decreasing function. $f_1(x)$ and $f_2(x)$ are image intensities inside and outside the zero level contour at the point x, respectively. Due to a kernel function in the data fitting term, intensity information in local regions is extracted to guide the motion of the contour, which thereby enables this model to cope with intensity inhomogeneity.

One of the most typical edge-based level set models is proposed by Li *et al* (2005), that eliminates the need of re-initialization procedure (without re-initialization). The energy functional can be formulated as

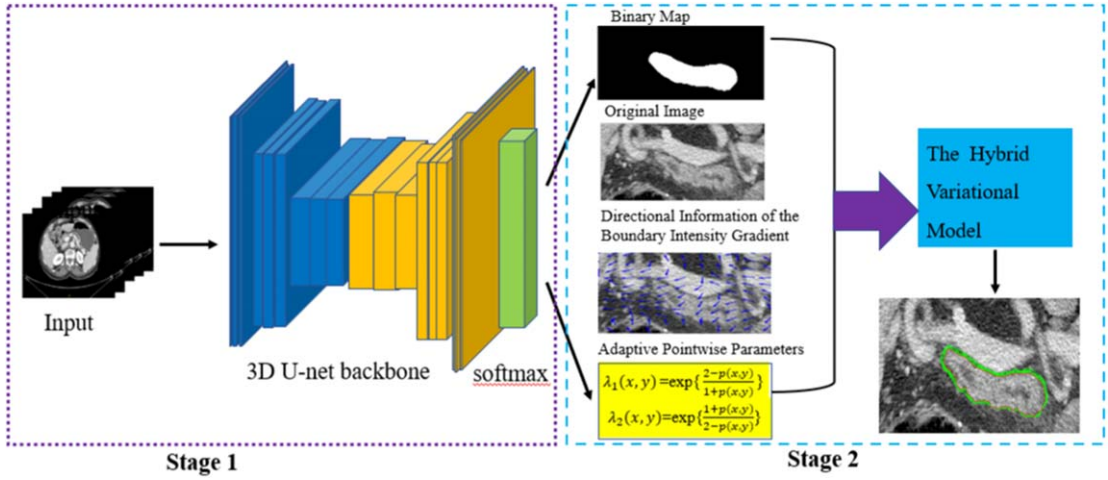

Figure 2. Flowchart of our proposed two-stage data-model driven pancreas segmentation framework. In stage 1, the network architecture employs a typical encoder–decoder network (3D U-Net) as the backbone. In stage 2, an adaptive pointwise parametric variational model with a new edge term containing the directional and magnitude information of the boundary intensity gradient is used to refine the predicted results from CNN. The green and red contours denote the zero level set of the predicted and the ground truth, respectively.

$$E_{WORE}(\phi) = \rho_1 \int_\Omega g(|\nabla I|)\delta(\phi)|\nabla\phi|dx + \rho_2 \int_\Omega g(|\nabla I|)H(\phi)dx + \rho_3 \int_\Omega \frac{1}{2}(|\nabla\phi| - 1)^2 dx, \quad (3)$$

where $g(|\nabla I|) = \frac{1}{1+|\nabla I|^2}$ is an edge detection function. δ is the Dirac delta function, which is the derivative of H. The energy functional can also be used to ensure the smoothness of the level set evolution and the regularity of the contour.

3. Method

In this section, we will elaborate our proposed two-stage framework which is illustrated in figure 2.

3.1. 3D convolutional neural network

Our backbone network is nnU-net (Isensee *et al* 2021) that obtains state-of-the-art results in 19 segmentation challenges with carefully designed U-net (Ronneberger *et al* 2015). nnU-net (no-new-U-net) is actually not a new network architecture (just like the U-net), but a medical image segmentation framework that integrates multiple tricks and data pre-processing and post-processing. It is demonstrated that image pre-processing, network topology and post-processing are more important than the network structure and that these most important parts of the nnU-net framework can be determined automatically (adaptive tuning). Without manually tuning, nnU-net can achieve better performances than most specialized deep learning pipelines (such as the use of residual connections, dense connections or attention mechanisms) in 11 international biomedical image segmentation challenges comprising 23 different datasets and 53 segmentation tasks (https://cremi.org).

Just like the 3D U-net, we follow the encoder–decoder pattern with skip connections. We use two plain convolutional layers between poolings in the encoder and transposed convolution operations in the decoder, and these convolutional layers followed by a ReLU. The cross-entropy loss and the Dice-coefficient loss are employed for backpropagation. Moreover, at the end of this network, we employ a convolution with the kernel size of $1 \times 1 \times 1$, which designs only two channels of feature maps corresponding to the foreground and background. The probability values are obtained by calculating the softmax function, which will be used to generate binary map as the initial contour of the proposed variational model and also can be used to design the adaptive pointwise weight parameters of internal and external area terms of the variational model.

In addition, considering that most of the useless background regions outside of the pancreas in the whole abdominal CT image, the predictions obtained from the 3D U-net are expanded into rectangular bounding boxes containing pancreas. Then we crop the original CT image into the same size as ROI, which will be fed into our variational model together. Hence, on the one hand, we can extract local information about pancreas; on the other hand, we can reduce computing cost and hardware memory.

