

**WestminsterResearch**

<http://www.westminster.ac.uk/westminsterresearch>

**Morphological and multi-level geometrical descriptor analysis in  
CT and MRI volumes for automatic pancreas segmentation  
Asaturyan, H., Gligorievski, A. and Villarini, B.**

NOTICE: this is the authors' version of a work that was accepted for publication in Computerized Medical Imaging and Graphics. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Computerized Medical Imaging and Graphics, 75 (2019), pp. 1-13, 2019.

The final definitive version in Computerized Medical Imaging and Graphics is available online at:

<https://dx.doi.org/10.1016/j.compmedimag.2019.04.004>

© 2019. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>

---

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

---

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail [repository@westminster.ac.uk](mailto:repository@westminster.ac.uk)



# Morphological and multi-level geometrical descriptor analysis in CT and MRI volumes for automatic pancreas segmentation

Hykoush Asaturyan<sup>a</sup>, Antonio Gligorievski<sup>b</sup>, Barbara Villarini<sup>a</sup>

<sup>a</sup>*University of Westminster, Computer Science Department, London, United Kingdom*

<sup>b</sup>*Ss. Cyril and Methodius University, Department of Radiology, Skopje, Macedonia*

---

## Abstract

Automatic pancreas segmentation in 3D radiological scans is a critical, yet challenging task. As a prerequisite for computer-aided diagnosis (CADx) systems, accurate pancreas segmentation could generate both quantitative and qualitative information towards establishing the severity of a condition, and thus provide additional guidance for therapy planning. Since the pancreas is an organ of high inter-patient anatomical variability, previous segmentation approaches report lower quantitative accuracy scores in comparison to abdominal organs such as the liver or kidneys. This paper presents a novel approach for automatic pancreas segmentation in magnetic resonance imaging (MRI) and computer tomography (CT) scans. This method exploits 3D segmentation that, when coupled with geometrical and morphological characteristics of abdominal tissue, classifies distinct contours in tight pixel-range proximity as “pancreas” or “non-pancreas”. There are three main stages to this approach: 1) identify a major pancreas region and apply contrast enhancement to differentiate between pancreatic and surrounding tissue; 2) perform 3D segmentation via continuous max-flow and min-cuts approach, structured forest edge detection, and a training dataset of annotated pancreata; 3) eliminate non-pancreatic contours from resultant segmentation via morphological operations on area, structure and connectivity between distinct contours. The proposed method is evaluated on a dataset containing 82 CT image volumes, achieving mean Dice Similarity coefficient (DSC) of

---

*Email addresses:* [h.asaturyan@my.westminster.ac.uk](mailto:h.asaturyan@my.westminster.ac.uk) (Hykoush Asaturyan),  
[b.villarini@westminster.ac.uk](mailto:b.villarini@westminster.ac.uk) (Barbara Villarini)

$79.3 \pm 4.4\%$ . Two MRI datasets containing 216 and 132 image volumes are evaluated, achieving mean DSC  $79.6 \pm 5.7\%$  and  $81.6 \pm 5.1\%$  respectively. This approach is statistically stable, reflected by lower metrics in standard deviation in comparison to state-of-the-art approaches.

*Keywords:* Automatic pancreas segmentation, Computer-aided diagnosis, Continuous max-flow and min-cuts, Contrast enhancement, Geometrical characteristics, Structured forest

---

## 1. Introduction

The computation of pancreas measurements, such as volume and curvature, can provide insight about the progression of type 2 diabetes mellitus (Burute et al., 2014, Asaturyan and Villarini, 2018) and assist detection of pancreatic neoplasms (Roth et al., 2015a). Studies have reported that variations in pancreas contouring can be linked to ductal adenocarcinoma (Cruickshank and Benbow, 1995) and enhanced contour analysis can facilitate stratification of normal variations against pancreatic tumours (Omeri et al., 2017). Obtaining such information firstly requires segmentation of the pancreas in radiological image volumes, such as computer tomography (CT) and magnetic resonance imaging (MRI). Although manual segmentation can produce very accurate results, it is time-consuming, sometimes prone to interobserver variability, but above all is difficult to replicate given the growing number of available medical image volumes. On the other hand, computer-aided diagnosis systems (CADx) that generate accurate, automatic pancreas segmentation on a scale involving thousands of abdominal medical scans, could support clinical studies to establish essential correlations between organ volume, curvature and anthropometric measures (Villarini et al., 2017).

That said, the pancreas has high structural variability and a full inspection from a scan is problematic since the organ lies just behind the stomach: the head of the pancreas touches the small intestine and often overlaps with surrounding abdominal fat, artery and veins. In general, the greyscale intensity of the pancreatic region is very similar to nearby tissue, and consequently this increases the challenge of accurately segmenting this organ.

Figure 1(a) displays a single axial slice from an abdominal CT scan. The top image shows the original slice and the bottom image highlights the ground-truth of the pancreas contoured in red. Notice the similarity in the greyscale intensity of surrounding tissue. Differing from CT imaging, the low

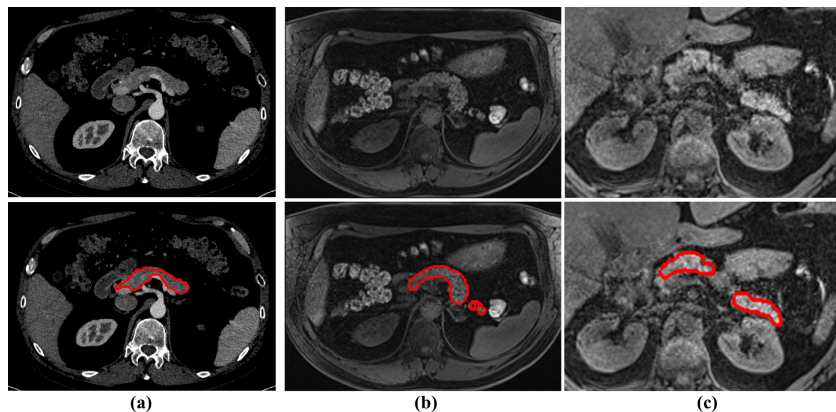


Figure 1: The expert-led manual (ground-truth) segmentation of the pancreas is contoured in red. Column (a) displays an axial slice in a CT scan. Columns (b) and (c) display two axial slices from two different MRI scans. Notice the finer image quality of CT in comparison to MRI that is significantly coarse and suffers from greater blurred boundaries between the pancreas, and duodenum.

resolution and slower imaging speed of MRI presents additional edge-based artefacts (Cai et al., 2016). Figure 1(b)(c) display axial MRI slices from two different scans for two different subjects. It is noted that the same imaging protocol was used to obtain both scans. Notice that in column (c), the slices appear “zoomed-in” to emphasise visibility variation of pancreatic features in a slice-by-slice inspection.

### 1.1. Related work

In recent research literature, various organ segmentation methods have been proposed in the scheme of multi-atlas approaches (Chu et al., 2013, Wolz et al., 2013, Wang et al., 2014, Okada et al., 2015, Tong et al., 2015) and convolutional neural networks (Roth et al., 2015a, Cai et al., 2016, Farag et al., 2017). The common factor amongst these methods is a reliance upon expert-led manually annotated organs in medical image volumes. Also, most segmentation methods have been performed on CT modality as opposed to MRI, which presents additional difficulties of image artefacts and greater blurred boundaries between organs.

The approach reported in Chu et al. (2013) describes an automatic multi-organ segmentation methodology that is based on spatially-divided probabilistic atlases and performs Markov random field (MRF)-based registration.

Wolz et al. (2013) describes a locally weighted atlas selection and patch-based scheme to perform multiple abdominal organ segmentation. A post-processing stage utilises graph-cuts using a learned intensity model. Wang et al. (2014) presents a patch-based label propagation that uses relative geodesic distances to define patient-specific coordinate systems as spatial context. This approach is said to overcome the problem of incorrect patch selection, which arises from registration errors after image alignment. In the approach reported in Okada et al. (2015) statistical shape modelling is performed using conditional shape-location and unsupervised intensity priors. The method described in Tong et al. (2015) employs voxel-wise local atlas selection, dictionary learning and sparse coding techniques to produce target specific priors for segmentation.

All of the above methods utilise datasets containing CT image volumes, and although produce promising segmentation results, these methods suffer from poor statistical stability with high standard deviation (>14%) and require 2 to 4 hours in computation time to process a single image volume (Chu et al., 2013, Wolz et al., 2013).

Recent advances in convolutional neural networks (CNNs) have been successfully woven into medical image segmentation algorithms, especially for abdominal organs that are highly deformable and possess vague edge boundaries. Roth et al. (2015a) presents an approach that uses dense labelling of local superpixel image patches via probability-based CNN and nearest neighbour fusion. Next, a regional CNN samples a set of bounding boxes around each image superpixel at different scales of contexts. The CNN models are trained to assign class probabilities for each superpixel region as either “pancreas” or “non-pancreas”. The methodology proceeds by employing another regional CNN that influences the joint space of CT image pixel intensities and the initial CNN outcome probability maps. In order to produce the final segmentation, post-processing integrates an application of 3D Gaussian smoothing and 2D conditional random fields (CRFs).

The use of CNN models is extended in Cai et al. (2016) where pancreatic detection and boundary segmentation of MRI volumes is performed using two types of CNN methods. Firstly, a region detection stage separates local image regions as “pancreas” or “non-pancreas” with spatial intensity context; secondly, the boundary detection stage extracts the semantic boundaries of the pancreas. Resulting outcomes from these two networks are merged as the input to a CRF, which provides the final segmentation result.

Farag et al. (2017) describes a bottom-up approach method by firstly

classifying image patches at different resolutions and cascading superpixels. Next, dense image patch labelling is performed using two methods: forest classification on image histogram and texture features; and deep convolutional neural network classification on image patches of greater spatial contexts.

With the exception of Okada et al. (2015), the above CNN methods yield DSC results that outperform the reported multi-atlas based approaches and produce a tighter standard deviation. However, such convolutional neural networks are prone to suffer from imbalance between classes (i.e. pancreas versus non-pancreas) and overfitting during the network training phase (Milletari et al., 2016), and thus ignore features that are related to the organ of interest during the prediction phase of test volumes.

