Discriminative Dictionary Learning for Abdominal Multi-Organ Segmentation

Tong Tong[a,*], Robin Wolz[a], Zehan Wang[a], Qinquan Gao[a], Kazunari Misawa[b], Michitaka Fujiwara[c], Kensaku Mori[d], Joseph V. Hajnal[e], Daniel Rueckert[a]

[a]*Biomedical Image Analysis Group, Department of Computing, Imperial College London, 180 Queen's Gate, London, SW7 2AZ, UK*
[b]*Aichi Cancer Center, Nagoya 464-8681, Japan*
[c]*Nagoya University Hospital, Nagoya 466-0065, Japan*
[d]*Information and Communications, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan*
[e]*Center for the Developing Brain, Division of Imaging Sciences and Biomedical Engineering, King's College London, St. Thomas Hospital, London, SE1 7EH, UK*

Abstract

An automated segmentation method is presented for multi-organ segmentation in abdominal CT images. Dictionary learning and sparse coding techniques are used in the proposed method to generate target specific priors for segmentation. The method simultaneously learns dictionaries which have reconstructive power and classifiers which have discriminative ability from a set of selected atlases. Based on the learnt dictionaries and classifiers, probabilistic atlases are then generated to provide priors for the segmentation of unseen target images. The final segmentation is obtained by applying a post-processing step based on a graph-cuts method. In addition, this paper proposes a voxel-wise local atlas selection strategy to deal with high inter-subject variation in abdominal CT images. The segmentation performance of the proposed method with different atlas selection strategies are also compared. Our proposed method has been evaluated on a database of 150 abdominal CT images and achieves a promising segmentation performance with Dice overlap values of 94.9%, 93.6%, 71.1%, and 92.5% for liver, kidneys, pancreas, and spleen, respectively.

Keywords: abdominal multi-organ segmentation, discriminative dictionary learning, local atlas selection, patch based

1. Introduction

CT-based clinical assessment of abdominal organs relies on quantitative measures and comprehensive analysis of multiple organs in order to identify disorders (Linguraru et al., 2012). The segmentation of multiple abdominal organs enables quantitative analysis of different organs, providing invaluable input for computer aided diagnosis (CAD) systems. For instance, liver segmentation is helpful in the automatic detection and definition of focal lesions (Liu et al., 2004). The segmentation of the pancreas facilitate the diagnosis of dilated pancreatic ducts or inflamed pancreatic tissues (Shimizu et al., 2010). The measurement of the size of the kidney is useful in evaluating its conditions. Other applications like radiotherapy planning as well as cancer detection and staging also require the accurate segmentation of abdominal organs. An automated segmentation approach can eliminate the need for manual labeling by trained observers, i.e. radiologists. Many segmentation approaches have been developed for abdominal computed tomography (CT) scans in recent years. Most of these approaches are based on statistical shape models (Heimann et al., 2006; Okada et al., 2008b; Spiegel et al., 2009; Wimmer et al., 2009; Cerrolaza et al., 2014) or multi-atlas

*Corresponding author. Tel: +44 20 7852 1982
Email address: t.tong11@imperial.ac.uk (Tong Tong)

segmentation (Park et al., 2003; Okada et al., 2008b; Shimizu et al., 2011; Wolz et al., 2013; Wang et al., 2014). In these methods, an shape model or a probability atlas is calculated by averaging shape or location priors of multiple spatially aligned atlases. Such statistical shape models or probabilistic atlases can then provide prior knowledge for the segmentation of organs in the target image. A combination of statistical shape models and probabilistic atlases has also been proposed (Okada et al., 2008a; Wang et al., 2012; Okada et al., 2013) to incorporate both shape and location priors for segmentation tasks.

Since the introduction of statistical shape models in the early 1990s (Cootes et al., 1995), these models have been proven very effective in various image segmentation applications. Gao et al. (1998) presented early work using statistical shape models for segmentation of abdominal organs. Later, Heimann et al. (2006) showed a successful application of an active shape model in the segmentation of the liver in CT scans. An automated segmentation method using statistical shape models was proposed by Shimizu et al. (2010) to successfully segment the pancreas. An enhanced shape model approach that integrates a hierarchical framework was also proposed in Bagci et al. (2012) for improving the segmentation accuracy. Another interesting study was presented in Cerrolaza et al. (2014) which introduces a generalized multiresolution hierarchical shape model to efficiently describe the shape variability of different organs to improve the segmentation performance of statistical shape models.

Early work using probabilistic atlases was described in Park et al. (2003), where a statistical atlas of the liver and the kidneys was shown to be helpful for the segmentation of these organs. Recent work have incorporated spatial priori knowledge for different abdominal organs (Okada et al., 2008a; Linguraru et al., 2010; Shimizu et al., 2011; Oda et al., 2012; Linguraru et al., 2012; Wolz et al., 2013; Wang et al., 2014). Notably, Okada et al. (2008a) constructed a hierarchical multi-organ statistical atlas for improving segmentation performance. In order to generate more specific atlases, Oda et al. (2012) separated an atlas database into several clusters and multiple probabilistic atlases were generated. Also recently, inter-organ spatial relations have been incorporated into the probabilistic atlases to perform multi-organ segmentation (Okada et al., 2013; Cerrolaza et al., 2014; Wang et al., 2014).

The aim in building models from a population in the form of statistical shape models or probabilistic atlases is that the constructed models can match to the shape or appearance of the anatomical structure of interest of new images. However, the average models calculated from a given population describe the full variability in this specific dataset, potentially leading to a low specificity with respect to individual appearance. The generality of such average models may hamper the segmentation of a specific target image due to large inter-subject variability. For example, difficuties may arise in the segmentation of the target images whose anatomical shapes or locations differ significantly from the average model. To address these shortcomings, more recent approaches are based on subject-specific shape models (Wang et al., 2010) or subject-specific probabilistic atlases (Shimizu et al., 2010; Wolz et al., 2013; Chu et al., 2013), generating subject-specific priors for unlabeled images instead of sharing the same average shape or location priors. The subject-specific models are generated by identifying a number of suitable atlases and then fusing their priors. In order to generate good subject-specific priors for segmentation, two steps are crucial in these methods: selecting similar atlases to the target images and performing accurate pairwise registrations.

Previous studies (Aljabar et al., 2009) show that the segmentation performance of multi-atlas based methods is highly dependant on the selected atlases for the target image. Most atlas selection methods define a global mask region to include multiple organs of interest. Then, global similarity measures are calculated in this predefined mask between the target and atlas images to select suitable atlases. More advanced methods (Wolz et al., 2010; Cao et al., 2011) transfer the global similarities into a manifold and perform the atlas selection in the learnt manifold. However, the global similarities represent the overall differences in the mask, which are dominated by the large organs. For example, in our application of abdominal segmentation, the atlas selection is likely to be dominated by the liver since the liver is much larger than other organs. This means that the selected "similar" atlases may not be similar in some local regions such as in the pancreas. A region-wise local atlas selection strategy (van Rikxoort et al., 2010; Shi et al., 2010; Wolz et al., 2013) has been utilized to overcome this shortcoming by selecting suitable atlases at each local region. However, these approaches require the separation of the whole image into different local regions and non-rigid registrations are performed over these local regions for accurate label fusion. Since different anatomical patterns exist at different locations, a voxel-wise local comparison strategy may provide

a better way to select similar atlases at each location.