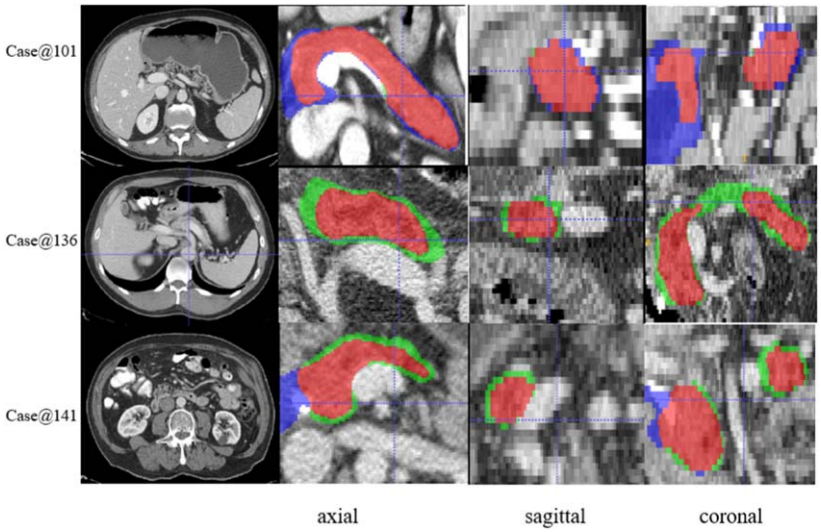

Figure 3. The different rows represent the segmentation results of the 3D U-net for different cases. The first column is the original image, The second to fourth columns represent the segmentation results in the axial, sagittal and coronal planes, respectively. Red indicates correct-fitting, green indicates over-fitting, blue indicates under-fitting. It can be observed that neural network method does not perform well for pancreas segmentation at the boundary.

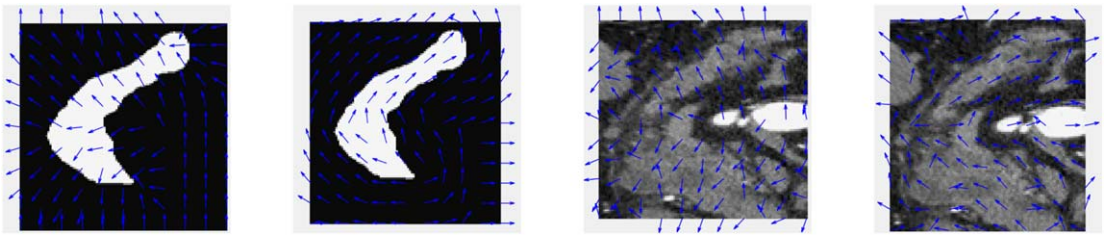

Figure 4. The two figures on the left show the gradient direction and the direction perpendicular to the gradient at the boundary of pancreas mask respectively. The two figures on the right show the gradient direction and the direction perpendicular to the gradient at the boundary of original image of pancreas respectively. We can observe that the normal vector at the boundary of pancreas is aligned with the gradient direction of the gray scale.

3.2. The hybrid variational model embedding the directional information of the boundary intensity gradient and deep adaptive pointwise parameters

As mentioned previously, although the above CNN can successfully extract texture information, it can not effectively extract the edge information of some target images, especially those of weak or (and) blurred boundary object images, see figure 3. In order to overcome this difficulty, further intensity information associated to the target boundaries should be captured. Therefore, we propose a new edge term embedding the directional and magnitude information of the boundary intensity gradient as follows:

$$E_{\mathrm{DI}}(\phi) = \oint_C g(|\nabla I|)|\langle \nabla I, \vec{N}\rangle| ds, \tag{4}$$

where $\vec{N}$ is the outward normal vector of the boundary curve C. Note that the energy of GAC model $E_{\mathrm{GAC}} = \oint_\alpha g(|\nabla I|)ds$ only considers locally magnitude of the gradient, which is suitable for images with clear boundaries and has bad performance about weak edges (Caselles *et al* 1997). Kimmel and Bruckstein (2003) proposed the fundamental edge integration problem for object segmentation in a geometric variational framework (Law and Chung 2010, Peyré *et al* 2010, Benmansour and Cohen 2011). Inspired by Kimmel and Bruckstein (2003), our proposed edge energy term E_{DI} takes into account not only the magnitude of the gradient, but also the directional information of the boundary intensity gradient. In fact, the gradient direction of the intensity value at the edge is the direction of the maximum intensity change. While the direction of the maximum intensity change should be from the interior of the pancreas to the exterior, which is also the normal direction of the level set function curve, as shown in figure 4. Therefore, we let the normal vector at the edge of pancreas align with the gradient direction of the intensity, and thus effectively extract the edge information.

Therefore, we propose an energy function $E(\phi)$ by

$$\begin{aligned} E(\phi) = & \int_\Omega \int_\Omega \lambda_1(x, y) K_\sigma(x - y)|I(y) - f_1(x)|^2 H(\phi(y)) dydx \\ & + \int_\Omega \int_\Omega \lambda_2(x, y) K_\sigma(x - y)|I(y) - f_2(x)|^2 (1 - H(\phi(y))) dydx \\ & + \nu \int_\Omega \frac{1}{2}(|\nabla\phi| - 1)^2 dx + \gamma \int_\Omega g(|\nabla I|) H(\phi) dx \\ & + \omega \int_\Omega g(|\nabla I|)\delta(\phi)|\nabla\phi| dx + \eta \oint_C g(|\nabla I|)|\langle \nabla I, \vec{N}\rangle| ds, \end{aligned} \tag{5}$$