### *1.2. Contributions*

The following list presents the technical and methodological novelties, which, in addition to further contributions, are presented in this sub-section: (a) a learned intensity model for digital contrast enhancement in image volumes; (b) a hybrid energy-minimisation segmentation approach exploits edge detection to yield detailed, optimal contouring of the pancreas; (c) a post-processing stage integrates principal geometric descriptors that characterises tissue and employs radiological expert-knowledge for refined tissue classification.

Although multi-atlas based approaches (Tong et al., 2015, Okada et al., 2015, Wolz et al., 2013) have achieved high quantitative accuracy scores (>90%) on organs such as the liver and kidneys, segmentation of pancreas is far lower in quantitative accuracy. Not only does the pancreas account for less than 1% in a given abdomen scan, but the selection of atlases with high similarity in the pancreas region is not successful because surrounding tissue, such as the duodenum, have similar greyscale intensity. In this paper, a pre-processing stage improves the distributional characterisation of intensities between pancreatic and surrounding tissue boundaries. A digital contrast intensity model is proposed to increase greyscale differences in an image volume, particularly in pancreatic contouring. Optimum parameters for tissue enhancement are achieved through analysing intensities in a corresponding training imaging dataset. The motivation is to differentiate the nearby splenic vein, superior mesenteric artery and surrounding tissue that are often overlapping, or in close proximity to the pancreas. Consequently,

this stage reduces the number of “non-pancreas” region predictions in later processing.

The methodologies reported in Roth et al. (2015a), Cai et al. (2016), Farag et al. (2017) perform convolutional neural network training and segmentation on 2D images or slice-by-slice basis. As a result, such methods can exhibit discontinuity or high variation in predicting pancreatic regions between consecutive slices. In this paper, the main segmentation approach employs a 3D based algorithm that performs an initial (rough) segmentation of the pancreas. Comparatively, this produces greater consistency in spatial smoothness and prediction among successive slices. In order to further overcome problems caused by the low contrast between organ boundaries, a pretrained edge detection model is utilised to “strengthen” the boundaries of distinct pancreas contours and reduce edges that appear within these closed contours, consequently improving tissue classification at the post-processing stage. Thus, the major pancreas region and surrounding tissue are extracted using a hybrid segmentation approach that integrates random forest probability-wise predictions of superpixels; structured forest edge detection for local regions; and energy-minimising 3D continuous max-flow and min-cuts, together which produces very detailed boundary preservation.

The resulting pancreas contouring has been verified by two independent experts in radiology and radiography. Moreover, this approach has shown to be equally effective across image volumes with varying noise and distortion.

Unlike previous approaches (Roth et al., 2015a, Wang et al., 2014, Tong et al., 2015, Farag et al., 2017) that label fixed image patches as “pancreas” or “non-pancreas” towards or during the final methodology stage, the approach presented in this paper demonstrates the effectiveness of classifying distinct contours with diverse size and structure. A multi-level stage of post-processing stage is proposed for refined tissue classification: morphological operations are performed to obtain geometrical descriptors in an image volume, and positioning of distinct contours are analysed to determine the likelihood of being pancreatic tissue. A combinations-based method of elimination aims to remove non-pancreatic contours that are likely to represent a fraction of the inferior vena cava, common hepatic artery, splenic artery and vein, superior mesenteric, stomach, left kidney or spleen. This detail of analysis, which combines expert-knowledge of radiological and anatomical structural information, preserves the original contouring of the pancreas while eliminating surrounding tissue separated by a tight pixel range.

The proposed method is evaluated on two imaging modalities, including



a CT dataset and two MRI datasets. The approach outperforms in Dice similarity coefficient (DSC) in comparison to previous state-of-the-art approaches Roth et al. (2015a), Cai et al. (2016), Wolz et al. (2013), Okada et al. (2015), Tong et al. (2015), Farag et al. (2017), Karasawa et al. (2017) and is relatively comparable to methods Roth et al. (2017), Cai et al. (2017), Zhou et al. (2016) with a lower standard deviation. The volumetric Jaccard Index (JI), precision and recall results are also reported.

Every stage of the methodology is reproducible and scalable depending on the number of training image volumes and the depth of refinement. Although the proposed approach has been tested on pancreas segmentation alone, the methodology can be generalisable to other 3D organ or muscular tissue segmentation tasks.

In Section 2, the methodology for 3D pancreas segmentation and refinement is covered. Section 3 presents the experiments and evaluation achieved, including key program parameters and numerical implementation. Section 4 discusses the segmentation results’ outcome with a comparison to those reported in recent literature and strategies for further optimisation. Section 5 provides a conclusion.

## 2. Methods

The methodology of the proposed approach, as illustrated in Figure 2, progresses through three main stages, each one of which is discussed below in further depth.

### 2.1. Major pancreas region detection and intensities analysis

An effective application of contrast enhancement in CT or MRI can differentiate pancreatic tissue and boundaries against background classes of blood vessels, stomach fundus and the first section of the small intestine (duodenum). Following an application of Non Local-means (Buades et al., 2005) for noise reduction, a sigmoid function is applied to a given test image volume by incorporating a gain,  $g$ , which controls the actual contrast, and a cut-off value,  $c$ , which represents the (normalised) greyscale value about which contrast level is changed. Every  $i$ -th slice,  $s_i$ , in the image volume undergoes contrast enhancement,  $C(s_i)$ , as described in Equation 1:

$$C(s_i) = \frac{1}{1 + \exp [g(c-s_i)]} \quad (1)$$

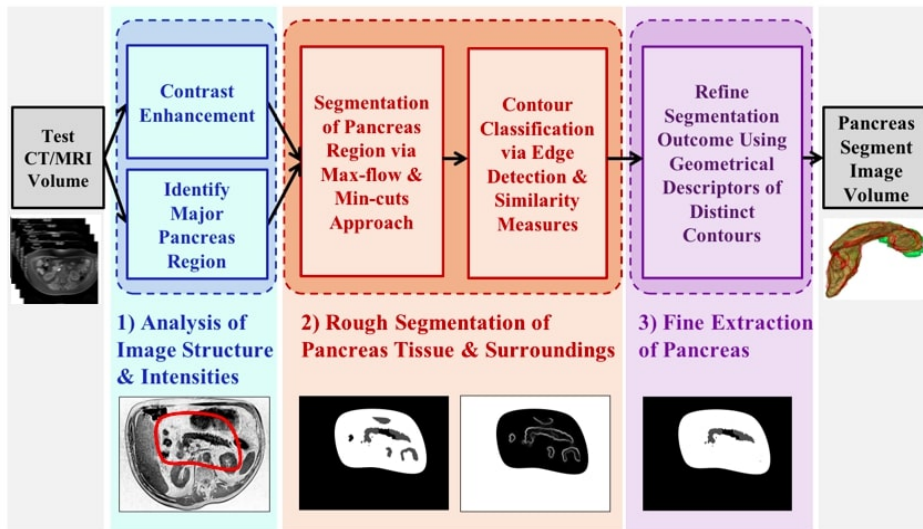


Figure 2: Overview of methodology: three main stages are proposed for automatic pancreas segmentation in abdominal CT and MRI scans (volumes).

Figure 3(a)(b) depicts a slice from an image volume before and after contrast enhancement, respectively. Empirically testing, the values for gain and cut-off were assigned to image volumes in the training dataset by considering the respective mean greyscale intensity value. From here, a non-linear regression model is developed to predict the gain for the test volume. Similarly, a linear regression model is developed for predicting the cut-off value. Overall, the value of the gain and cut-off value increases in proportion to the mean pixel intensity of the image volume.

Next, the major pancreas region is identified using the method reported in Farag et al. (2017). A random forest is trained on a selection of extracted features in the training data; afterwards, the trained forest predicts the likelihood of a region in the test image volume as “pancreas” given a probability threshold.

Essentially, every image volume in the training dataset is initially converted to superpixels of 32 pixel region size. From here, feature information is extracted in image patches of  $25 \times 25$  pixels, describing texture and the probability of a patch as “pancreas” based on voxel intensity analysis. Equation 2 describes this probability as:

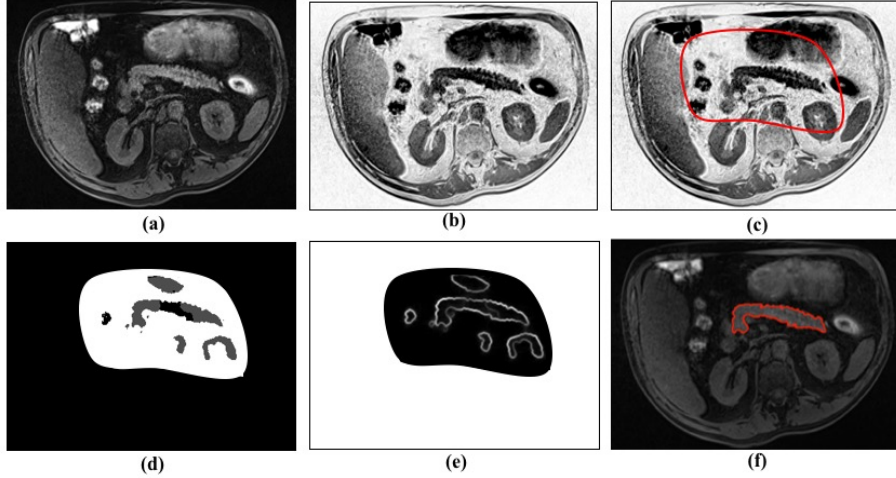


Figure 3: Visualisation of results for a slice from image volume. (a) Original MRI slice, (b) MRI slice after contrast enhancement, (c) Red bounding outline depicts major pancreas region, (d) Segmentation following max-flow and min-cuts approach, (e) Boundary detection using structured forest learning, (f) Final contour segmentation after refinement.

$$p^+ = \frac{(f^+(M^+))}{(f^+(M^+) + f^-(M^-))} \quad (2)$$

where  $f^+(M^+)$  and  $f^-(M^-)$  are kernel density estimators representing the intensity distributions of the positive  $\{M^+\}$  and negative  $\{M^-\}$  pancreas and non-pancreas voxels from an image volume, respectively.

Two further features for each axial slice are the normalised relative  $x$ -axis and  $y$ -axis positions in the range  $[0, 1]$ , which are calculated at each image patch centre against the segmented abdomen area. From here, a total of 46 image patch-level features per superpixel are employed to train the random forest classifier. The image patch labels are acquired by extracting information from their patch centre pixels in accordance to their ground-truth segmentation.