The other drawback of traditional multi-atlas based methods is that accurate pairwise registrations are needed to acquire good segmentation results. This can be problematic in the case of high inter-subject variability. Another challenge that arises from using non-rigid registration is the highly computational burden. Previous studies (Wolz et al., 2013; Chu et al., 2013) have demonstrated that the computational complexity is largely defined by the computational time required for the non-rigid registration step. Recently, nonlocal patch based segmentation (PBS) method (Coupé et al., 2011; Rousseau et al., 2011) has been proposed to avoid the need of accurate non-rigid registration and demonstrated the successful applications on the segmentation of brain MR images. However, the patch-based segmentation method cannot be directly applied to the segmentation of abdominal organs, because i) unlike the human brain, the anatomy in abdominal region shows great variability. There is significant variation in the shapes, sizes and locations of the abdominal organs especially the pancreas, making the overall image alignment particularly challenging (Wolz et al., 2013; Wang et al., 2014). This will pose difficulties for the segmentation methods that rely on image registrations. Although only affine registration is required for the patch based segmentation method, Rousseau et al. (2011) argued that more accurate registration is beneficial to improve the segmentation accuracy of patch-based segmentation methods; ii) the computational complexity becomes a significant problem for large abdominal organs.

To address the above problems, a novel patch-based segmentation framework is presented for the abdominal multi-organ segmentation. In our previous work (Tong et al., 2013), a dictionary learning technique was introduced to improve the segmentation performance of the patch-based methods. However, this approach was limited to binary segmentation and only evaluated on the hippocampus labeling. In this paper, we extend our previous method (Tong et al., 2013) for the simultaneous segmentation of multiple structures. Furthermore, we evaluate the approach on abdominal multi-organ segmentation from CT images. Specifically, dictionaries and classifiers are learnt from the selected training atlases, which will then be utilized to generate a subject-specific probabilistic atlas for each unlabeled target image. The final segmentation is obtained by applying a post processing step based on graph-cuts in combination with the generated subject-specific probabilistic atlases (Wolz et al., 2013; Chu et al., 2013). The main contributions of this work can be summarized as follows: (1) The extension of discriminative dictionary learning for segmentation (DDLS) algorithm for the segmentation of multiple organs in CT images; (2) A local voxel-wise atlas selection in order to capture local information for segmentation and to tackle the high inter-subject variability; (3) A comparison between different atlas selection strategies; (4) A multi-resolution strategy for gaining computational efficiency. In the remainder of the paper, we will first introduce the datasets used in our work in Section 2.1. The methodology of DDLS for multi-organ segmentation is introduced in Section 2.4 and different atlas selection strategies are also presented. The performance of the proposed method is analyzed in Section 3. Finally, we discuss the strengths and weaknesses of the proposed method and conclude this paper.

2. Material and Methods

2.1. Dataset

150 3-D abdominal CT scans acquired from 36 female and 114 male subjects were used for our experiments. All scans were acquired between 2004 and 2009 at Nagoya University hospital by a TOSHIBA Aquilion 64 scanner and obtained under typical clinical protocols for the purpose of laparoscopic resection of the stomach and gallbladder glands or colon. Among the 150 CT scans, 141 subjects had early or advanced gastric cancer, one subject had cholecystitis cancer and eight subjects had colorectal cancer. All subjects were aged between 26 and 84 years with a mean age of 62.8±12.0. Scans have a resolution of 512×512 voxels in plane and contain between 238 and 1061 slices depending on the field-of-view and the slice thickness. Voxel sizes range from 0.55 to 0.82 mm and the slice spacing varies from 0.4 to 0.8 mm. The X-ray tube voltage is 120 kV and the X-ray tube current is 350-400 mAs. All of the images were acquired in portal venous phase (20-30 s delayed from starting point). The starting point of scanning was chosen according to the following rules: for patients who were younger than 60 years, the starting point was set as 25 s delayed

from the injection point; for other patients, the scan started after 7 s when the intensity of the aorta is over 80 HU. Scanning control is performed by utilizing the Toshiba Real Prep System. Images were acquired under typical clinical conditions and therefore show typical contrast variations. Images start anterior at the lungs and are automatically cropped at 25 cm in the axial direction.

Reference segmentations are available for the liver, spleen, pancreas and the kidneys. The segmentations are used as atlases. All 150 subjects were segmented by one out of three trained raters. The reference segmentations are based on interactive region growing, where a spherical element is utilized to prevent excess segmentation of a target region, or graph-cut segmentation, where a set of foreground and background voxels are manually set as seed points. After the semi-automated segmentation, a manual correction step was performed on the axial, coronal, or sagittal slices.

Algorithm 1 DDLS with Local Atlas Selection [a]

Input: A target image I_t; A set of training Atlases: images $A = \{A_1, A_2, \cdots, A_N\}$ and labels $S = \{S_1, S_2, \cdots, S_N\}$; Parameters: β_1 and β_2.

Output: A label map S_t.

1: Affinely align atlases A to the target space.
2: **for** each target voxel in I_t **do**
3: Perform local atlas selection.
4: Extract training patches in a constraint search neighborhood from the selected atlas images and form a training patch library: $P_L = [p_1, p_2, \cdots, p_n] \in R^{m \times n}$
5: **Discriminative dictionary training:**
6: $\left\langle \hat{D}, \hat{W}, \hat{\alpha} \right\rangle = \underset{D,W,\alpha}{\arg\min} \|\mathrm{P_L} - D\alpha\|_2^2 + \beta_1 \|H - W\alpha\|_2^2 + \beta_2 \|\alpha\|_1.$
7: **Sparse coding for the target patch p_t:**
8: $\hat{\alpha}_t = \underset{\alpha_t}{\arg\min} \left\| p_t - \hat{D}\alpha_t \right\|_2^2 + \beta_2 \|\alpha_t\|_1$
9: **Probabilistic labels estimation :**
10: h_t=$\hat{\mathrm{W}}\hat{\alpha}_t$
11: **end for**
12: Obtain final segmentation label maps S_t using graph cuts.

[a] Algorithm of the proposed discriminative dictionary learning for segmentation (DDLS) with local atlas selection strategy (L-DDLS). β_1 and β_2 are parameters in the dictionary learning and sparse coding process. P_L is the training patch library extracted from selected atlases. $\hat{D}$ and $\hat{W}$ represent the learned dictionary and the classifier respectively from P_L. p_t is the target patch under study and h_t is the estimated probabilistic labels for p_t.

2.2. Overview

There are three major steps in the proposed method. *Step 1*: After all the training atlases are affinely aligned to the target space, atlas selection is performed. Similar atlases can be selected by calculating similarity measures over the whole image or in a local mask. *Step 2*: Training patches are extracted from the selected atlas images within a search volume to form a training patch library P_L. Then, a dictionary D with reconstruction power and a classifier W with discriminative ability are learnt simultaneously from P_L and their corresponding labels. A probabilistic label is then estimated for each target voxel. In the end, a subject-specific probabilistic atlas is generated for each target image. *Step 3*: Based on the subject-specific probabilistic atlas, the final segmentation is obtained by using the graph-cuts method as proposed in Wolz et al. (2013). Algorithm 1 outlines the major steps of the proposed DDLS approach with a local atlas selection strategy. Details of these three steps are described in the following sections.

2.3. Atlas selection

In traditional atlas selection (Aljabar et al., 2009), the whole image is treated as a single entity for calculating inter-subject pairwise similarity measures. As a result, the selected atlases are shared by all voxels in the target image. However, it may not be optimal to utilize the same atlases at different locations.