where the adaptive pointwise weight parameters $\lambda_1(x, y)$ and $\lambda_2(x, y)$ are based on the results of the previous CNN. In fact, the binary map of pancreas generated by the previous CNN is obtained by comparing the last output probability value $p(x, y)$ of each pixel (x, y) with the threshold 0.5. Specifically, if $p(x, y)$ is greater than 0.5, then it takes 1, and if $p(x, y)$ is smaller than 0.5, then it takes 0. However, for these pixels whose probability values $p(x, y)$ are close to 0.5, the above processing scheme is not accurate enough, which is one of the reasons for over-fitting and under-fitting in the first stage. In order to provide a more accurate pointwise evolutionary trend of the level set contour, we design adaptive pointwise weight parameters $\lambda_1(x, y)$, $\lambda_2(x, y)$ instead of constants λ_1, λ_2, which is also inspired by the approach of Hatamizadeh (2020)

$$\begin{aligned} \lambda_1(x, y) &= \exp\left(\frac{2 - p(x, y)}{1 + p(x, y)}\right) \\ \lambda_2(x, y) &= \exp\left(\frac{1 + p(x, y)}{2 - p(x, y)}\right). \end{aligned} \tag{6}$$

If $0.5 \ll p(x, y) < 1$, it indicates that the pixel belongs to the interior of the pancreas region and $\lambda_1(x, y) < \lambda_2(x, y)$, then the driving force outside the region is greater, which will make the level set curve expand outward, so as to reach the target boundary.

If $0 < p(x, y) \ll 0.5$, it indicates that the pixel belongs to the exterior of the pancreas region and $\lambda_1(x, y) > \lambda_2(x, y)$, then the drving force inside the region is greater, which will make the level set curve contract inward, so as to reach the target boundary.

If $p(x, y) \approx 0.5$, it indicates that the pixel is near the boundary of pancreas and $\lambda_1(x, y) \approx \lambda_2(x, y)$, then the driving force inside and outside the region is the same, and the level set curve can be stabilized at the boundary of pancreas without too much expansion or contraction.

Minimizing the energy function $E(\phi)$ with respect to ϕ, we derive the gradient descent flow:

$$\begin{aligned} \frac{\partial\phi}{\partial t} = & -\delta_\varepsilon(\phi)(\lambda_1(x, y)e_1 - \lambda_2(x, y)e_2) - \gamma g(|\nabla I|)\delta_\varepsilon(\phi) \\ & + \nu\left(\nabla^2\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right) + \mu g(|\nabla I|)\delta_\varepsilon(\phi) div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) \\ & + \omega\delta_\varepsilon(\phi) g(|\nabla I|) sign\left(\left\langle \nabla I, \frac{\nabla\phi}{|\nabla\phi|}\right\rangle\right) \triangle I, \end{aligned} \tag{7}$$

where

$$\begin{aligned} e_i &= \int_\Omega K_\sigma(y - x)|I(x) - f_i(y)|^2 dy, \\ f_i(x) &= \frac{K_\sigma(x)*[M_i(\phi(x))I(x)]}{K_\sigma(x)*M_i(\phi(x))} (i = 1, 2), \\ M_1(\phi) &= H(\phi),\ M_2(\phi) = 1 - H(\phi). \end{aligned} \tag{8}$$

In addition, in practice, the Heaviside function H is approximated by a smooth function H_ε defined by

$$H_\varepsilon(x) = \frac{1}{2}\left[1 + \frac{2}{\pi} arctan\left(\frac{x}{\varepsilon}\right)\right]. \tag{9}$$

The Dirac function δ in equation (5) is approximated by the derivative of H_ε

$$\delta_\varepsilon(x) = H'_\varepsilon(x) = \frac{1}{\pi}\frac{\varepsilon}{\varepsilon^2 + x^2}. \tag{10}$$

4. Experiments

4.1. Dataset and evaluation measures

In this paper, we will use three public datasets and a private dataset of pancreas segmentation for training and validating the proposed 3D CNN. The detailed introduction is as follows:

(1) AbdomenCT-1K dataset: it consists of 1112 3D CT scans from five existing datasets: LiTS (201 cases), KiTS (300 cases), MSD Spleen (61 cases) and Pancreas(420 cases), NIH Pancreas (80 cases), and a new dataset from Nanjing Drum Tower Hospital (50 cases). It is the largest abdominal CT organ segmentation dataset, which includes multi-center, multi-phase, multi-vendor, and multi disease cases. We randomly selected 200 cases from this dataset. In addition, since this dataset contains multi-organs, we strip out the annotations of other organs before using it and keep only the annotations of pancreas (Ma *et al* 2021).

(2) NIH Pancreas-CT dataset: it performs 82 abdominal contrast enhanced 3D CT scans from subjects of different genders. The CT scans have resolutions of 512×512 pixels with varying pixel sizes and slice thickness between 1.5 ± 2.5 mm.

(3) MSD Pancreas-CT dataset: it contains 281 CT scans with labeled pancreas masks, with pancreas tumor annotations as well. The CT scans have resolutions of 512×512 pixels with varying pixel sizes and slice thickness about 2.5 mm.

(4) CM50 Pancreas-CT dataset: it contains 50 CT scans from Jiangsu Province Hospital of Chinese Medicine, with labeled pancreas masks and pancreas tumor annotations as well. The CT scans have resolutions of 512×512 pixels with varying pixel sizes and slice thickness between 1.5 ± 5 mm.

In addition, we use four metrics to quantitatively evaluate the segmentation performance. Let G, R denote the ground truth and the segmentation result respectively.