The trained random forest classifier distinguishes the negative or “non-pancreas” class patches such as the liver and greater outer region. Figure 3(c) displays a red outline over a sample slice that embodies the area predicting “pancreas” at a probability of 0.85 per superpixel - this represents the major pancreas region for that particular slice in a test image volume.

## 2.2. Segmentation of pancreas tissue and surroundings

The major region of interest as a volume is processed through an unsupervised 3D segmentation algorithm (Yuan et al., 2010), which uses maximal-flow and minimum graph-cuts approach. Allow  $\Omega$  to serve as a closed and continuous 3D domain representing the major pancreas region volume as a graph. At every position,  $x \in \Omega$ , the spatial flow passing  $x$  can be written as  $q(x)$ . Additionally, the directed source flow from terminal  $s$  to  $x$  can be denoted by  $q_s(x)$ , and the directed sink flow from  $x$  to terminal  $t$  by  $q_t(x)$ . The continuous max-flow and min-cuts model can be described through the introduction of a multiplier known as the “dual variable”, denoted by  $\mu$  to some flow conservation:

$$\max_{q_s, q_t, q} \min_{\mu} \int_{\Omega} q_s dx + \int_{\Omega} \mu (\nabla \cdot q - q_s + q_t) dx \quad (3)$$

such that  $q_s(x) \leq C_s(x)$ ,  $q_t(x) \leq C_t(x)$  and  $|q(x)| \leq C(x)$ , where  $C(x)$ ,  $C_s(x)$  and  $C_t(x)$  describe given capacity functions, and  $\nabla \cdot q$  calculates the total spatial flow nearby  $x$ . From here, the “Multiplier-Based Maximal-Flow Algorithm” described in Yuan et al. (2010) is employed to perform unsupervised image segmentation on the entire volumetric major pancreas region. Figure 3(d) displays the resulting segmentation of a single slice in a given test image volume.

*Edge detection and boundary matching.* The segmented image volume undergoes a transformation via structured forest learning (Dollár and Zitnick, 2013) where the boundaries or edges of pancreatic issue and surroundings are detected. The edges of segments in each slice are measured against the boundaries of equally sized pancreas segments provided in the training dataset. The measure of similarity between these edges are performed via modified Hausdorff distance (MHD) (Dubuisson and Jain, 1994) and structural similarity (SSIM) index (Dosselmann and Yang, 2011).

An MHD measure provides a relatively local comparison between two pancreatic boundary points in comparison to SSIM, which tends to capture a much more global similarity between two image patches. For example, an MHD captures small circular patterns or strokes of pancreatic contours, whereas SSIM provides an overall morphology ‘appearance’ similarity between contours being compared. Hence, combining these two metrics provides the advantages of both global and local measures of similarity.

Whenever the error between a region in the training data and its corresponding region in the segmentation slice falls below empirically identified threshold values of 0.15 (MHD) and 0.29 (SSIM), a boundary match is assigned to a compilation of pancreas contour similarities, otherwise the contour is discarded. Figure 3(e) depicts the boundaries of different tissue in a segmented slice after max-flow and min-cuts segmentation. Notice the variation in contour intensity against the background: the greater outline of the pancreas has been heavily detected whereas the edges inside are less enhanced, thus preventing unnecessary segment division in later processing.

### *2.3. Refined extraction of pancreas*

Once an image volume of rough segmentation has been extracted, a stage of post-processing eliminates surrounding contours identified as “non-pancreas”. Figure 3(f) displays the final segmentation outcome for a single slice in a test image volume following three levels of refinement. Each level progressively targets surrounding tissue that is located in closer proximity to the pancreas. The first level performs a shallow-based removal of distinct non-pancreatic contours such as whole or remaining fractions of the aorta, portal vein, duodenum, stomach, and the ascending and descending colon. The second level aims to further tackle non-pancreatic tissue that do not lie within the radiologically described embodiment of a pancreas shape (Cruickshank and Benbow, 1995). Last, but certainly not least, the third level of refinement analyses the positioning of distinct contours that are very close to but are not part of the pancreas’ head and body, such as the inferior vena cava, common hepatic artery, splenic artery and vein, superior mesenteric, a fraction of the stomach; and close to but not part of the pancreas’ tail including a fraction of the left kidney and spleen. Depending on the slice position in the segmented image volume, such tissue can be separated by just a few pixels from pancreatic tissue and thus requires further analysis to increase the likelihood of correct tissue classification. In order to emphasise the importance of this post-processing stage, Figure 4 highlights the major pancreas region in three slices (from three different image volumes) that progress through three consecutive levels of refinement. Notice that close-range non-pancreatic contours are effectively removed in the third (and final) level.

### 2.3.1. Level 1: morphological operations on distinct contours

Analysing a given set of annotated training image volumes, the following is deduced: the mean of ranges (where the pancreas is visible throughout successive slices in a volume), with careful consideration to heavy outliers; the mean slice number where the pancreas features become visible for the first time,  $s_{start}$ , reach maximum area,  $s_{max}$ , and the last slice after which pancreas features are not visible any longer,  $s_{end}$ . By considering the total number of slices in each image volume,  $t_s$ , a discrete set of four slice ranges are established:  $r_1 : [1, s_{start}-1]$ ,  $r_2 : [s_{start}, s_{max}-1]$ ,  $r_3 : [s_{max}, s_{end}-1]$  and  $r_4 : [s_{end}, t_s]$ , where  $r_1, r_2, r_3$  and  $r_4 \in \mathbb{Z}$ . A k-medoids cluster approach is employed in order to generate  $N_{r_1}, N_{r_2}, N_{r_3}$  and  $N_{r_4}$  number of groups of constraints for every slice range. Every group of constraints includes a measure of area, triangularity and ratio of spatial dimensions. For each slice range, individual contours in the segmentation image volume are simultaneously measured against corresponding groups of constraints; if the error for each observation is greater than an empirically assigned threshold,  $t_m$ , then this contour is regarded as “non-pancreas” and removed from the overall segmentation result, otherwise it is retained. Let  $N_{r_i}(\alpha)$ ,  $N_{r_i}(\beta)$  and  $N_{r_i}(\gamma)$  represent individual constraints of area, triangularity and ratio of spatial dimensions. Thus, for every  $N_{r_i}$ , this operation can be defined as:

$$p_c = \begin{cases} 1 & \text{if } 0 \leq E_{r_i}^\alpha, E_{r_i}^\beta, E_{r_i}^\gamma \leq t_m \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $p_c$  represents a segmentation contour whose value of 1 corresponds to “pancreas” and a value of 0 corresponds to “non-pancreas”.  $E_{r_i}^\alpha$ ,  $E_{r_i}^\beta$  and  $E_{r_i}^\gamma$  represent the error between  $N_{r_i}(\alpha)$ ,  $N_{r_i}(\beta)$  and  $N_{r_i}(\gamma)$  and an observed segmentation contour’s similar measures, respectively.

Another morphological operation is the mean curvature (J. Williams and Shah, 1992) of distinct contours in each segmentation slice. Suppose that  $C_p$  represents a contour with natural parameter,  $t$ . Let the angle between the tangents to  $C_p$  at points  $q_1$  and  $q_2$  be  $\theta(q_1, q_2)$ , and let the length of the segment of the curve between  $q_1$  and  $q_2$  be  $s(q_1, q_2)$ . The curvature,  $\kappa_p$ , of the curve  $C_p$  to  $q_1$  is defined in Equation 5 as:

$$\kappa_p = \lim_{q_2 \rightarrow q_1} \frac{\theta(q_1, q_2)}{s(q_1, q_2)} \quad (5)$$

If an observed contour falls below a threshold,  $t_c$ , which is based on

the curvature analysis of unique contours in the training dataset, then it is discarded from the resultant segmentation volume. The top image in Figure 5(a) illustrates an example where several distinct contours in a slice are labelled with area, ratio and triangularity; the bottom image displays the resultant slice after removal of contours deemed as “non-pancreas”.

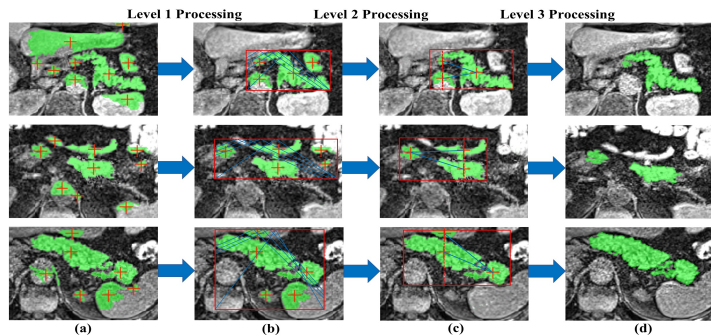


Figure 4: Visualisation of three levels for fine pancreas extraction in three different slices from three different image volumes. Column (a): original slice after initial segmentation. Column(b): resultant slice after first level of refinement, now contained in a bounding box against a “trail map”. Column (c): resultant slice after second level of refinement, now contained in a bounding box that identifies gradient between combinational distinct contours. Colum (d): resultant slice after third level of refinement.

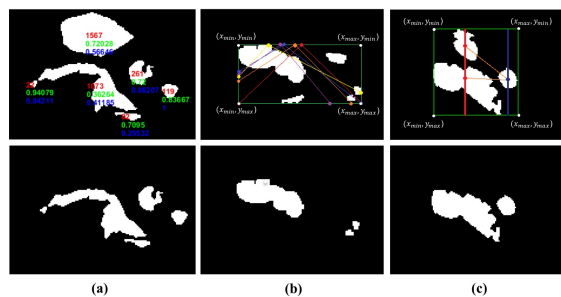


Figure 5: Visualisation of three levels for fine pancreas extraction in three different slices (from three different image volumes). Top row and bottom row: distinct contours in slice before and after processing, respectively. Column (a): first level of morphological operations on distinct contours, highlighting measurements of area, spatial aspect ratio and triangularity. Column (b): second level of positioning contours on “trail map” within bounding box containing all contours. Column (c): combinational connectivity between centre landmark points and respective gradients between all distinct contours.