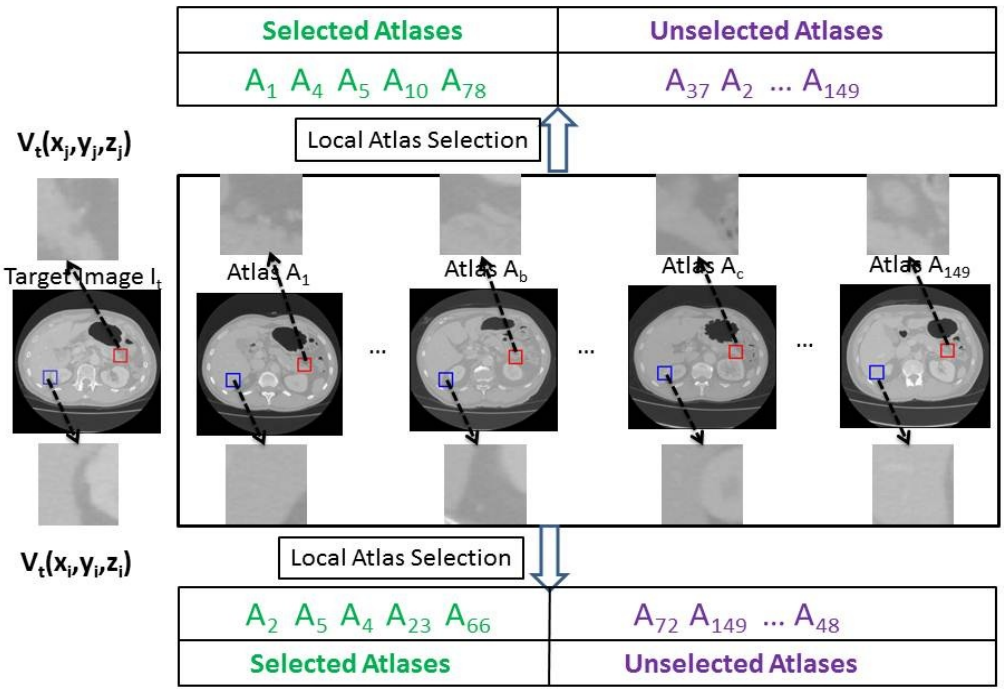

Figure 1: Demonstration of the voxel-wise local atlas selection strategy. At different locations in the target image I_t, different subsets of atlases are selected. Atlases A_2, A_5, A_4, A_{23} and A_{66} are selected at location (x_i, y_i, z_i) since these atlases have similar local intensity patterns with that of the target image at this location. When the target voxel v_t is at location (x_j, y_j, z_j), atlases A_1, A_4, A_5, A_{10} and A_{78} are selected.

Assume that a target voxel $v_t\,(x_i, y_i, z_i)$ is in the liver region of a target image I_t as shown in Figure 1, and atlas A_b is selected because A_b has a similar liver and shows similar anatomical patterns at location (x_i, y_i, z_i). If the target voxel $v_t\,(x_j, y_j, z_j)$ moves into the pancreas region of image I_t, it is possible that atlas $A_c\,(c \neq b)$ is selected at location (x_j, y_j, z_j) because atlas A_c contains more similar anatomical patterns with the target image I_t at this location than atlas A_b. Therefore, we propose to use a voxel-wise local atlas selection strategy to capture the important local information for segmentation. Figure 2 shows an example of the local atlas selection at different locations for a target image. As can be seen from this figure, the most similar atlas is different at different locations. In addition, same atlases are selected at neighboring voxels in homogeneous regions. The extent of the local mask influences the behaviour of the atlas selection: Larger masks mean that the atlas selection is more global (in the limit the mask can be the size of the image) and smaller masks lead to more local behaviour (atlases are selected based on more local intensity patterns). If the size of the mask is as large as the image, local atlas selection will be equivalent to the global atlas selection. In this case, the selected atlases are the same at all locations in the target image.

In this paper, we propose novel DDLS methods that allow either global or local atlas selection strategies, which are denoted as G-DDLS and L-DDLS respectively. In G-DDLS, a global mask (i.e. the whole image) is first defined. Then, a set of atlases is selected for each target image according to the similarities between the atlas images and the target image within this mask. In contrast to this, in L-DDLS, a voxel-wise atlas selection is carried out to select similar atlases locally at different locations in the target image. This means that different sets of atlases can be selected at different locations in the target image. For a target voxel v_t at location (x, y, z), a local neighborhood is defined as shown in Figure 2. Then, pairwise similarities at this location between atlas images and the target image are calculated within this local mask. Different similarity measures such as the squared intensity differences (SSD), cross-correlation or mutual information (Pluim et al., 2003) can be used. Finally, K atlases are selected at location (x, y, z) for the target voxel v_t according to the local similarity measures.

In our previous work (Tong et al., 2013), DDLS with a fixed-atlas strategy was proposed which we denoted as F-DDLS. In F-DDLS, a subgroup of the whole dataset is randomly selected as the fixed training atlases. Discriminative dictionaries are then trained from these randomly selected training atlases offline. After that, the segmentation is performed on the remaining test subjects online. In contrast to G-DDLS and L-DDLS which select subject-specific atlases for training, F-DDLS uses fixed atlases for training. The advantage of F-DDLS is that it can yield a significant speed-up in the segmentation process since the dictionaries are learnt offline and kept fixed.

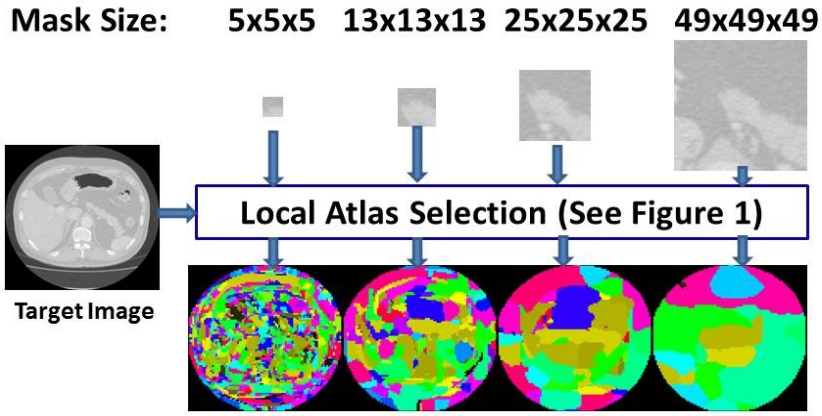

Figure 2: Example demonstrating the local atlas selection for different local mask sizes. The color maps show the most similar atlas selected at different locations in the target image. Different colors mean that different atlases are selected at different locations.