(1) Dice coefficient: Dice, is used to evaluate the region overlapping ratio of predicted result and ground truth masks

$$\text{Dice} = \frac{2|R \cap G|}{|R| + |G|}. \tag{11}$$

(2) Jaccard index: Jaccard, is similar to Dice, evaluates the overlapping ratio

$$\text{Jaccard} = \frac{|R \cap G|}{|R \cup G|}. \tag{12}$$

(3) Hausdorff distance: HD, is used to determine the maximum error margin

$$H(G, R) = \max\{h(G, R), h(R, G)\}, \tag{13}$$

where

$$\begin{aligned} h(G, R) &= \max_{a \in G}\{\min_{b \in R} \|a - b\|\}, \\ h(R, G) &= \max_{b \in R}\{\min_{a \in G} \|b - a\|\}. \end{aligned} \tag{14}$$

(4) Volume difference: VD, is used to evaluate the volumetric similarity between predicted result and ground truth masks

$$\text{VD} = \frac{||V_R| - |V_G||}{|V_G|}. \tag{15}$$

Thus, for Dice and Jaccard, higher scores indicate better segmentation performance. For HD and VD, lower scores indicate better segmentation performance.

4.2. Experimental design

We design four kinds of experiments to evaluate our proposed two-stage data-model driven pancreas segmentation strategy.

Experiment 1: result comparison between stages. We show and compare the results between stages of our two-stage pancreas segmentation scheme separately to analyze the contribution of each stage. These experiments are performed on the AbdomenCT-1K dataset.

Experiment 2: ablation study. To illustrate the effect of directional information term and adaptive pointwise parameters choice strategy in our model, we conduct three ablation experiments with different term combinations to demonstrate their performance. These experiments are also performed on the AbdomenCT-1K dataset.

Experiment 3: comparison with state-of-the-art methods. We compare different methods to validate the contributions of our proposed two-stage data-model driven approach. We compare our approach on the common public dataset of pancreas with the state-of-the-art methods of pancreas segmentation.

Experiment 4: inter-institutional study. We use the proposed two-stage strategy to train on AbdominCT-1K dataset and test on CM50 dataset.

4.3. Experimental details

In stage 1, we adopt the default training protocols of nnU-net. The initial learning rate takes 0.01 and is updated by the polyLR learning rate (Chen *et al* 2017, Ma *et al* 2020a, 2020b). The optimizer is stochastic gradient descent with a momentum (0.99). All models are trained for 1000 epochs and batch size 2. Each model is trained on one NVIDIA TITAN V100 GPU with 2–3 d.

In stage 2, the equation (5) is solved by the iterative process of the gradient descent algorithm as

$$\phi^{n+1} = \phi^n + \Delta t \frac{\partial \phi}{\partial t}, \tag{16}$$

where ϕ^n is the level set function at the nth iteration, and Δt is the time step. We use the following parameters in our variational model of different experiments: $\Delta t = 0.4$, $\nu = \eta = 0.04$, $\sigma = 3.0$, $\gamma = 3.0$, $\varepsilon = 2$ and $\omega = \mu = 0.004 * 255 * 255$, as Li *et al* (2010). λ_1, λ_2 are set as discussed in section 3.2.

5. Experiment results and discussion

5.1. Experiment 1: result comparison between stages

In figure 5, we show the 3D segmentation results obtained from each stage of our proposed strategy. It is clear that the performance of segmentation results of 3D U-net are improved after refining by our proposed variational model. Additionally, as seen from the second column of figure 5, the segmentation results of stage 1 have both under-fitting and over-fitting at the boundaries (green indicates over-fitting, blue indicates under-fitting, and red indicates correct fitting). Some regions show more under-fitting,while some regions show more over-fitting. Therefore, in order to alleviate both under-fitting and over-fitting problems at the edge of the pancreas, we propose a new edge term that takes into account direction information of the boundary intensity gradient, and region terms with adaptive pointwise parameters in the variational model. As seen from the third column of figure 5, both under-fitting and over-fitting problems have been improved to some extent. We select three cases with a high increase in Dice score. But it is noted that not every case has a higher Dice score in the second stage than in the first stage. When there is a large error in the prediction of the first stage, it may provide the wrong shape prior and adaptive pointwise parameters to the second stage, resulting in a lower Dice score in the second stage. Table 1 shows quantitative indexes of two stages in our segmentation strategy. The average Dice score, Jaccard index, Hausdorff distance, volume difference of 3D U-net segmentation results in stage 1 are 87.56%, 78.67%, 14.98% and 11.09%, respectively. After refining by our variational model in stage 2, these metrics yield to 89.61%, 80.22%, 11.94% and 7.91% respectively, leading to approximately 2% improvement.

In addition, our segmentation results are also better than those of Ma *et al* (2021). In Ma *et al* (2021), they select 225 cases from the MSD Pan(225) dataset for training to segment the pancreas and then test separately on MSD Pan(56), LiTS(131), KiTS(210) and Spleen(41) datasets which are all contained in AbdomenCT-1K dataset, yielded Dice scores of 86.1% $\pm$ 6.59%, 86.6% $\pm$ 12.2%, 80.9% $\pm$ 10.5% and 86.6% $\pm$ 8.80%, respectively. While we use a five-fold cross-validation training strategy on 200 randomly selected cases from AbdomenCT-1K dataset, yield a Dice score of 87.56% $\pm$ 6.58% in stage 1 and 89.61% $\pm$ 5.70% in stage 2.