### 2.3.2. Level 2: localisation and positioning of contours

The slice-by-slice inspection of pancreatic regions in the training dataset reveals that whole or distinct pancreatic contours are embodied in a shape resembling a horseshoe, an inverted-V, transverse, sigmoidal (Hagen-Ansert, 2017) but more commonly, oblique or L-shaped (Cruikshank and Benbow, 1995). Therefore, consider a localisation bounding box to contain all the contours in each segmentation slice, as described in 6:

$$F_s = [x_{min}, y_{min}, (x_{max} - x_{min}), (y_{max} - y_{min})] \quad (6)$$

where  $(x_{min}, y_{min})$  represents the top left-vertex of the bounding box,  $(x_{max} - x_{min})$  is the width and  $(y_{max} - y_{min})$  is the height that follows from this point of reference. From here, it is possible to generate an L-shaped template that behaves like a “trail-map” for identifying contours deemed as “pancreas” or otherwise. This trail-map can be viewed as a collection of neighbouring paths that begin from a set of points,  $B_{XY} = \{(x_1, y_{max}), \dots, (x_n, y_{max})\}$ , on the bottom horizontal of the bounding box and rise by corresponding angles,  $\theta = \{\theta_1, \dots, \theta_n\}$ , to respective points on the top horizontal of the bounding box. From here, the trail descends by angles,  $\phi = \{\phi_1, \dots, \phi_n\}$ , to respective terminating points (on the bounding box). It is noted that  $B_X = \{x_1, \dots, x_n\}$  are values that refer to a set of  $n$  distances measured from the bottom right-hand vertex, i.e.  $(x_{max}, y_{max})$ , hence,  $B_X \in \mathbb{R} \mid x_{min} \leq B_X \leq x_{max}$ . Values of  $\theta$  and  $\phi$  are co-dependent on the width and height of the bounding box.

The top image in Figure 5(b) illustrates an example where several oblique trails run across distinct segments. An observed contour in a segmentation slice that does not lie on the trail-map is removed from the image volume as displayed in the bottom image of Figure 5(b), otherwise it is retained to progress to the next post-processing level.

### 2.3.3. Level 3: centre landmarks of distinct contours

A final level of refinement examines the gradient between pairs of contours. For example, in any given slice in the training dataset, it is unusual, if not uncommon to observe two pancreatic contours where the gradient between their landmark centre points is infinity. Therefore, if the centre points of two pancreatic contours in a segmentation slice lie on the same imaginary vertical line, then it is safe to suggest that one of two is likely “non-pancreas” and opt for removal. The top image in Figure 5(c) illustrates an example where a pair of contours with gradient infinity have been identified; the bot-



tom image displays the resultant slice after removal of the contour deemed “non-pancreas”.

In order to identify which contour to eliminate from the resultant segmentation, a number of computations are performed to generate multiple groups of “segmentation characteristics”. These groups are analysed against multiple groups of similar measurements in the training data, otherwise known as “ground-truth characteristics”, and contours identified as “non-pancreas” are eliminated to form the final segmentation result.

*Generate groups of “ground-truth characteristics”.* Firstly, for every image volume in a training dataset, the slices that contain two, three or four distinct pancreas contours are analysed. For each slice, it is possible to generate its respective “ground-truth characteristics” that contains the following: the areas of individual contours, the gradient, Euclidean distance and ratio-of-areas between pairs of distinct contours, which are computed in a combinational manner (without repetition).

For example, as highlighted in Figure 6(a), if an image slice has three distinct contours with areas,  $A$ ,  $B$  and  $C$ , then three combination gradients,  $g_{AB}$ ,  $g_{AC}$  and  $g_{BC}$  are computed, as well as corresponding Euclidean distances,  $d_{AB}$ ,  $d_{AC}$  and  $d_{BC}$ . Furthermore, the ratios of the former area to the latter, i.e.,  $A/B$ ,  $A/C$  and  $B/C$  are important descriptors, providing an indication of overall contour shape as opposed to area alone.

Four groups,  $R_1$  to  $R_4$ , describing “ground-truth characteristics” of different combinations of contours, are generated. Each group is confined to its respective slice range,  $r_1$  to  $r_4$ . As highlighted in Figure 6(b),  $R_1$  describes a set containing groups of “ground-truth characteristics” from a combination of two contours.  $R_2$  describes two sets: one set has groups of “ground-truth characteristics” from a combination of two contours and another set has groups of “ground-truth characteristics” from a combination of three contours.  $R_3$  describes three sets, each set contains groups of “ground-truth characteristics” from a combination of two, three and four distinct contours separately.

Each set of groups of “ground-truth characteristics” can be represented by  $f_p^i$  where  $i$  refers to the  $i$ -th slice range group,  $R_i$ , and  $p$  refers to the number of contours. Figure 6(b) provides a visual summary of all sets and defined below is an example of  $f_3^2$ :

$$\begin{aligned}
f_3^2 = \{ & (A, B, C, g_{AB}, g_{AC}, g_{BC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_1, \\
& A, B, C, g_{AB}, g_{AC}, g_{BC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_2, \\
& \dots \\
& A, B, C, g_{AB}, g_{AC}, g_{BC}, d_{AB}, d_{AC}, d_{BC}, A/B, A/C, B/C)_n \}.
\end{aligned} \tag{7}$$

where  $n$  is a chosen number of groups in a set that is based on k-medoids clustering of these measurements in the training dataset.

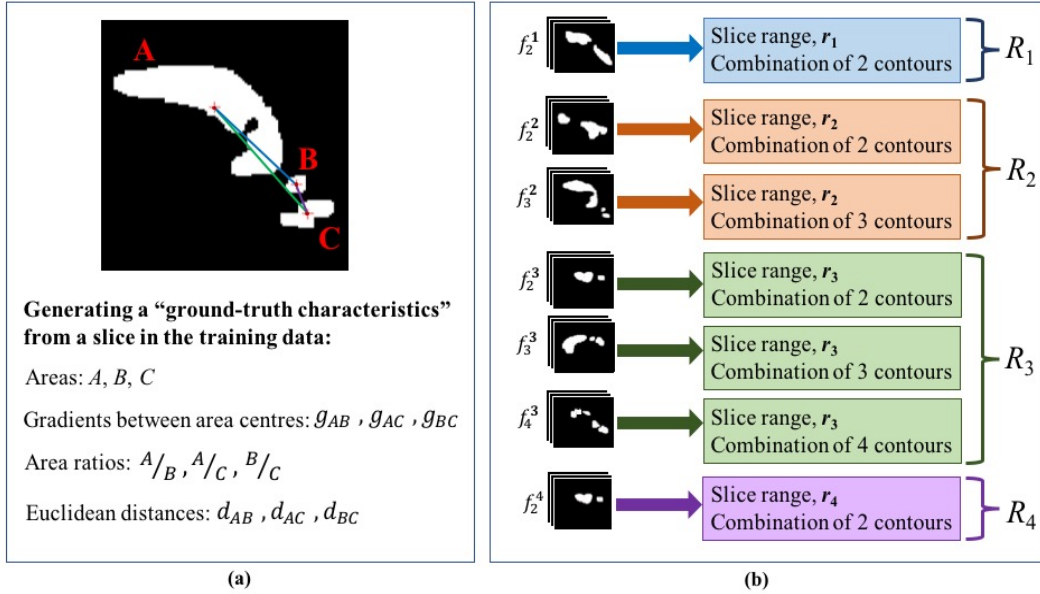


Figure 6: Column (a) provides a visual example of a slice from an image volume in a training dataset. Computations include the gradient and distance between combinational area centres and the respective ratios of combinational areas. Column (b) provides a summarised visual representation of "ground-truth characteristics" generated for sets  $R_1, R_2, R_3$  and  $R_4$ .

*Generate groups of "segmentation characteristics".* It should be noted that the focus lies on analysing two, and only two contours in each segmentation slice whose centres lie on the same imaginary vertical line. This limits computational time, but the general technique can be extended to a greater number of contours. Thus, in each segmentation slice, let  $M$  represent the total number of separate contours and let  $N$  represent the total number of

vertical lines that have two contours, i.e., where the gradient between respective centre points is infinity. Therefore, it is of interest to retain  $T = (M - N)$  contours in the slice. Next, the total number of possible combinations of  $T$  from  $M$  contours is computed, as denoted by  $C_{TM}$  and defined in Equation 8 as:

$$C_{TM} = \frac{M!}{T!(M - T)!} \quad (8)$$

For every  $C_{TM}(j)$ -th combination of contours, where  $j \in \mathbb{Z} : j \in [1, C_{TM}]$ , the gradients of all possible paired contours are calculated, in addition to respective distances and ratios.

For example, given a slice range group  $R_3$ , if  $M = 5$  and  $T = 3$ , then  $C_{TM} = 10$  different combinations of three contours from five contours (where two are presumed invalid). It is important to note that a number of combinations will be irrelevant since the gradient between two centres is infinity, and thus can be discarded. Hence, for every remaining (relevant) three contour combination, e.g. let  $C_{TM}(7)$  involve areas,  $B$ ,  $C$ ,  $D$ , a “segmentation characteristics” group is generated consisting of three gradients, areas, Euclidean distances and ratios-of-areas. This group is simultaneously evaluated against a set of “ground-truth characteristics” from the same slice range group and  $T$  contours of interest. Figure 7 provides a visual representation of this example, where set  $f_3^3$  would be utilised.

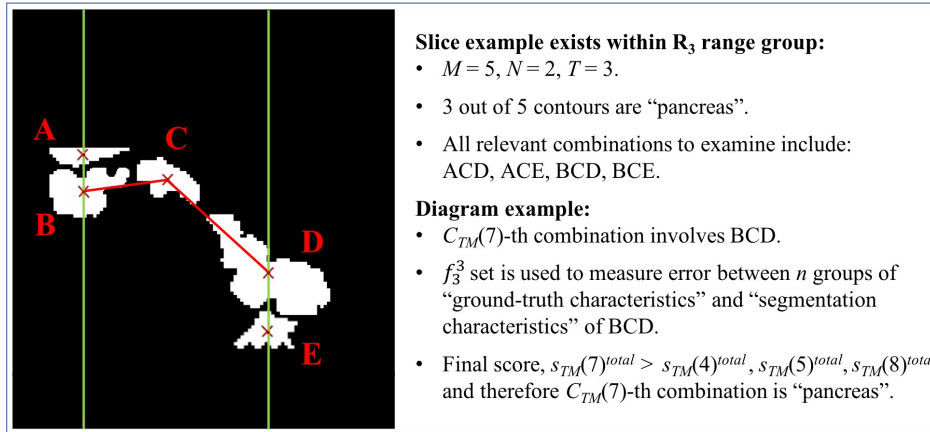


Figure 7: Visual example of three contour combination,  $BCD$ , which represents the 7th in 10 different combinations from five distinct contours. Three out of five contours are presumed valid, i.e. “pancreas” and the other two otherwise.