2.4. DDLS for Multiple Structures

For labeling a target voxel v_t in the target image I_t, the surrounding patch of v_t is extracted and denoted as the target patch $p_t \in R^{m\times 1}$. Here, the m intensity values within the patch are arranged into a m-dimensional feature vector. A search volume is defined in each selected atlas image A_i. All template patches in the search volume across the K selected similar atlases are extracted to form a training patch library P_L. Assuming that the patch library contains n training patches, the patch library can then be represented as $P_L = [p_1, p_2, \cdots, p_n] \in R^{m\times n}$. A reconstructive dictionary $\hat{D} \in R^{m\times d}$ with d atoms can be learnt from the input patch library $P_L \in R^{m\times n}$ by solving the following problem:

$$\left\langle \hat{D}, \hat{\alpha} \right\rangle = \arg\min_{D,\alpha} \|\mathrm{P_L} - D\alpha\|_2^2 + \beta\|\alpha\|_1 \tag{1}$$

where the first term is the reconstructive term and the second term adds the sparsity constraint over the coding coefficients α, forcing that each training patch in P_L is represented by a linear combination of a few atoms in D. This means that all the training patches in P_L can be reconstructed using the learnt dictionary $\hat{D}$. Equation (1) can be solved by using the K-SVD algorithm (Aharon et al., 2006) or via the online dictionary learning algorithm (Mairal et al., 2009). However, the learnt dictionary only has reconstructive power, lacking of discriminative ability for our segmentation task. Since we know the segmentation labels of the training patches in P_L, we can use this prior information to learn a classifier that predicts labels for the target patch p_t. As in (Tong et al., 2013), a linear classifier $f(\alpha, W) = W\alpha$ is added to the objective function:

$$\left\langle \hat{D}, \hat{W}, \hat{\alpha} \right\rangle = \arg\min_{D,W,\alpha} \|\mathrm{P_L} - D\alpha\|_2^2 + \beta_1 \|H - W\alpha\|_2^2 + \beta_2\|\alpha\|_1 \tag{2}$$

where a labeling error term $\|H - W\alpha\|_2^2$ is added to Equation (1). Each column of H is a label vector corresponding to a training patch in P_L. Each label vector is defined as $h_i = [0, 0 \ldots 1 \ldots 0, 0]$, where the

non-zero entry position indicates the label of the center voxel in the corresponding training patch p_i. W denotes the linear classifier parameters and β_1 controls the trade-off between the reconstruction error term and the labeling error term. Here, we use the online dictionary learning algorithm as in (Tong et al., 2013) to solve Equation (2). After this equation is solved, a dictionary $\hat{D}_t$ and a classifier $\hat{W}$ are learnt from P_L and their labels H. The target patch p_t can be represented by the learnt dictionary $\hat{D}$ as:

$$\hat{\alpha}_t = \arg\min_{\alpha_t} \left\| p_t - \hat{D}_t \alpha_t \right\|_2^2 + \beta_2 \|\alpha_t\|_1 \tag{3}$$

where $\hat{\alpha}_t$ are the coding coefficients of the target patch p_t. Probabilistic labels of the target voxel v_t can then be estimated by the linear predictive classifier $\hat{W}$ and the coding coefficients $\hat{\alpha}_t$:

$$h_t = \hat{\mathrm{W}} \hat{\alpha}_t \tag{4}$$

Here h_t is the estimated probabilistic label vector for the target voxel v_t. Values in h_t represent the probability of the target voxel v_t belonging to different organs and are normalized to sum to one. Ideally, h_t will be $\{0, 0, \cdots, 1, \cdots, 0, 0\}$ with only one non-zero element, indicating the label of the structure. The final label at each voxel v_t can directly be determined by finding the index of the largest element in the probabilistic label vector h_t.

2.5. Speedup with Multi-resolution Framework

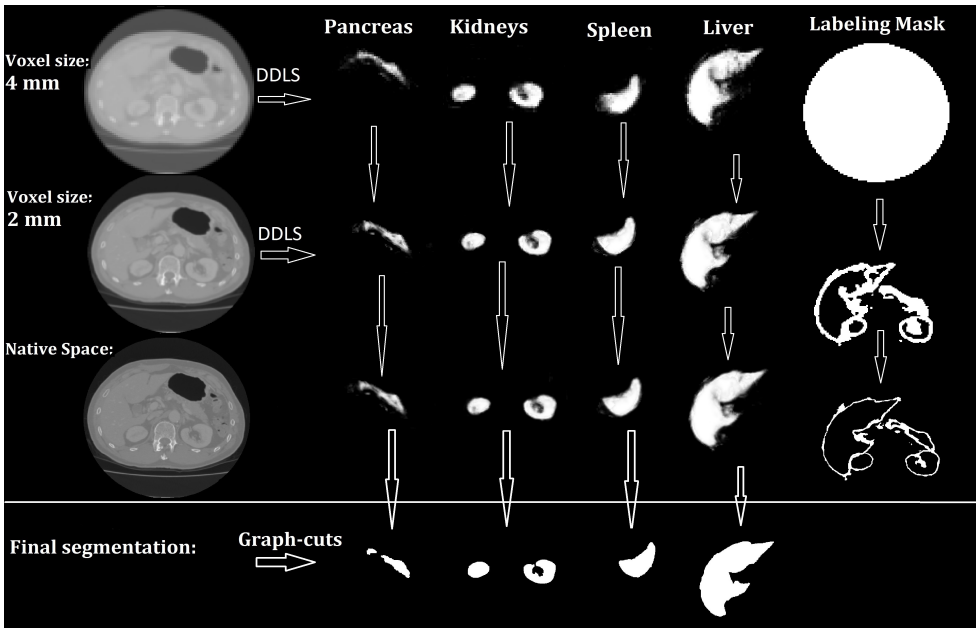

Figure 3: The multiresolution segmentation process. DDLS is performed to generate probabilistic atlas for each organ, which propagates across resolutions. The final segmentation is achieved by using the graph-cuts algorithm in the native space.

Previous patch-based approaches including our DDLS algorithm were evaluated over the segmentation of small structures like the hippocampus (Coupé et al., 2011; Tong et al., 2013), which can be computed efficiently. However, when these methods are applied to the segmentation of large structures such as the whole brain or the abdominal organs, the computational complexity becomes extremely high. For example, it takes more than 42 hours for a whole brain segmentation by using the nonlocal patch based segmentation as reported in Eskildsen et al. (2011). To overcome this problem, a multi-resolution framework (Eskildsen et al., 2011; Wang et al., 2013) can be used to gain computational efficiency. In order to make efficient

multi-organ segmentation possible, we also integrated a multi-resolution dictionary learning framework into our proposed DDLS algorithm as shown in Figure 3.

Multiple resolutions of the target image and all atlases are created by constructing Gaussian image pyramids offline. Using DDLS, a probabilistic atlas is obtained for the target image at the lowest resolution, which contains the initial probabilities of each voxel belonging to different organs. If the largest probability value at a location is lower than a defined confidence level γ, the probabilities at this location will be re-calculated at next resolution; otherwise, the probabilities at current location will be retained. This enables propagation of probabilistic atlas across resolutions by using the resulting probabilistic atlas at the current resolution to initialize the probabilistic atlas at next resolution. In this manner, the segmentation mask at next resolution is limited to the voxels with uncertain segmentations at the current resolution, forming a computationally-efficient way to process images through increasing resolutions.

2.6. Refinement with Graph Cuts

The above DDLS algorithm generates a probabilistic segmentation that serves as a subject-specific probabilistic atlas. This in turn, provides the spatial prior for obtaining the final segmentation. Previous studies (van der Lijn et al., 2008; Wolz et al., 2009, 2013) demonstrate that further improvements can be achieved by combining the target intensity information and the spatial prior. In the work of (Wolz et al., 2009), the graph-cuts algorithm is used to obtain the final segmentation S_t of the target image I_t by solving an MRF-based energy function:

$$E(S_t) = \lambda \sum_{v_i \in I_t} D_{v_i}(S_t(v_i)) + \sum_{\{v_i, v_j\} \in N} E_{v_i, v_j}(S_t(v_i), S_t(v_j)) \tag{5}$$

where v_i and v_j are voxels in a neighborhood N in the target image I_t. The data term D_{v_i} measures the disagreement between a prior probabilistic model and the observed data, which is a combination of the target intensity information and the spatial prior. E_{v_i, v_j} is a smoothness term penalizing discontinuities in a neighborhood N. A more detailed description of the energy function is given in Appendix A. The parameter λ controls the influence of these two terms, which was set to 1 in all experiments as in Wolz et al. (2013). The setting of λ was not optimized for the current dataset. Since the graph-cuts algorithm is applied to each organ independently, a fusion step is applied to obtain the final segmentation. In this step, equivocal voxels are assigned the label that has the largest value in the probabilistic label vector h_t.