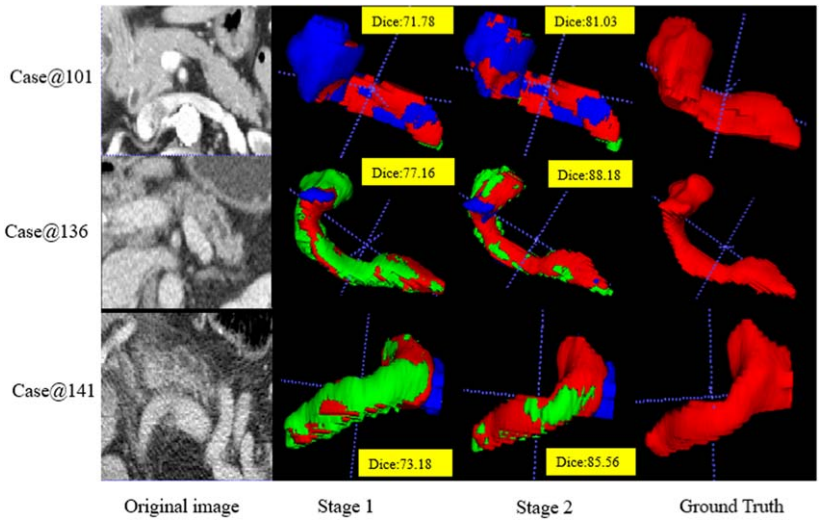

Figure 5. Three segmentation examples from AbdomenCT-1K dataset. The first to fourth columns represent the original images, segmentation results of stage 1, segmentation results of stage 2 as well as the corresponding ground-truths.

Table 1. The contribution of each stage in the proposed segmentation framework on AbdomenCT-1K dataset.

Method	Dice ↑	Jaccard ↑	HD ↓	VD ↓
Stage 1	87.56 ± 6.58	78.67 ± 8.89	14.98 ± 7.73	11.09 ± 12.04
Stage 2	89.61 ± 5.70	80.22 ± 8.36	11.94 ± 6.91	7.91 ± 9.88

Table 2. Comparisons of the individual effects of different terms on AbdomenCT-1K dataset ($p < 0.01$).

Method	Dice ↑	Jaccard ↑	HD ↓	VD ↓
$E(\phi)$ w.o. DI	88.15 ± 6.30	78.96 ± 8.26	13.51 ± 7.01	9.69 ± 9.05
$E(\phi)$ w.o. APP	87.94 ± 5.71	78.73 ± 8.35	13.97 ± 6.20	10.53 ± 9.63
$E(\phi)$ w.o. APP and DI	87.43 ± 6.33	77.94 ± 8.82	14.66 ± 8.70	11.67 ± 11.46

5.2. Experiment 2: ablation study

In order to analysis the individual impact of different terms in equation (5), we use hybrid variational models with different term combinations to refine predicted results on AbdomenCT-1K dataset with five-fold cross-validation. Table 2 summarizes the quantitative results, and figure 6 shows some examples of pancreas segmentation results.

(1) Effectiveness of directional information: without incorporating directional information, the performance drops on AbdomenCT-1K dataset, leading to decrease 1.46% in Dice, 1.26% in Jaccard, and increase in 1.57% in HD, 1.78% in VD.

(2) Effectiveness of adaptive pointwise parameter weights: without using adaptive pointwise parameters, it leads to larger performance drop in all four metrics, which highlights the importance of adaptive pointwise parameters. From figure 6, it can be found that the segmentation results would have large boundary errors and obvious segmentation outliers when using constant parameters. It means that adopting adaptive pointwise parameters can obtain more accurate results in boundaries and fewer segmentation errors.

(3) Effectiveness of directional information and adaptive pointwise parameters: it can be observed that the model without the directional information and adaptive pointwise parameters leads to the worst performance, indicating that both the proposed directional information and adaptive pointwise parameters are important to obtain performance improvements.