*Elimination process of “non-pancreas” tissue.* A method of elimination follows: if the mean error between the  $C_{TM}(j)$ -th “segmentation characteristics” group and a “ground-truth characteristics” group falls below threshold,  $t_p$ , then this combination is assigned a point score of 1, otherwise 0. Once all measures of error have been exhausted, the combination with the highest score is retained as “pancreas” and the others are discarded. Equation 9 mathematically describes the process of assigning a score to the  $j$ -th combination,  $s_{TM}(j)^k$ , where  $k \in \mathbb{Z} : k \in [1, n]$ .

$$s_{TM}(j)^k = \begin{cases} 1 & \text{if } 0 \leq E_{f_p^i}^k \leq t_p \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where  $E_{f_p^i}^k$  is the mean error between the  $C_{TM}(j)$ -th combination “segmentation characteristics” and  $(f_p^i)_k$  which is the  $k$ -th “ground truth characteristics”. The final score,  $s_{TM}(j)^{total}$  can be described in Equation 10 as:

$$s_{TM}(j)^{total} = \sum_{k=1}^n s_{TM}(j)^k \quad (10)$$

In the case where two or more combinations have attained the same score, the combination that encapsulates these two contours scoring the least mean error is retained.

### 3. Experiments and results

#### 3.1. Datasets

The performance of this approach is evaluated on three pancreas datasets, all of which have been manually annotated by an expert operator. The first two datasets contain 216 (MRI-A) and 132 (MRI-B) T2-weighted (fat suppressed) abdominal 3D MRI scans, obtained using a Philips Intera 1.5T scanner and a Siemens Trio 3T scanner respectively. The CT-NIH dataset is publicly available and contains 82 abdominal contrast-enhanced CT 3D scans, acquired on Philips and Siemens MDCT scanners (120 kVp tube voltage) <http://dx.doi.org/10.7937/K9/TCIA.2016.tNB1kqBU>.

Every image volume (scan) in MRI-A consists of 50 slices with spacing  $2mm$ , each slice has spatial size  $384 \times 384$  and  $0.9766mm$  pixel interval in the axial and sagittal direction. In MRI-B, every image volume consists

of 80 slices with  $1.6mm$  spacing, each slice has spatial size  $320 \times 260$  and  $1.1875mm$  pixel interval in the axial and sagittal direction. The CT-NIH dataset contains image volumes of spatial size  $512 \times 512$  pixels (between 181 and 466 slices) and slice thickness ranging  $1.5 - 2.5mm$ .

The volunteers who underwent the MRI scan were over 18 years of age and displayed early signs of type 2 diabetes. The CT scans were acquired from 53 male and 27 female subjects. A radiologist selected sixty-five subjects from patients who did not have any major abdominal pathologies nor pancreatic cancer lesions; the remaining seventeen subjects were healthy kidney donors scanned prior to nephrectomy. The subjects' ages range from 18 to 76 years with a mean age of  $46.8 \pm 16.7$ .

### 3.2. Evaluation

The training and testing image volumes have been randomly split into 196 and 20 for MRI-A, 112 and 20 for MRI-B, and 62 and 20 for CT-NIH.

For each experiment in Section 2.3.1, the values for  $N_{r_1}$ ,  $N_{r_2}$ ,  $N_{r_3}$  and  $N_{r_4}$  are 32, 52, 64 and 16 respectively. The thresholds include  $t_m = 0.15$  and  $t_c = 0.27$ . In Section 2.3.2, the value of  $n$  is 6, and in Section 2.3.3, the value of threshold  $t_p = 0.15$  and  $n$  is set to 16, 48, 36 and 12 for slice groups  $R_1$ ,  $R_2$ ,  $R_3$  and  $R_4$  respectively.

The performance of the proposed approach is evaluated using the Dice Similarity Coefficient (DSC) and Jaccard Index (JI) method (Taha and Hanbury, 2015), in addition to precision (PC) and recall (RC) (Yeghiazaryan and Voiculescu, 2015). If  $G$  represents the volumetric ground-truth annotation and  $S$  represents the corresponding automatic segmentation result, then the DSC accuracy percentage of  $S$  relative to  $G$  can be defined as:  $DSC = 2(|G \cap S|)/(|G| \cup |S|)$ . Similarly, the JI accuracy percentage of  $S$  relative to  $G$  can be defined as:  $JI = (|G \cap S|)/(|G \cup S|)$ . The precision normalises the correct segmented region against the segmentation:  $PC = (|S \cap G|)/|S|$ . The recall (i.e. sensitivity) normalises  $S \cap G$  against the ground-truth,  $G$  and can be defined as:  $RC = (|S \cap G|)/|G|$ .

Table 1 displays the DSC, JI, PC and RC for the CT-NIH dataset as mean  $\pm$  standard deviation [lowest, highest], in comparison to other automatic approaches reported in research literature that employ CT modality. Similarly, Table 2 displays the quantitative accuracy scores for datasets MRI-A and MRI-B, in comparison to automatic approaches reported in research literature that use MRI modality.

Table 1: Overall DSC, JI, PC and RC shown as mean  $\pm$  standard deviation [lowest, highest] for automatic pancreas segmentation methods in CT modality image volumes. The value of N represents the dataset size.

| Method                   | DSC(%)                             | JI(%)                           | PC(%)                          | RC(%)                          | N   |
|--------------------------|------------------------------------|---------------------------------|--------------------------------|--------------------------------|-----|
| Roth et al. (2015a)      | 71.8 $\pm$ 10.7<br>[25.0, 86.9]    | -                               | -                              | -                              | 82  |
| Karasawa et al. (2017)   | 78.5 $\pm$ 14.0                    | 66.3 $\pm$ 15.5                 | -                              | -                              | 150 |
| Tong et al. (2015)       | 71.1 $\pm$ 14.7                    | 56.90 $\pm$ 15.2                | -                              | -                              | 150 |
| Okada et al. (2015)      | 73.4 $\pm$ 15.1                    | 60.4 $\pm$ 16.7                 | -                              | -                              | 134 |
| Wolz et al. (2013)       | 69.6 $\pm$ 16.7                    | 55.50 $\pm$ 17.1                | 67.9 $\pm$ 18.2                | 74.1 $\pm$ 17.1                | 150 |
| Wolz et al. (2012)       | 65.5                               | 49.6                            | 70.7                           | 62.9                           | 100 |
| Roth et al. (2017)       | 81.27 $\pm$ 6.27                   | -                               | -                              | -                              | 82  |
| Cai et al. (2017)        | 82.4 $\pm$ 6.7<br>[60.0, 90.1]     | 70.6 $\pm$ 9.00<br>[42.9, 81.9] | -                              | -                              | 82  |
| Zhou et al. (2016)       | 82.37 $\pm$ 5.68<br>[62.43, 90.85] | -                               | -                              | -                              | 82  |
| Wang et al. (2014)       | 65.5 $\pm$ 18.6<br>[2.4, 90.2]     | -                               | -                              | -                              | 100 |
| Chu et al. (2013)        | 69.1 $\pm$ 15.3                    | 54.6 $\pm$ 15.9                 | -                              | -                              | 100 |
| Oda et al. (2016)        | 75.1 $\pm$ 15.4                    | 62.1 $\pm$ 16.6                 | -                              | -                              | 147 |
| Saito et al. (2016)      | 74.4 $\pm$ 20.2                    | 62.3 $\pm$ 19.5                 | -                              | -                              | 140 |
| Shimizu et al. (2010)    | -                                  | 57.9                            | -                              | -                              | 98  |
| Roth et al. (2015b)      | 68.0 $\pm$ 10.0<br>[43.0, 80.0]    | -                               | -                              | -                              | 82  |
| Roth et al. (2016a)      | 78.0 $\pm$ 8.2                     | -                               | -                              | -                              | 82  |
| Farag et al. (2014)      | 68.8 $\pm$ 25.6                    | 57.2 $\pm$ 25.4                 | 71.5 $\pm$ 30.0                | 72.5 $\pm$ 27.2                | 80  |
| Farag et al. (2017)      | 70.7 $\pm$ 13.0                    | 57.9 $\pm$ 13.6                 | 71.6 $\pm$ 10.5                | 74.4 $\pm$ 15.1                | 80  |
| Hammon et al. (2013)     | -                                  | 61.2 $\pm$ 9.08                 | -                              | -                              | 40  |
| Okada et al. (2012)      | -                                  | 46.6                            | -                              | -                              | 28  |
| <b>CT-NIH (Proposed)</b> | 79.3 $\pm$ 4.4<br>[72.8, 86.0]     | 66.1 $\pm$ 6.2<br>[58.2, 75.5]  | 88.6 $\pm$ 6.9<br>[74.5, 99.5] | 97.6 $\pm$ 1.9<br>[94.7, 99.8] | 82  |

Figure 8 displays the final segmentation results in six MRI scans, equally split between MRI-A and MRI-B. Despite columns (a, b, c) being part of MRI-A, notice the variation between intensity and image contrast in the original axial MRI slices. Columns (d, e, f) corresponds to exemplars from dataset MRI-B. In particular, the contouring of the pancreas can appear less-well defined with blurred boundaries between the organ and surround-

Table 2: Overall DSC, JI, PC and RC shown as mean  $\pm$  standard deviation [lowest, highest] for automatic pancreas segmentation methods in MRI modality image volumes. The value of N represents the dataset size.

| Method                  | DSC(%)                          | JI(%)                           | PC(%)                           | RC(%)                           | N   |
|-------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|-----|
| Cai et al. (2016)       | 76.1 $\pm$ 8.7<br>[47.4, 87.1]  | -                               | -                               | -                               | 78  |
| Cai et al. (2017)       | 80.5 $\pm$ 6.70<br>[59.1, 89.4] | 67.9 $\pm$ 8.90<br>[41.9, 80.9] | -                               | -                               | 79  |
| <b>MRI-A (Proposed)</b> | 79.6 $\pm$ 5.7<br>[68.8, 87.5]  | 66.5 $\pm$ 7.9<br>[52.4, 77.8]  | 85.7 $\pm$ 11.5<br>[71.5, 98.9] | 94.7 $\pm$ 10.7<br>[50.5, 99.7] | 216 |
| <b>MRI-B (Proposed)</b> | 81.6 $\pm$ 5.1<br>[71.3, 88.0]  | 69.2 $\pm$ 7.1<br>[55.5, 78.6]  | 82.2 $\pm$ 5.8<br>[73.1, 93.5]  | 84.8 $\pm$ 8.3<br>[69.4, 96.0]  | 132 |

ing tissue, and therefore affect the overall segmentation accuracy. Similarly, Figure 9 displays the visualisations of four different pancreas segmentation results in the CT-NIH dataset. There are clear, evident differences in the shape and structure of slice-by-slice pancreatic contours as well as the overall 3D reconstruction.