3. Experiments and Results

The proposed methods were evaluated on 150 abdominal CT scans as described in Section 2.1. For the G-DDLS and L-DDLS methods, a leave-one-out procedure was utilized in the validation. Each scan was segmented by treating the remaining 149 subjects as atlases. Atlas selection was performed over the remaining 149 atlas database. Two resolution levels with isotropic voxel spacing respectively of 4 mm and 2 mm were utilized to speed up the process of the proposed methods as shown in Figure 3. After the probabilistic atlases were generated in the native spaces, they were treated as the input of the graph cuts algorithm to achieve the final segmentations. All parameters were empirically set (see Table 1) according to previous studies (Eskildsen et al., 2011; Tong et al., 2013). The influence of the mask size in the local atlas selection on the segmentation performance was evaluated in Section 3.3.

The Dice overlap was calculated between automated and manual segmentations for the evaluation of our proposed method. Paired (for the same group) or non-paired (for different groups) two-tailed t-tests were performed with the Dice overlaps to assess the statistical significance of different results. In order to compare with state-of-the-art methods, the Jaccard index (JI) as well as the Dice overlap were computed. Given the true positive (TP), false positive (FP) as well as false negative (FN) fraction, these two measures are defined as:

Resolution (mm^3)	Patch size (voxels)	Search volume (voxels)	Dictionary atoms	β_1	β_2	γ
$4 \times 4 \times 4$	$5 \times 5 \times 5$	$5 \times 5 \times 5$	256	1	0.15	0.9
$2 \times 2 \times 2$	$5 \times 5 \times 5$	$11 \times 11 \times 11$	256	1	0.15	-

Table 1: Parameter settings on different resolutions. Two resolution levels ($4 \times 4 \times 4mm^3$ and $2 \times 2 \times 2mm^3$) are used. The patch size is set to $5 \times 5 \times 5$ voxels at different resolutions and for all experiments. The search volume is the defined neighborhood in the atlases for extracting training patches. The number of atoms in dictionaries is set to 256. β_1 and β_2 are parameters in the dictionary learning and sparse coding step. γ is the defined confidence level for the propagation of probabilistic atlas across resolutions in the multi-resolution framework.

$$JI = \frac{TP}{TP + FP + FN}$$
$$Dice = \frac{2TP}{2TP + FP + FN} \tag{6}$$

3.1. Advantage of discriminative dictionary learning

We first evaluated the segmentation performance of G-DDLS compared with majority voting (MV) labeling (Heckemann et al., 2006). Global atlas selection was performed by comparing pairwise similarities between atlases and the target images over the whole CT scan. After 20 similar atlases were selected globally, the G-DDLS and MV approaches were used to perform the labeling of the target images. Furthermore, the graph cuts algorithm was utilized as a post-processing step of the G-DDLS and MV approaches, denoted as G-DDLS-GC and MV-GC respectively. Figure 4 shows a comparison of these four different methods. Since only affine registrations were used, the MV method cannot provide accurate segmentation results. Especially for the pancreas, the segmentation results are quite poor due to the significant registration errors resulting from the large variation in the shapes and locations of the pancreas. In comparison with MV, G-DDLS can generate more accurate results even though only affine registrations were used. By applying graph cuts as a post-processing step, both the G-DDLS and MV approaches gain further improvements. Therefore, we utilized the graph cuts refinement in all the following experiments.

The segmentation performance of G-DDLS is also compared with the non-local patch-based segmentation (PBS) method as proposed in Coupé et al. (2011). The results are shown in Table 2. As can be seen from Table 2, G-DDLS can achieve significant improvements over MV and PBS on all the four organs. The great variability of abdominal organs result in large registration errors, which may degrade the segmentation performance of the PBS method.

3.2. Advantage of local atlas selection

The proposed method was also validated with different atlas selection strategies. Figure 5 compares the segmentation performances of G-DDLS and L-DDLS. It can be seen that L-DDLS can achieve more accurate segmentation results than G-DDLS on the four structures when the same number of atlases are selected. Especially in the case of a small number of selected atlases (i.e. 5), the improvements of L-DDLS over G-DDLS is significant. It has been reported in Aljabar et al. (2009) that the segmentation accuracy of multi-atlas methods in terms of Dice overlap rises from a low value to a maximum and then gradually declines as the number of selected atlases increases. This is due to the fact that the population represented by a large atlas database is heterogeneous, for example in terms of age, morphology or pathology (Aljabar et al., 2009). Our proposed DDLS method follows this trend, but the segmentation accuracy of L-DDLS converges to the maximum much more quickly than that of G-DDLS as suggested by the results in Figure 5. This is attractive because it is possible to achieve the best segmentation performance of the proposed DDLS method by using only a small number of atlases.

Table 3 shows the segmentation results using G-DDLS, L-DDLS and F-DDLS over the four organs. In order to perform the F-DDLS method, 150 images were affinely transformed to a template space. Here, we chose the first image in our dataset as the template image. After that, 50 subjects were randomly