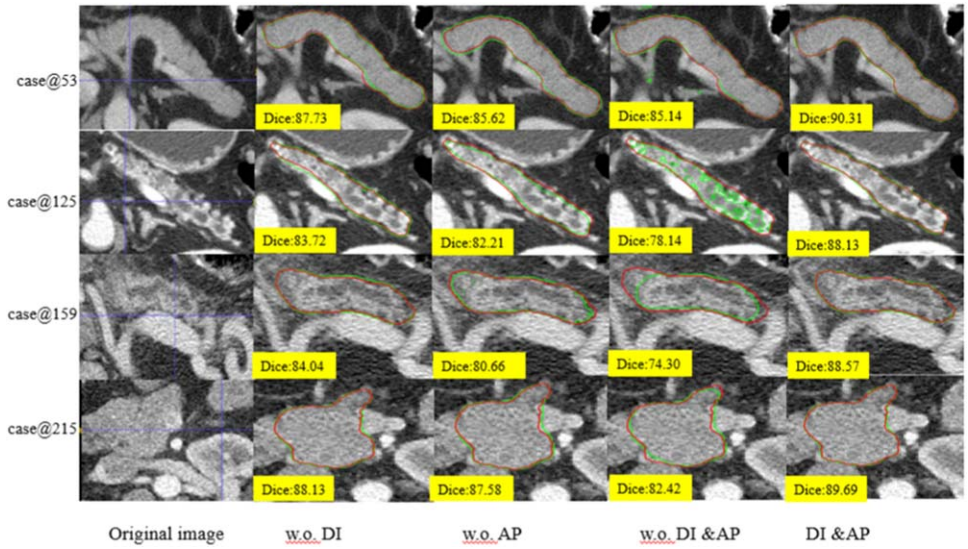

Figure 6. Four segmentation examples from AbdomenCT-1K dataset. The first to fifth columns represent the original images, segmentation results without directional information (w.o.DI), segmentation results without adaptive pointwise parameters (w.o. APP), segmentation results without directional information and adaptive pointwise parameters (w.o.DI and APP) as well as segmentation results with directional information and adaptive pointwise parameters (DI and APP). The red and green contours represent ground truths and the predicted segmentation results, respectively. We can observe that it leads to larger performance drop and has larger marginal errors without using adaptive pointwise parameters which means that the effectiveness of adaptive pointwise parameters is greater than that of directional information.

Table 3. Comparisons of different methods on the NIH dataset, in terms of Dice score statistic.

Method	Mean Dice	Max Dice	Min Dice
Zhou *et al* (2017)	82.37 ± 5.68	90.85	62.43
Zhu *et al* (2018)	84.59 ± 4.86	91.45	69.62
Hu *et al* (2020)	85.49 ± 4.77	91.64	67.19
Zhang *et al* (2021b)	85.8 ± 4.80	91.54	70.61
Chen *et al* (2022)	86.38 ± 3.18	92.28	74.63
Lim *et al* (2022)	87.00 ± 7.40	—	—
Huang and Wu (2022)	82.87 ± 1.00	—	—
The proposed	87.67 ± 5.70	92.43	74.13

Table 4. Comparisons of different methods on the MSD dataset, in terms of Dice score statistic.

Method	Mean Dice	Max Dice	Min Dice
Yu *et al* (2018a)	82.16	91.37	66.53
Zhang *et al* (2021b)	82.74	—	—
Zhang *et al* (2021a)	85.56	93.01	69.16
Chen *et al* (2022)	84.97	91.87	67.33
The proposed	87.13	95.93	69.57

5.3. Experiment 3: comparison with state-of-the-art methods

We compare our proposed method with the state-of-the-art methods about pancreas segmentation in term of Dice score on two popular pancreas-CT public datasets, i.e. NIH and MSD. In table 3, it is observed that the proposed segmentation method outperforms those state-of-the-art results by about 2%–5% on NIH dataset. Meanwhile, compared our results with the second-best results, the min Dice score is improved by about 1% and the max Dice score is improved by about 4%. Furthermore, in table 4, we evaluate the proposed method on MSD dataset. The results demonstrate the effectiveness of our proposed method.

Meanwhile, we conduct three new experiments using the recent cascaded or hybrid deep learning-based models (3D cascade U-net, UNETR and ResU-Net respectively), then use our proposed variational model to

Table 5. The comparisons of different methods on different datasets, in terms of Dice score statistic ($p < 0.01$).

		3D Cascade U-net		The proposed	
Dataset	Dice(stage1:coarse)	Dice(stage1:fine)	Dice(stage2)	Dice(stage1)	Dice(stage2)
AbdomenCT-1K	85.56	85.70	87.89	87.56	89.61
NIH	85.49	85.63	87.21	86.39	87.67
MSD	80.20	80.33	83.37	85.71	87.13

Table 6. The comparisons of different methods on the NIH dataset, in terms of Dice score statistic ($p < 0.01$).

Method	Dice(stage 1)	Dice(stage 2)
UNETR	82.33	85.06
ResU-Net	85.76	86.95
The proposed	86.09	87.67

Table 7. The performance of different datasets, in terms of Dice score statistic ($p < 0.01$).

Dataset	Dice(stage 1)	Dice(stage 2)
CM50	88.75	90.72
AbdominCT-1K	87.56	89.61

refine the prediction results of the three models. The framework of 3D cascade U-net combining the variational model we proposed is divided into two stages. Firstly, a low-resolution 3D U-net is used to coarsely segment the whole abdominal CT images with the aim of obtaining the location of pancreas (i.e. ROI of the pancreas). Secondly, a high-resolution 3D U-net is used to further get the result of fine segmentation from that of coarse segmentation and the original ROI of the pancreas as 2-channel inputs for this stage. Thirdly, we use a novel edge term in a hybrid variational model with adaptive parameters to recover the missing contents especially the edge parts and enhance the quality of boundaries for post-processing. In table 5, it can be observed that there is almost no increase in the Dice value in the fine stage compared to the coarse stage, which is why we do not choose a cascade network as backbone. In addition, the Dice values of the 3D cascade U-net are all lower than those of our proposed method on the three different datasets.

The framework of UNETR combining the variational model we designed is divided into two stages. Specifically, in the first stage, UNETR utilizes a transformer as the encoder to learn sequence representations of the input volume and effectively capture the global multi-scale information, while also following the successful U-shaped network design for the encoder and decoder (Hatamizadeh *et al* 2022). Then we use our proposed variational model to refine the prediction results of the UNETR in the second stage. Similar to UNETR, the framework of ResU-Net combining the variational model we design is also divided into two stages. In the first stage, we add residual connections in the 3D U-Net as backbone, then we use our proposed variational model to refine the prediction results of the ResU-Net in the second stage. In table 6, it is observed that the Dice values of UNETR and ResU-Net are all lower than those of our proposed method on the NIH dataset. It is demonstrated that image pre-processing, network topology and post-processing are more important than the network structure, which we believe is the reason why nnU-net performs well. It also shows that our two-stage data-model driven strategy is effective for dealing with the segmentation of small organs like the pancreas.