Similar to state-of-the-art automatic segmentation methods, every test image volume is processed using an approach generated by its corresponding training dataset. The morphological operations performed in the post-processing Section 2.3 involve principal geometric computations that are reproducible. The implementation of this section is also scalable, depending on the amount of training data and the number chosen for  $n$  clusters in the k-medoids stage.

The first two levels of refinement described in Sections 2.3.1 and 2.3.2 improve the overall mean DSC, proving considerably effective for individual cases that require varying degrees of refinement as opposed to some cases that do not require as much post-processing. After the first two levels of refinement, implementing the third level in Section 2.3.3 raises the accuracy in mean DSC by 4.59%, 4.74% and 4.68% for MRI-A, MRI-B and CT-NIH respectively. Empirically testing, in some cases, minor pancreatic contours were eliminated however incorporating this level of refinement also reduced the standard deviation by 2.73%, 2.45% and 2.58% for MRI-A, MRI-B and CT-NIH respectively.

The Hausdorff distance (HSD) metric (Huttenlocher et al., 1993) is presented as mean  $\pm$  standard deviation in CT-NIH, MRI-A and MRI-B as  $12.52 \pm 1.98mm$ ,  $13.59 \pm 4.34mm$  and  $14.28 \pm 4.02mm$ , respectively. This

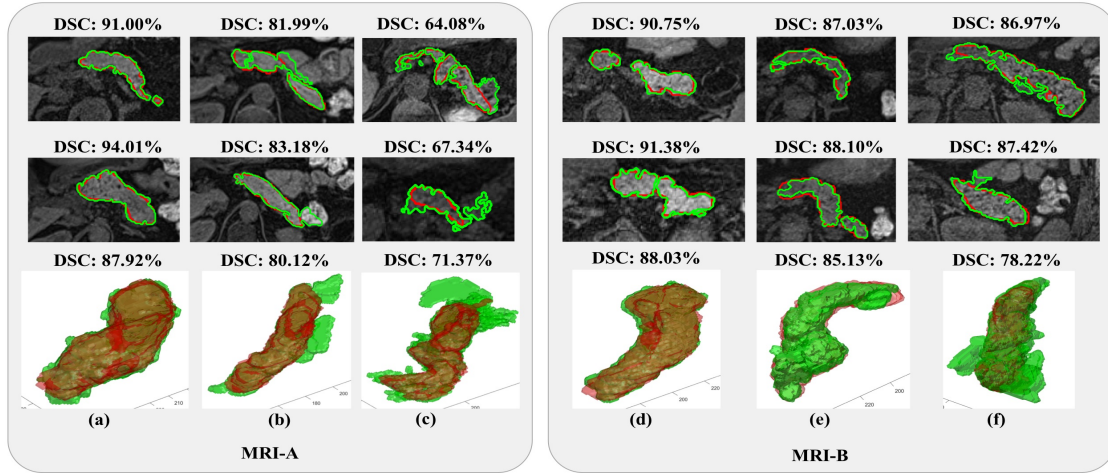


Figure 8: Segmentation results in six different MRI scans (volumes). Every column corresponds to a single MRI volume. From left, first row displays sample MRI axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed overall DSC.

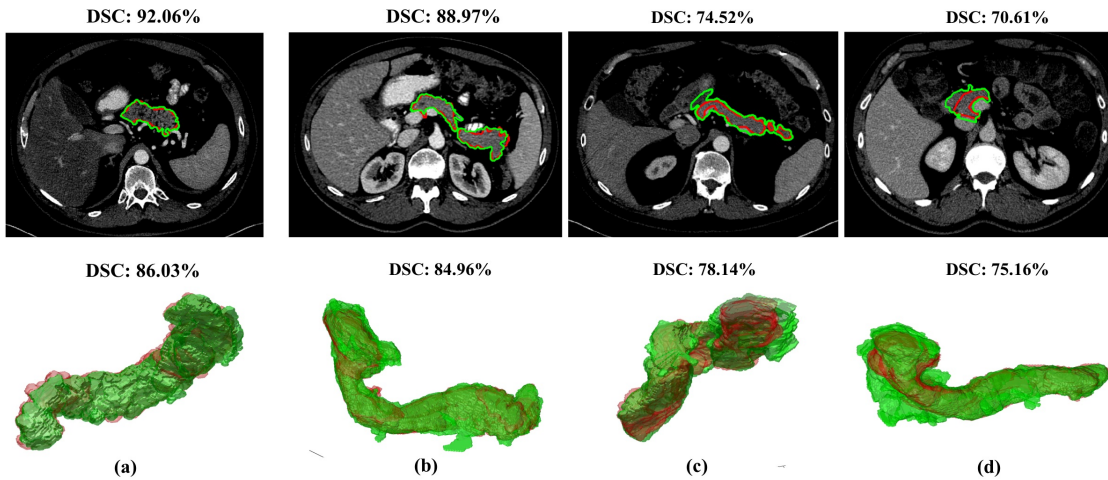


Figure 9: Segmentation results in four different CT image scans (volumes). Every column corresponds to a single CT volume. From left, first row displays sample CT axial slices with segmentation outcome (green) against ground-truth (red), and computed DSC; second row displays 3D reconstruction of entire pancreas (green) segmentation against its ground-truth (red) with computed overall DSC.



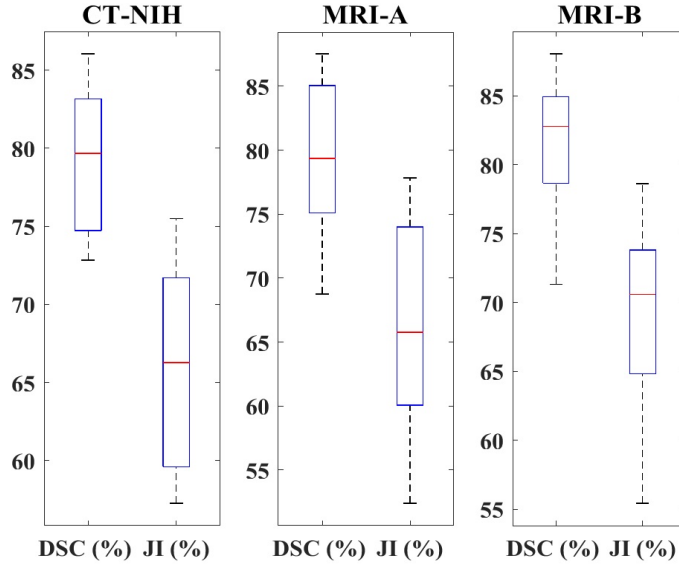


Figure 10: Dice Score Coefficient and Jaccard Index box plots for datasets CT-NIH, MRI-A and MRI-B.

metric, which represents the maximum deviation between two surfaces and is very sensitive to outliers, indicates that the approach is slightly more consistent for CT data. The relatively lower HSD in CT implies that slices have greater co-dependency, and therefore the proposed approach delivers a stable performance in predicting between successive slices. In contrast, the higher HSD value in MRI reflects less co-dependency between successive slices and higher shape variability. In such cases, the major pancreas region included excess background tissue that formed part of the final segmentation result, or eliminated pancreas tissue (mostly comprising part of the tail or body).

A box and whisker plot representation for all three datasets is displayed in Figure 10. The standard deviation is a reflection of intensity variation of surrounding tissue coming into contact with the final, resultant pancreas segmentation. In CT-NIH and MRI-B, 60% of all segmentation results score 80% or above in DSC, compared to 45% of all segmentation results in MRI-A. This difference in accuracy scores highlight the performance of the approach with respect to image quality and distribution of greyscale intensity in test image volumes. The broader range between median and upper quartile for MRI-A, in comparison to MRI-B, suggests a higher degree of variation in

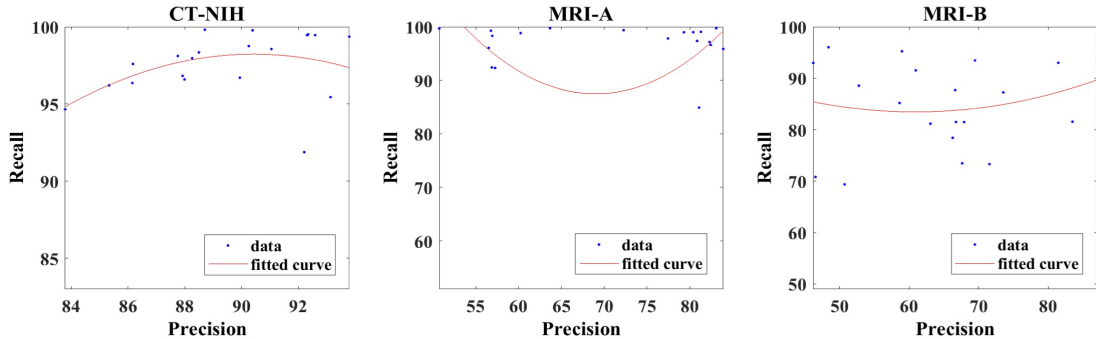


Figure 11: Recall against precision plots for datasets CT-NIH, MRI-A and MRI-B.

test volumes’ shape, size and level of coarse image noise. Thus, a tighter interquartile range in MRI-B indicates a similar degree of error within that dataset for all resultant segmentations. The relatively lower standard deviation in CT-NIH reflects a lesser degree of variation between image volumes as opposed to MRI-A and MRI-B that suffer from greater image noise and artefacts.