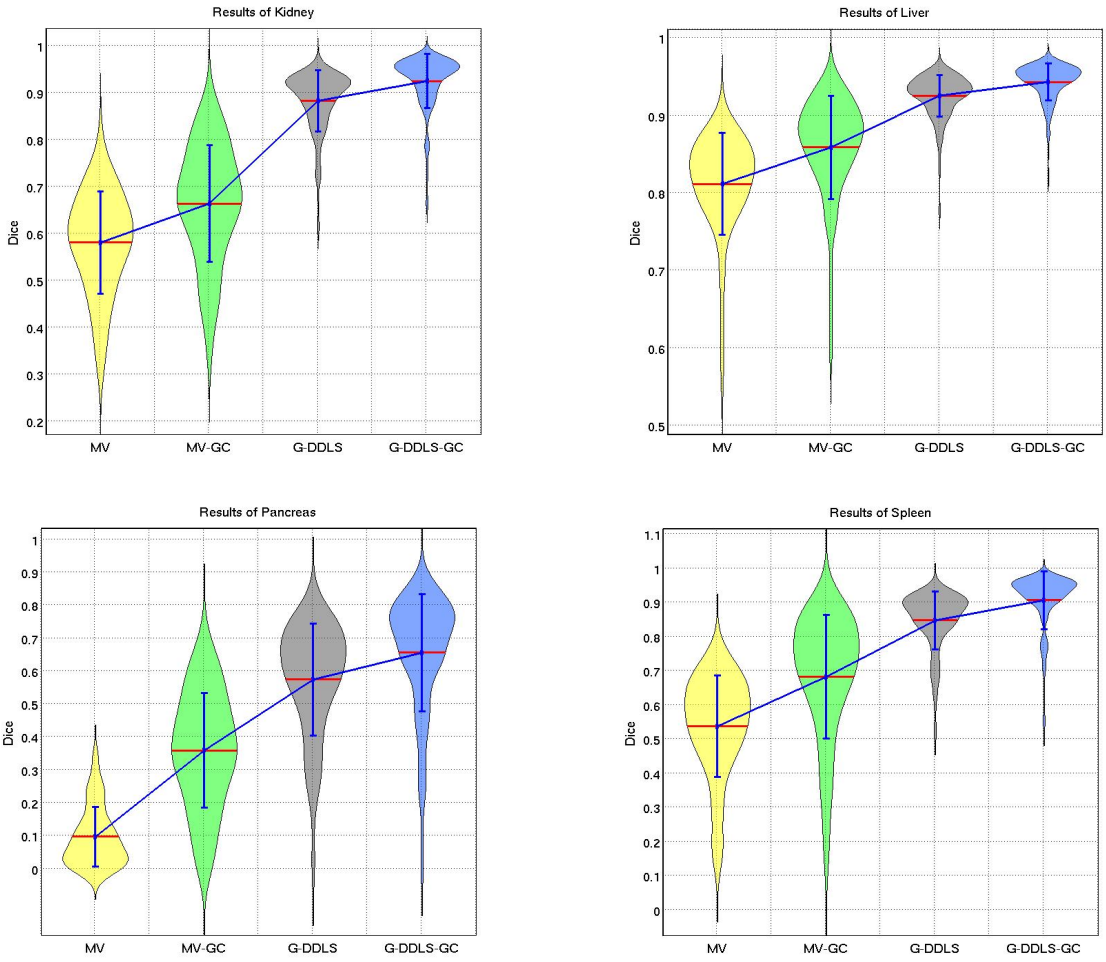

Figure 4: Comparison of different approaches. The global atlas selection strategy was utilized and 20 atlases were selected for the segmentation of each target image.

selected as training atlases. Dictionaries and classifiers were then trained offline in the template space using the randomly selected 50 atlases. The segmentations of the remaining 100 images were carried out in the template space by using the learnt dictionaries and classifiers. Finally, the segmentation results were transformed back to the target spaces for calculating the Dice overlaps. This evaluation was repeated 10 times and the average Dice overlaps were calculated. As shown in Table 3, F-DDLS achieved the lowest Dice overlaps among the three different methods because F-DDLS does not utilize an atlas selection step but learns an average model from the randomly selected subset of the database. In contrast with F-DDLS, G-DDLS and L-DDLS select similar atlases for each target image and generate a subject-specific probabilistic atlas for segmentation, which results in a significant improvement in the segmentation accuracy. In terms of Dice overlap, L-DDLS has an improvement of 3% over that of G-DDLS in the segmentation of the pancreas. However, the improvement on the segmentations of the liver is limited. This is due to the fact that both G-DDLS and L-DDLS can select similar atlases in the liver region since the liver is the largest organ, but only L-DDLS can select similar atlases in the pancreas region. It is observed that there is significant variation in the shapes and locations of the pancreas. The improvement of L-DDLS over G-DDLS in the segmentation of the pancreas suggests that the local atlas selection strategy can handle this high inter-subject variability to some extent. Figure A.7 shows exemplar segmentations for the four organ of a subject by using different

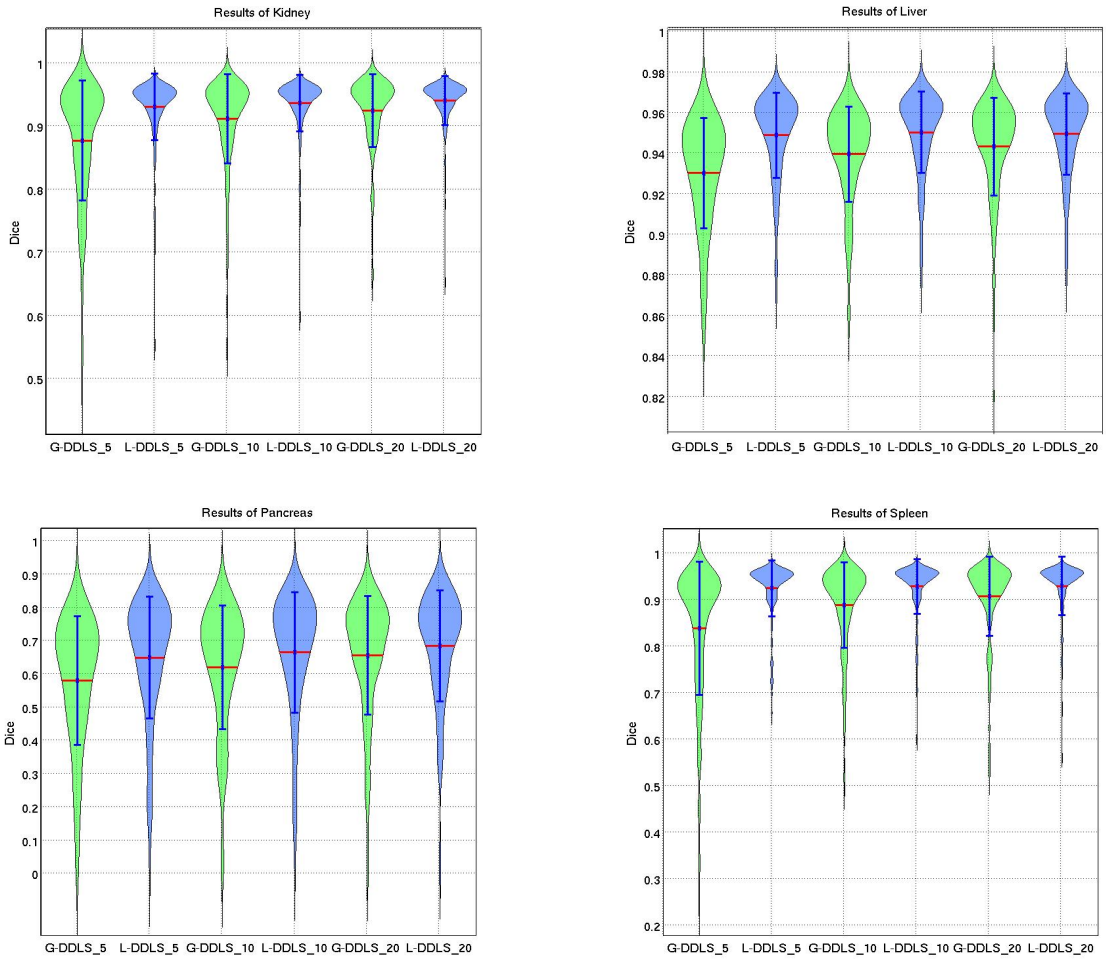

Figure 5: Comparison of G-DDLS and L-DDLS on the segmentation of four structures using different numbers of selected atlases. L-DDLS_5 means that the local atlas selection strategy is used in DDLS and 5 similar atlases are selected for labeling one target voxel. The mask size was set to $11 \times 11 \times 11$ voxels in L-DDLS at all resolutions. The other parameters in G-DDLS and L-DDLS were set as shown in Table 1.

atlas selection strategies.

3.3. Influence of mask size in L-DDLS

In L-DDLS, a local mask is defined at every voxel in the target image for selecting similar atlases at different locations adaptively. The influence of the mask size on the segmentation accuracy is shown in Figure 6. The G-DDLS is an extreme case of L-DDLS by increasing the mask size to the image size. Due to the computational burden of the DDLS method, 5 atlases were selected in this evaluation. As the mask size increases from $7 \times 7 \times 7$ voxels to $31 \times 31 \times 31$ voxels, the segmentation accuracy of the liver remains roughly unchanged, but that of the pancreas gradually drops, indicating that the local atlas selection strategy has more influence in the segmentations of small organs with large inter-subject variability.