5.4. Experiment 4: inter-institutional study

We utilize nnU-net as backbone network to train on AbdominCT-1K dataset and test on another private dataset, namely CM50 dataset. Then, we use our proposed variational model to refine the prediction results tested on CM50 dataset. In table 7, it is demonstrated that our proposed two-stage approach is also feasible for multi-center datasets.

5.5. Limitations

We have successfully evaluated our proposed segmentation approach on three public datasets. But we admit that there are still some rooms for further improvements to our proposed method.

(1) In this study, the variational model is only acted on each 2D slice of the axial plane and then reconstructed into the 3D volume. Thus, some voxel information may be lost in this process. The variational model may be improved to 3D variational model in the next study.

(2) Our segmentation framework is to first segment pancreas using a 3D U-net and then use a variational model to refine missing edge parts, which is not end-to-end approach. We will incorporate the traditional active contour model into CNN-based segmentation architectures in an end-to-end way.

6. Conclusion

In general, this paper presents a two-stage pancreas segmentation method that combines both modern data-driven CNNs and classical model-driven active contour approaches to joinly extract local appearances, object global information, shape prior information and edge information. We firstly use 3D U-net to extract the local texture information of the pancreas, the location information as well as object global information in whole abdominal CT images. In order to extract more accurate shape representations (i.e. edge information), we embed the directional information of the gradient into the boundary term and select adaptive pointwise weight parameters relative to the region terms. It not only eliminates the trouble of manual parameter adjustment, but also provides a more accurate pointwise evolutionary trend of the level set contour. Thus, the novel variational model has an intrinsic ability of mitigating the trouble of weak boundaries and addressing the small segmentation errors, thereby extracts more accurate edge information. Different evaluations and comparisons on three public datasets indicate that the proposed method can lead to improvements in contrast to existing state-of-the-art methods. The traditional active contour approach is interpretable and can solve some specific problems. However, it has been criticized due to the trouble of tuning parameters. In this paper, we combine it with modern network methods to complement shortcomings of each other which is also a trend for further development of deep learning (Kobler *et al* 2017, Marcos *et al* 2018, Chen *et al* 2021).

Finally, our aim at proposing the two stage data-model driven pancreas segmentation strategy is to obtain a more accurate boundary of the pancreas for further understanding the changes of pancreatic morphology. Quantifying such changes aids in understanding the spatial heterogeneity of the pancreas and assists in the diagnosis and treatment planning of pancreatic cancer.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.12090023 and No.11971229) and Ministry of Science and Technology of China (No. 2020YFA0713800). The authors appreciate the anonymous reviewers for their great suggestions to improve this paper.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/ https://github.com/JunMa11/AbdomenCT-1K.

ORCID iDs

Liwen Zou https://orcid.org/0000-0003-4085-4003

References

Benmansour F and Cohen L D 2011 Tubular structure segmentation based on minimal path method and anisotropic enhancement *Int. J. Comput. Vision* **92** 192–210

Cai X, Chan R, Schonlieb C-B, Steidl G and Zeng T 2019 Linkage between piecewise constant mumford-shah model and rudin-osher-fatemi model and its virtue in image segmentation *SIAM J. Sci. Comput.* **41** B1310–40

Caselles V, Kimmel R and Sapiro G 1997 Geodesic active contours *Int. J. Comput. Vision* **22** 61–79

Chan T F and Vese L A 2001 Active contours without edges *IEEE Trans. Image Process.* **10** 266–77

Chen L-C, Papandreou G, Kokkinos I, Murphy K and Yuille A L 2017 Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs *IEEE Trans. Pattern Anal. Mach. Intell.* **40** 834–48

Chen X, Chen Z, Li J, Zhang Y-D, Lin X and Qian X 2021 Model-driven deep learning method for pancreatic cancer segmentation based on spiral-transformation *IEEE Trans. Med. Imaging* **41** 75–87

Chen Y, Xu C, Ding W, Sun S, Yue X and Fujita H 2022 Target-aware u-net with fuzzy skip connections for refined pancreas segmentation *Appl. Soft Comput.* **131** 109818