Precision and recall results in CT-NIH significantly outperform state-of-the-art approaches reported in Wolz et al. (2013, 2012), Farag et al. (2014, 2017). Therefore, comparatively high precision and high recall scores reflect a low false positive rate and low false negative rate. As shown in Figure 11, MRI-A and MRI-B report strong precision and recall results but the differences in standard deviation reflect higher variations in correct and incorrect classification of voxels. In general, the initial detection of the major pancreas region guarantees that the pancreas will be segmented (albeit with varying DSC accuracy) while maintaining higher mean precision and recall scores.

Unlike other publications that focus on one modality or one dataset for every modality, this method is robust to variations in noise, distortion, sharpness and changes in greyscale intensity distribution within multiple MRI datasets and a CT dataset. Figure 12 highlights three different slices from exemplar test volumes and their corresponding segmentation result. Notice the difference in levels of original noise and the comparable accurate segmentation outcome of approximately 85% in DSC.

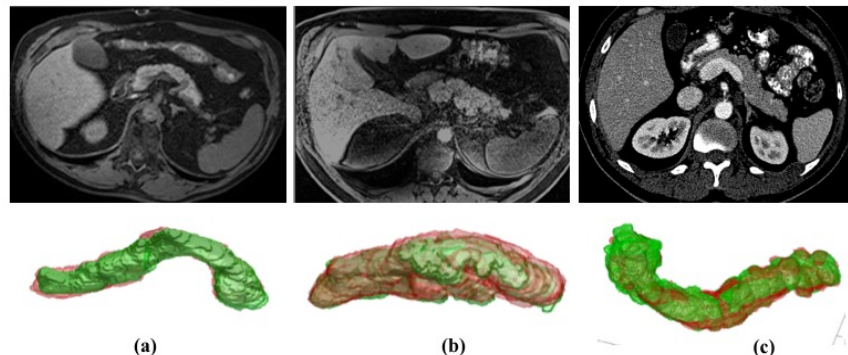


Figure 12: Columns (a) (b) and (c) highlights sample slice and corresponding segmentation outcome from MRI-A, MRI-B and CT-NIH respectively. Notice the variation in noise and image distortion between (a) and (b), and the difference in blur and sharpness between (a) and (c). All segmentation results (green) are mapped over the ground-truth (red) and have approximately 85% accuracy in DSC.

#### 4. Discussion

The approach presented in this paper reports better quantitative pancreas segmentation results in comparison to other state-of-the-art techniques (Roth et al., 2015a, Cai et al., 2016, Wolz et al., 2013, Okada et al., 2015, Tong et al., 2015, Farag et al., 2017, Karasawa et al., 2017).

With the exception of the publicly available CT dataset that has been utilised in publications (Roth et al., 2015a, 2017, Cai et al., 2017, Zhou et al., 2016, Roth et al., 2015b, 2016a) direct comparison with other methods in literature is difficult due to differences in modality, scanner imaging protocols and spatial resolution.

It should also be noted that the MRI data, evaluated for the approach proposed in this paper, was obtained using a scanner imaging protocol (sequence) that was not optimised for any particular organ, thus adding to the challenge of detecting the pancreas.

This proposed method delivers better quantitative accuracy and statistical stability, comparing 5.7% for MRI-A and 5.1% for MRI-B versus 8.7% (Cai et al., 2016) in DSC standard deviation. Although the mean MRI result reported in Cai et al. (2017) is relatively higher, the proposed method in this paper delivers a tighter standard deviation when compared to 6.7% in DSC as reported in the publication.

Employing 3D max-flow and min-cuts approach in combination with

meticulous geometrical descriptor analysis, the proposed approach advances the quantitative performance to mean DSC of 79.3% in testing CT-NIH, and reveals statistical significance (t-test,  $p < 0.0001$ ). This method performs close to the state-of-the-arts (Roth et al., 2017, Cai et al., 2017, Zhou et al., 2016) with approximate differences in mean DSC scores of 2%, 3% and 3% respectively.

The mean of listed DSC standard deviations from other reported publications in Table 1 is 13.1%, whereas the proposed method achieves approximately 3 times lower a value in standard deviation. In fact, the proposed method also delivers better statistical stability, comparing 4.40% versus 6.27% (Roth et al., 2017), 6.70% (Cai et al., 2017) and 5.68% (Zhou et al., 2016) in standard deviation of DSC scores.

A methodology reported in Suzuki et al. (2016) employs probabilistic atlas-based graph cut and achieves a mean JI coefficient of  $77.0 \pm 10.2\%$  for 15 CT volumes. Since the technique is interactive based, a medical expert identifies a “seed” (or location) from where the segmentation is performed. Similarly, the max-flow and min-cuts based approach, described in this paper, utilises rich statistical information about wide variations and irregularities in pancreas shape and size. However, the proposed method for this paper is fully automatic and does not require user intervention. Furthermore, the choice of modality also includes MRI and therefore tackles greater problems relating to image artefacts.

The approach reported in Gou et al. (2014) performs segmentation on MRI data, though it should be noted the images are dynamic MRI (contrast-enhanced). Nevertheless, this method also involves operator intervention where “seeds” are manually placed near the centre of the regions of interest (ROIs) prior to segmentation. Although the accuracy score for this method boasts a mean DSC of 77.0%, only two image volumes are used for testing purposes. The max-flow and min-cuts based approach produces comparably more robust results for two different MRI datasets that were generated using a scanner imaging protocol (sequence) that is not optimised for any organ, let alone the pancreas.

*Qualitative assessment.* The feedback received by an expert radiographer and an expert radiologist reveals that the proposed method produces detailed contouring of the pancreas, for every protrusion and indentation as opposed to an approximate or mean tracing of the organ. The primary source of segmentation error can be attributed to the *remaining* accumulated

surrounding pancreas tissue including the superior mesenteric vein, splenic artery and duodenum.

A potential method for optimisation includes removal of surrounding splenic artery and vessels *prior* to the max-flow and min-cuts segmentation stage. As reported in Erdt et al. (2011), a number of clinical support structures are implemented to detect the splenic artery by learning the spatial relationship between the vessel path, and local pancreas texture and position. Although the modality of choice is CT, giving relatively clearer distinction between organs and vessels in comparison to MRI, the concept can still extend to MRI data. Thus, incorporating further anatomical knowledge relating to abdominal organs, in combination with radiological information about tissue surrounding the pancreas, can potentially raise the overall accuracy score.

Another approach for method optimisation includes utilising a 3D deep learning network that can exploit and learn multilevel features of tissue surrounding or “touching” the pancreas. This would require the corresponding ground-truth information to train the network. Such a technique can be applied before the max-flow and min-cuts stage in order to further eliminate unwanted veins and artery.

Although deep learning methods deliver highly promising results for 2D image data or slice-by-slice image analysis, existing methods struggle to serve the 3D data that drives medical imaging, especially for MRI. Despite evidence to suggest that 3D deep learning networks suffer from overfitting (Qi et al., 2016, Roth et al., 2016) and would require more training data, recent publications suggest that data augmentation and 3D convolutional neural network models (Çiçek et al., 2016, Bui et al., 2017) can produce improved volumetric segmentation results without adding to the cost of computational time.

In many a case, the duodenum, which touches the head of the pancreas, often “sticks” to the resultant segmentation: this occurs whenever a distinction cannot be made between the greyscale intensities of the pancreas and duodenum at the max-flow and min-cuts segmentation stage. Often, the subtle detail separating the pancreas head and duodenum are detected by medical experts through analysis of the image slices in a back-and-forth manner, whereby identifying the features or slight contours required to establish organ separation. In fact, it can be possible to “strengthen” the detection of weak boundaries between the duodenum and pancreas head in a manner that is similar to the approach in Xiang et al. (2017) for the renal cortex and kidneys. Here, a statistical shape model for the area of interest is developed to automatically initialise the outer surface and inner surface.

Afterwards, a graph-cuts algorithm can be introduced to detect the optimal boundary of the organ.

Overall, an area to investigate includes incorporating such meticulous information (e.g. appearance texture descriptors) into a statistical model or a deep learning network that can identify key duodenum landmarks in an image *prior* to the main segmentation process. Therefore, such a pre-processing model that is generated using heavy data augmentation can potentially minimise false predictions and even reduce computation time at the post-processing stage.

*Computational efficiency.* The overall mean runtime per CT and MRI volume is 30 and 22 minutes respectively, and not two or four hours as in Chu et al. (2013), Wolz et al. (2013), Karasawa et al. (2017). Furthermore, the proposed approach does not involve the computationally expensive costs of training a network, as in 9 hours (Farag et al., 2017) or 55 hours (Roth et al., 2015a) before segmenting a test image volume. With the exception of a random forest, trained in just under 10 minutes to identify the major pancreas region, the approach described in this paper evaluates an image volume within 10 to 35 minutes via a workstation with i7-59-30k-CPU at 3.50 Ghz. In the future, this run-time can be potentially reduced by a factor of 10 by using a GeForce Titan X GPU.

## 5. Conclusion

This paper presents an approach for automatic pancreas segmentation in 3D abdominal computed tomography (CT), and magnetic resonance imaging (MRI) scans. The proposed method is based on a hierarchical pooling of information by classifying extracted image patches, superpixels and intensity distributions as pancreatic tissue or otherwise. A sequential process firstly improves the difference between pancreatic and background classes using optimised parameters from an intensity model. Next, the major pancreas region is identified using a trained random forest that makes probability-wise predictions of superpixels in each slice per image volume. This technique, followed by 3D max-flow and min-cuts segmentation and structured forest edge detection, produces very detailed contouring of the pancreas and heavy boundary-preservation. This voxel-based algorithm addresses the intensity consistency problem that is often the case when segmenting image volumes on a slice-by-slice basis. Furthermore, a novel post-processing stage optimises

tissue classification using morphological, anatomical and radiological knowledge about connectivity between pancreatic contours in an image volume. The final quantitative pancreas segmentation results are better than or close to state-of-the-art approaches for both CT and MRI modality, and report higher statistical stability with lower standard deviation metrics. Given the wide variation in datasets, this methodology can be further optimised and applicable to other abdominal MRI and CT sequences and also, generalisable to other organ or muscular tissue segmentation tasks.

### **Conflict of interest statement**

The authors declare that they have no conflict of interest.

### **Acknowledgment**

The authors would like to thank Professor Jimmy Bell and his group (at the University of Westminster, School of Life Sciences) for providing an extensive medical imaging database that fosters research for health and wellbeing.