3.4. L-DDLS with different similarity measures

The L-DDLS method was also evaluated using different similarity measures. Squared intensity differences (SSD), cross correlation (CC) and normalized mutual information (NMI) were used as similarity measures in

Methods	Liver	Kidneys	Pancreas	Spleen
MV	$85.8 \pm 6.7^{\dagger}$	$66.4 \pm 12.4^{\dagger}$	$35.9 \pm 17.4^{\dagger}$	$68.3 \pm 18.1^{\dagger}$
PBS	$91.4 \pm 7.1^{\dagger}$	$72.7 \pm 10.9^{\dagger}$	$48.7 \pm 15.7^{\dagger}$	$75.8 \pm 13.9^{\dagger}$
Proposed G-DDLS	94.3 ± 2.4	92.4 ± 5.8	65.5 ± 17.8	90.6 ± 8.5

Table 2: Comparison of Dice overlaps (MEAN ± STD (%)) using different segmentation approaches. MV and PBS represent majority voting (MV) labeling method (Heckemann et al., 2006) and non-local patch-based segmentation (PBS) method (Coupé et al., 2011) respectively. All the results were obtained by selecting 20 similar atlases in a leave-one-out procedure. Graph cuts refinement was applied in all the evaluations. The other parameters of PBS and G-DDLS were set as shown in Table 1. † means statistically significant different from the results of G-DDLS with $p < 0.0001$.

Methods	Liver	Kidneys	Pancreas	Spleen
F-DDLS	$93.1 \pm 5.2(0.0034)$	$89.6 \pm 10.8^{\dagger}$	$63.1 \pm 23.3^{\dagger}$	$89.7 \pm 11.5(0.0014)$
G-DDLS	$94.3 \pm 2.4(0.0831)$	$92.4 \pm 5.8(0.0249)$	$65.5 \pm 17.8(0.0192)$	$90.6 \pm 8.5(0.0107)$
L-DDLS	94.9 ± 2.1	93.8 ± 4.3	68.9 ± 15.8	92.8 ± 6.0

Table 3: Comparison of Dice overlaps (MEAN ± STD (%) (p value)) using different atlas selection strategies. The results of F-DDLS were obtained by randomly selecting 50 atlases for training and the remaining 100 subjects for testing, which was repeated 10 times. The other results were obtained by selecting the 20 similar atlases in a leave-one-out procedure. The mask size was set to $11 \times 11 \times 11$ voxels in L-DDLS at all resolutions. The other parameters in F-DDLS, G-DDLS, L-DDLS were set as shown in Table 1. † means statistically significant different from the results of L-DDLS with $p < 0.0001$.

the atlas selection step. Table 4 shows the results of the L-DDLS method with different similarity measures. The results using SSD, CC, and NMI are not significant different from each other on the segmentation of the liver, the spleen and the kidneys in a paired t-test. However, L-DDLS using CC and NMI as similarity measures can generate more accurate results on the segmentation of the pancreas than L-DDLS using SSD.

An experiment was also performed in order to assess the performance on lower quality image data. The dataset were downsampled in dorsoventral direction (slice-spacings were set to 5 mm) while in-plane voxel spacings were kept, simulating a typical low-resolution clinical protocol. The proprosed L-DDLS method was then validated on this downsampled dataset. Results are not significantly different from those on the high resolution dataset over the segmentation of the liver, the spleen and the kidneys in a paired t-test, except the pancreas. Since the pancreas is the smallest organ with high shape variability, the interpolation artefacts during downsampling may have more effect on it than other organs.

3.5. Comparison with state-of-the-art methods

It is always difficult to directly compare the segmentation performance with those of the state-of-the-art methods (Heimann et al., 2009; Shimizu et al., 2010; Chen et al., 2012; Linguraru et al., 2012; Bagci et al., 2012; Okada et al., 2013) due to different datasets for evaluation, different qualities of manual segmentations, and differences in the evaluation metrics used. Here, the results of three state-of-the-art methods (Chu et al., 2013; Wolz et al., 2013; Wang et al., 2014) which utilized the same dataset (Wolz et al., 2013) or a subset of our dataset (Chu et al., 2013; Wang et al., 2014) for evaluation and also the results of four other methods that used different datasets are shown in Table 5 for comparison. Table 5 shows that our proposed method

Similarity Measures	Liver	Kidneys	Right Kidney	Left Kidney	Pancreas	Spleen
SSD	94.9 ± 2.1	93.8 ± 4.3	93.4 ± 8.8	92.9 ± 9.6	68.9 ± 15.8	92.8 ± 6.0
NMI	95.0 ± 1.4	93.4 ± 4.2	93.1 ± 8.8	92.6 ± 9.1	70.7 ± 14.4	92.6 ± 6.5
CC	94.9 ± 1.9	93.6 ± 3.8	93.1 ± 8.7	93.0 ± 8.8	71.4 ± 14.7	92.5 ± 6.5
CC on Downsampled Dataset	95.1 ± 1.9	93.8 ± 3.8	93.2 ± 8.8	93.3 ± 8.8	67.1 ± 17.0	92.7 ± 6.1

Table 4: Influence of different similarity measures on the segmentation accuracy of L-DDLS. All the results were obtained by selecting 20 similar atlases in a leave-one-out procedure. The mask size was set to $11 \times 11 \times 11$ voxels in L-DDLS at all resolutions. The other parameters were set as shown in Table 1. It should be mentioned that the overall Dice of kidneys is not an average of the Dice of the left kidney and the Dice of the right kidney. All the dice values were calculated according to Equation 6.

Methods	Registration	Subjects	Organs	Dice Overlap (%)	Jaccard Index (%)	Computational time (hours)
Proposed L-DDLS_5	Global Affine	150	Liver	94.9 ± 1.8	90.4 ± 3.2	0.5
			Kidneys	93.4 ± 3.8	87.9 ± 6.1	
			Pancreas	69.8 ± 14.5	55.3 ± 14.8	
			Spleen	91.9 ± 6.7	85.7 ± 9.8	
Proposed L-DDLS_20	Global Affine	150	Liver	94.9 ± 1.9	90.1 ± 3.3	2
			Kidneys	93.6 ± 3.8	88.3 ± 6.1	
			Pancreas	71.1 ± 14.7	56.9 ± 15.2	
			Spleen	92.5 ± 6.5	86.7 ± 9.7	
Wang et al. (2014)	Gobal Affine	100	Liver	94.5 ± 2.5	-	14
			Kidneys	92.4 ± 7.7	-	
			Pancreas	65.5 ± 18.6	-	
			Spleen	92.5 ± 8.4	-	
Wolz et al. (2013)	Local non-rigid	150	Liver	94.0 ± 2.8	88.9 ± 4.8	51
			Kidneys	92.5 ± 7.2	86.8 ± 10.5	
			Pancreas	69.6 ± 16.7	55.5 ± 17.1	
			Spleen	92.0 ± 9.2	86.2 ± 12.7	
Chu et al. (2013)	Local non-rigid	100	Liver	95.1 ± 1.0	90.6	-
			Kidneys	90.1 ± 5.0	82.3	
			Pancreas	69.1 ± 15.3	54.6	
			Spleen	91.4 ± 5.7	84.5	
Shimizu et al. (2007)	Affine	28	Liver	-	89.0	-
			Kidneys	-	88.2	
			Pancreas	-	46.6	
			Spleen	-	82.5	
Okada et al. (2012)	Non-rigid	10	Liver	-	89.1	-
			Kidneys	-	82.5	
			Pancreas	-	35.0	
			Spleen	-	83.5	
Bagci et al. (2012)	Affine	20	Liver	92.2 ± 1.0	-	-
			Kidneys	93.4 ± 1.0	-	
			Spleen	93.5 ± 1.3	-	
Linguraru et al. (2012)	Non-rigid	20	Liver	94.0 ± 1.2	-	3
			Kidneys	92.6 ± 2.4	-	
			Spleen	89.6 ± 2.7	-	

Table 5: Comparison with state-of-the-art methods (Top group: the proposed L-DDLS method with different number of selected atlases; Middle group: methods using the same dataset; Bottom group: methods using other dataset). The results of L-DDLS were obtained in a leave-one-out procedure and cross correlation was used as the similarity measure for local atlas selection. L-DDLS_5 and L-DDLS_20 represent that 5 and 20 atlases were selected in L-DDLS respectively. The computational time is the runtime of the segmentation of one target image without parallelization (single core).