Farag A, Lu L, Roth H R, Liu J, Turkbey E and Summers R M 2016 A bottom–up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling *IEEE Trans. Image Process.* **26** 386–99
Fortun D, Storath M, Rickert D, Weinmann A and Unser M 2018 Fast piecewise-affine motion estimation without segmentation *IEEE Trans. Image Process.* **27** 5612–24
Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann F A and Brendel W 2018 Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness arXiv:1811.12231
Hatamizadeh A 2020 *Deep Learning of Unified Region, Edge, and Contour Models for Automated Image Segmentation* (Los Angeles: University of California)
Hatamizadeh A, Hoogi A, Sengupta D, Lu W, Wilcox B, Rubin D and Terzopoulos D 2019 Deep active lesion segmentation *Int. Workshop on Machine Learning in Medical Imaging* (Springer) pp 98–105
Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, Roth H R and Xu D 2022 Unetr: transformers for 3d medical image segmentation *Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision* pp 574–84
Hu P, Li X, Tian Y, Tang T, Zhou T, Bai X, Zhu S, Liang T and Li J 2020 Automatic pancreas segmentation in ct images with distance-based saliency-aware denseaspp network *IEEE J. Biomed. Health Inform.* **25** 1601–11
Huang M-L and Wu Y-Z 2022 Semantic segmentation of pancreatic medical images by using convolutional neural network *Biomed. Signal Process. Control* **73** 103458
Isensee F, Jaeger P F, Kohl S A, Petersen J and Maier-Hein K H 2021 nnu-net: a self-configuring method for deep learning-based biomedical image segmentation *Nat. Methods* **18** 203–11
Kass M, Witkin A and Terzopoulos D 1988 Snakes: active contour models *Int. J. Comput. Vision* **1** 321–31
Khadidos A, Sanchez V and Li C-T 2017 Weighted level set evolution based on local edge features for medical image segmentation *IEEE Trans. Image Process.* **26** 1979–91
Kimmel R and Bruckstein A M 2003 Regularized laplacian zero crossings as optimal edge integrators *Int. J. Comput. Vision* **53** 225–43
Kobler E, Klatzer T, Hammernik K and Pock T 2017 Variational networks: connecting variational methods and deep learning *German Conf. Pattern Recognition* (Springer) pp 281–93
Law M W and Chung A 2010 An oriented flux symmetry based active contour model for three dimensional vessel segmentation *European Conference on Computer Vision* (Springer) pp 720–34
Li C, Kao C-Y, Gore J C and Ding Z 2008 Minimization of region-scalable fitting energy for image segmentation *IEEE Trans. Image Process.* **17** 1940–9
Li C, Xu C, Gui C and Fox M D 2005 Level set evolution without re-initialization: a new variational formulation *2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR'05)* vol 1, pp 430–6
Li C, Xu C, Gui C and Fox M D 2010 Distance regularized level set evolution and its application to image segmentation *IEEE Trans. Image Process.* **19** 3243–54
Lim S-H, Kim Y J, Park Y-H, Kim D, Kim K G and Lee D-H 2022 Automated pancreas segmentation and volumetry using deep neural network on computed tomography *Sci. Rep.* **12** 4075
Ma J *et al* 2021 Abdomenct-1k: is abdominal organ segmentation a solved problem? *IEEE Trans. Pattern Anal. Mach. Intell.* **44** 6695–714
Ma J, He J and Yang X 2020a Learning geodesic active contours for embedding object global information in segmentation cnns *IEEE Trans. Med. Imaging* **40** 93–104
Ma J, Nie Z, Wang C, Dong G, Zhu Q, He J, Gui L and Yang X 2020b Active contour regularized semi-supervised learning for covid-19 ct infection segmentation with limited annotations *Phys. Med. Biol.* **65** 225034
Man Y, Huang Y, Feng J, Li X and Wu F 2019 Deep q learning driven ct pancreas segmentation with geometry-aware u-net *IEEE Trans. Med. Imaging* **38** 1971–80
Marcos D, Tuia D, Kellenberger B, Zhang L, Bai M, Liao R and Urtasun R 2018 Learning deep structured active contours end-to-end *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 8877–85
Min H, Jia W, Zhao Y, Zuo W, Ling H and Luo Y 2018 Late: a level-set method based on local approximation of taylor expansion for segmenting intensity inhomogeneous images *IEEE Trans. Image Process.* **27** 5016–31
Peyré G *et al* 2010 Geodesic methods in computer vision and graphics *Foundations and Trends® in Computer Graphics and Vision* (Boston: Now Publishers) vol 5, pp 197–397
Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 234–41
Roth H R, Lu L, Lay N, Harrison A P, Farag A, Sohn A and Summers R M 2018 Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation *Med. Image Anal.* **45** 94–107
Xue J, He K, Nie D, Adeli E, Shi Z, Lee S-W, Zheng Y, Liu X, Li D and Shen D 2019 Cascaded multitask 3D fully convolutional networks for pancreas segmentation *IEEE Trans. Cybern.* **51** 2153–65
Yu Q, Xie L, Wang Y, Zhou Y, Fishman E K and Yuille A L 2018a Recurrent saliency transformation network: incorporating multi-stage visual cues for small organ segmentation *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 8280–9
Yu Z, Liu W, Zou Y, Feng C, Ramalingam S, Kumar B and Kautz J 2018b Simultaneous edge alignment and learning *Proc. of the European Conf. on Computer Vision (ECCV)* pp 388–404
Zhang D, Zhang J, Zhang Q, Han J, Zhang S and Han J 2021a Automatic pancreas segmentation based on lightweight dcnn modules and spatial prior propagation *Pattern Recognit.* **114** 107762
Zhang W, Wang X, You W, Chen J, Dai P and Zhang P 2019 Resls: region and edge synergetic level set framework for image segmentation *IEEE Trans. Image Process.* **29** 57–71
Zhang Y, Wu J, Liu Y, Chen Y, Chen W, Wu E X, Li C and Tang X 2021b A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set *Med. Image Anal.* **68** 101884
Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, Li L, Wei W and He J 2022 Cancer incidence and mortality in China, 2016 *J. Natl Cancer Center* **2** 1–9
Zhou Y, Xie L, Shen W, Wang Y, Fishman E K and Yuille A L 2017 A fixed-point model for pancreas segmentation in abdominal ct scans *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 693–701
Zhu Z, Xia Y, Shen W, Fishman E and Yuille A 2018 A 3D coarse-to-fine framework for volumetric medical image segmentation *2018 Int. Conf. on 3D Vision (3DV)* (IEEE) pp 682–90