### **References**

- Asaturyan, H. and Villarini, B. (2018), Hierarchical framework for automatic pancreas segmentation in MRI using continuous max-flow and min-cuts approach, *in* ‘Image Analysis and Recognition’, Springer International Publishing, pp. 562–570.
- Buades, A., Coll, B. and Morel, J.-M. (2005), A non-local algorithm for image denoising, *in* ‘Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on’, Vol. 2, IEEE, pp. 60–65.
- Bui, T. D., Shin, J. and Moon, T. (2017), ‘3D densely convolutional networks for volumetric segmentation’, *CoRR* **abs/1709.03199**.
- Burute, N., Nisenbaum, R., Jenkins, D. J., Mirrahimi, A., Anthwal, S., Colak, E. and Kirpalani, A. (2014), ‘Pancreas volume measurement in patients with Type 2 diabetes using magnetic resonance imaging-based planimetry’, *Pancreatology* **14**(4), 268–274.

- Cai, J., Lu, L., Xie, Y., Xing, F. and Yang, L. (2017), ‘Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function’, *CoRR* **abs/1707.04912**.
- Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L. and Yin, Q. (2016), *Pancreas Segmentation in MRI Using Graph-Based Decision Fusion on Convolutional Neural Networks*, Springer International Publishing, Cham, pp. 442–450.
- Chu, C., Oda, M., Kitasaka, T., Misawa, K., Fujiwara, M., Hayashi, Y., Nimura, Y., Rueckert, D. and Mori, K. (2013), Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images, *in* ‘Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013’, Springer Berlin Heidelberg, pp. 165–172.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. and Ronneberger, O. (2016), ‘3D U-Net: Learning dense volumetric segmentation from sparse annotation’, *CoRR* **abs/1606.06650**.
- Cruickshank, A. H. and Benbow, E. W. (1995), *Pathology of the Pancreas*, Springer-Verlag London, pp. 10–15.
- Dollár, P. and Zitnick, C. L. (2013), Structured forests for fast edge detection, *in* ‘proceedings of the 2013 IEEE International Conference on Computer Vision’, ICCV ’13, IEEE Computer Society, Washington, DC, USA, pp. 1841–1848.
- Dosselmann, R. and Yang, X. D. (2011), ‘A comprehensive assessment of the structural similarity index’, *Signal, Image and Video Processing* **5**(1), 81–91.
- Dubuisson, M. and Jain, A. (1994), A modified hausdorff distance for object matching, *in* ‘Proceedings of the 12th IAPR International Conference’, Vol. 1, pp. 566–568.
- Erdt, M., Kirschner, M., Drechsler, K., Wesarg, S., Hammon, M. and Cavallo, A. (2011), ‘Automatic pancreas segmentation in contrast enhanced ct data using learned spatial anatomy and texture descriptors’, *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* pp. 2076–2082.



- Farag, A., Lu, L., Roth, H. R., Liu, J., Turkbey, E. and Summers, R. M. (2017), ‘A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling’, *IEEE Transactions on Image Processing* **26**(1), 386–399.
- Farag, A., Lu, L., Turkbey, E., Liu, J. and M. Summers, R. (2014), ‘A bottom-up approach for automatic pancreas segmentation in abdominal CT scans’, *CoRR abs/1407.8497*, 103–113.
- Gou, S., Wu, J., Liu, F., Lee, P., Rapacchi, S., Hu, P. and Sheng, K. (2014), ‘Feasibility of automated pancreas segmentation based on dynamic MRI’, *Br J Radiol* **87**(1044).
- Hagen-Ansert, S. (2017), *Textbook of Diagnostic Sonography*, Mosby, pp. 305–308.
- Hammon, M., Cavallaro, A., Erdt, M., Dankerl, P., Kirschner, M., Drechsler, K., Wesarg, S., Uder, M. and Janka, R. (2013), ‘Model-based pancreas segmentation in portal venous phase contrast-enhanced ct images’, *Journal of Digital Imaging* **26**(6), 1082–1090.
- Huttenlocher, D. P., Klanderman, G. A. and Rucklidge, W. J. (1993), ‘Comparing images using the hausdorff distance’, *IEEE Transactions on pattern analysis and machine intelligence* **15**(9), 850–863.
- J. Williams, D. and Shah, M. (1992), ‘A fast algorithm for active contours and curvature estimation’, *CVGIP: Image Understanding* **55**, 14–26.
- Karasawa, K., Oda, M., Kitasaka, T., Misawa, K., Fujiwara, M., Chengwen, C., Zheng, G., Rueckert, D. and Mori, K. (2017), ‘Multi-atlas pancreas segmentation: Atlas selection based on vessel structure’, *Medical Image Analysis* **39**, 18–28.
- Milletari, F., Navab, N. and Ahmadi, S. A. (2016), V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in ‘2016 Fourth International Conference on 3D Vision (3DV)’, pp. 565–571.
- Oda, M., Shimizu, N., Karasawa, K., Nimura, Y., Kitasaka, T., Misawa, K., Fujiwara, M., Rueckert, D. and Mori, K. (2016), Regression forest-based atlas localization and direction specific atlas generation for pancreas

- segmentation, *in* ‘Medical Image Computing and Computer-Assisted Intervention’, Springer International Publishing, pp. 556–563.
- Okada, T., Linguraru, M. G., Hori, M., Summers, R., Tomiyama, N. and Sato, Y. (2015), ‘Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors’, *Medical Image Analysis* **26**, 1–18.
- Okada, T., Linguraru, M. G., Yoshida, Y., Hori, M., Summers, R. M., Chen, Y.-W., Tomiyama, N. and Sato, Y. (2012), Abdominal multi-organ segmentation of ct images based on hierarchical spatial modeling of organ interrelations, *in* ‘Abdominal Imaging. Computational and Clinical Applications’, Springer Berlin Heidelberg, pp. 173–180.
- Omeri, A. K., Matsumoto, S., Kiyonaga, M., Takaji, R., Yamada, Y., Kosen, K., Mori, H. and Miyake, H. (2017), ‘Contour variations of the body and tail of the pancreas: evaluation with mdct’, *Jpn J Radiol* **35**(6), 310–318.
- Qi, C. R., Su, H., Niessner, M., Dai, A., Yan, M. and Guibas, L. J. (2016), ‘Volumetric and multi-view CNNs for object classification on 3D data’, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 5648–5656.
- Roth, H., Farag, A., Lu, L., Sohn, A. and Summers, R. M. (2016a), Spatial aggregation of holistically-nested networks for automated pancreas segmentation, *in* ‘MICCAI’, pp. 451–450.
- Roth, H., Farag, A., Lu, L., Turkbey, E. and Summers, R. M. (2015b), Deep convolutional networks for pancreas segmentation in CT imaging, *in* ‘Medical Imaging: Image Processing’, Vol. 9413.
- Roth, H., Lu, L., Farag, A., Shin, H.-C., Liu, J., Turkbey, E. and Summers, R. (2015a), *DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation*, Springer International Publishing, pp. 556–564.
- Roth, H. R., Lu, L., Lay, N., Harrison, A. P., Farag, A., Sohn, A. and Summers, R. M. (2017), ‘Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation’, *CoRR* **abs/1702.00045**.

- Roth, H. R., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., Kim, L. and Summers, R. M. (2016), ‘Improving computer-aided detection using convolutional neural networks and random view aggregation’, *IEEE Transactions on Medical Imaging* **35**(5), 1170–1181.
- Saito, A., Nawano, S. and 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’, *Medical Image Analysis* **28**, 46–65.
- Shimizu, A., Kimoto, T., Kobatake, H., Nawano, S. and Shinozaki, K. (2010), ‘Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography’, *Int J Comput Assist Radiol Surg* **5**(1), 85–98.
- Suzuki, T., Takizawa, H., Kudo, H. and Okada, T. (2016), ‘Interactive segmentation of pancreases from abdominal CT images by use of the graph cut technique with probabilistic atlases’, *BioMed Res. Int* **45**, 575–584.
- Taha, A. A. and Hanbury, A. (2015), ‘Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool’, *BMC Medical Imaging* **15**(1), 29.
- Tong, T., Wolz, R., Wang, Z., Gao, Q., Misawa, K., Fujiwara, M., Mori, K., Hajnal, J. V. and Rueckert, D. (2015), ‘Discriminative dictionary learning for abdominal multi-organ segmentation’, *Med Image Anal* **23**(1), 92–04.
- Villarini, B., Asaturyan, H., Thomas, E., Mould, R. and Bell, J. (2017), A framework for morphological feature extraction of organs from MR images for detection and classification of abnormalities, *in* ‘2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)’, IEEE, pp. 666–671.
- Wang, Z., Bhatia, K. K., Glocker, B., Marvao, A., Dawes, T., Misawa, K., Mori, K. and Rueckert, D. (2014), *Geodesic Patch-Based Segmentation*, Springer International Publishing, pp. 666–673.
- Wolz, R., Chengwen, C., Misawa, K., Fujiwara, M., Mori, K. and Rueckert, D. (2013), ‘Automated abdominal multi-organ segmentation with subject-specific atlas generation’, *IEEE Transactions on Medical Imaging* **32**(9), 1723–1730.

- Wolz, R., Chu, C., Misawa, K., Mori, K. and Rueckert, D. (2012), Multi-organ abdominal ct segmentation using hierarchically weighted subject-specific atlases, *in* ‘Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012’, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 10–17.
- Xiang, D., Bagci, U., Jin, C., Shi, F., Zhu, W., Yao, J., Sonka, M. and Chen, X. (2017), ‘Cortexpert: A model-based method for automatic renal cortex segmentation’, *Medical Image Analysis* **42**, 257–273.
- Yeghiazaryan, V. and Voiculescu, I. (2015), An overview of current evaluation methods used in medical image segmentation, Technical report, Tech. Rep. CS-RR-15-08, Department of Computer Science, University of Oxford, Oxford, UK.
- Yuan, J., Bae, E. and Tai, X.-C. (2010), A study on continuous max-flow and min-cut approaches, *in* ‘Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, pp. 2217–2224.
- Zhou, Y., Xie, L., Shen, W., Fishman, E. K. and Yuille, A. L. (2016), ‘Pancreas segmentation in abdominal CT scan: A coarse-to-fine approach’, *CoRR* **abs/1612.08230**.