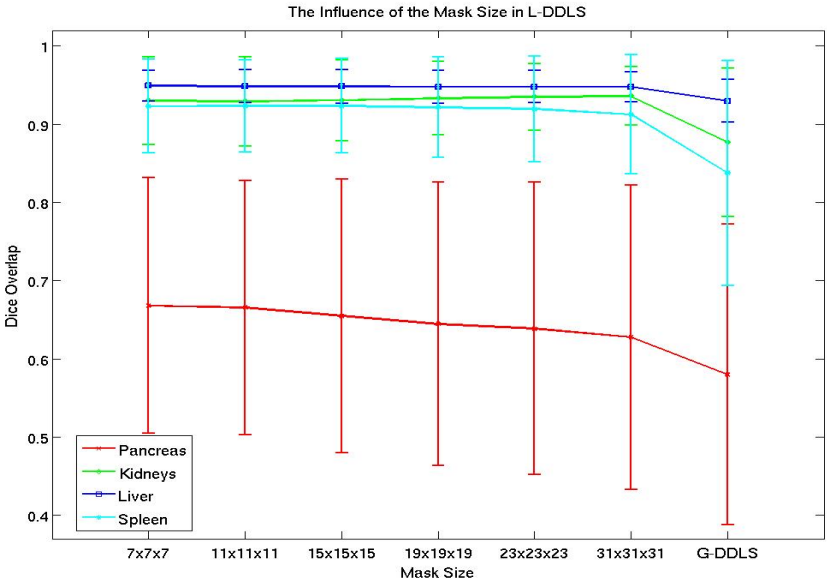

Figure 6: Influence of the mask size on the segmentation accuracy of L-DDLS. The results were obtained by selecting the 5 most similar atlases in a leave-one-out procedure. G-DDLS is an extreme case of L-DDLS by increasing the mask size to the image size.

achieves competitive performance with these state-of-the-art methods. In addition, the proposed L-DDLS method can be implemented very efficiently as shown in Table 5, which can be attractive in clinical practice.

3.6. Computational time

The runtimes of our proposed G-DDLS and L-DDLS methods increase approximately linearly with the number of the selected atlases during training. In our implementation, all the experiments were carried out with eight Intel Xeon cores clocked at 3 GHz and 32 GB RAM. It takes around half an hour to segment the four organs of an abdominal scan when 5 atlases are selected for training dictionaries. However, if the number of selected atlases increases to 20, the runtime increases to around 2.5 hours. For G-DDLS, the number of selected atlases yields significant differences in the segmentation accuracy as the Dice overlap values increase significantly from selecting 5 atlases to 20 atlases as shown in Figure 5. However, L-DDLS does not have this problem as its segmentation accuracy reaches the maximum much earlier than that of G-DDLS. Therefore, it takes much more time for G-DDLS to achieve the best segmentation results, as more atlases are needed compared with L-DDLS. Using F-DDLS, segmentations can be performed quite efficiently. It takes approximately 15 minutes per scan using F-DDLS since the dictionaries and classifiers have been trained offline and only the sparse coding step is needed in the segmentation stage, which can speedup the process significantly.

4. Discussion and Conclusion

In this paper, we developed discriminative dictionary learning techniques for the multi-organ abdominal segmentation in CT images. A large dataset of 150 abdominal CT images was used for evaluation. Experimental results show that the proposed DDLS method achieves significantly more accurate results than the traditional multi-atlas segmentation method based on MV label fusion (Heckemann et al., 2006) and the nonlocal patch based segmentation method (Coupé et al., 2011). It provides a comparable segmentation accuracy to those of the state-of-the-art methods (Okada et al., 2012; Linguraru et al., 2012; Bagci et al., 2012; Chu et al., 2013; Wolz et al., 2013; Wang et al., 2014). In addition, our proposed DDLS method

achieves promising segmentation results by only using global affine registration. Since only affine registration is required, our method can be implemented efficiently, demonstrating the potential for real-time clinical applications and in challenging datasets where accurate registration is difficult to achieve.

Different atlas selection strategies were implemented and compared with the DDLS method. Among them, the F-DDLS method employs an average model as in approaches based on statistical shape models. In statistical shape models, ideal mean shapes of different organs are constructed from a specific dataset. In F-DDLS, fixed dictionaries and classifiers are learnt from a given subset (i.e. randomly selected 50 atlases). The advantage of F-DDLS is that the segmentation can be performed quite efficiently since the average model (fixed dictionaries and classifiers in F-DDLS) has been learnt offline. However, approaches based on the average model from a specific dataset may be challenged by diverse testing datasets, where high inter-subject variability exists. In comparison with F-DDLS, the G-DDLS and L-DDLS methods automatically select suitable atlases for an unlabeled target image and then learn target specific priors for segmentation. This can result in significant improvement in the segmentation performance, especially for the segmentation of the pancreas as shown in Table 3.

The L-DDLS method takes full avantage of the whole dataset and adapts to each location in the target image individually. The atlases most suitable to the current location under consideration are automatically selected. Atlases that have different local anatomical patterns at the current location are not taken into account, but still available for other locations in the target image. In comparison with G-DDLS, there are three advantages of L-DDLS: (1) One can achieve promising segmentation results with fewer atlases by using local atlas selection strategy in comparison with using normal global atlas selection. For example, the L-DDLS method can segment the liver, kidneys, pancreas, and spleen with Dice overlap values of 94.8%, 92.9%, 66.6%, and 92.4% respectively by selecting 5 atlases locally. Although only 5 atlases are selected, the most similar atlases have already been found at each location by using L-DDLS, which can then provide reliable prior information for label estimation. In comparison, the Dice overlap values of G-DDLS using 20 atlases (as shown in Table 3) are still lower than those of L-DDLS with 5 atlases. (2) Since less training atlases are needed for labeling a target image in L-DDLS, the computational burden can also be significantly reduced. The runtime of DDLS is around 30 minutes by selecting 5 atlases, while this increases to approximately 2.5 hours by using 20 atlases. (3) L-DDLS can handle the high inter-subject variability of small organs like the pancreas much better than G-DDLS. This is due to the fact that G-DDLS selects atlases according to global similarity between atlases and the target image. This global similarity, however, is dominated by the similarity in large structures like the liver, weakening the influence of the similarity in small organs like the pancreas. By treating the similarity at each location equally, L-DDLS achieves an improvement of 3% in terms of Dice overlap over that of G-DDLS in the segmentation of the pancreas, which is the most challenging structure.

The number of selected atlases K is an important parameter in multi-atlas segmentation methods. In our work, K was predefined globally, which means that the same number of atlases are selected at each location in the target image. However, it is observed (Aljabar et al., 2009) that K required for the highest segmentation accuracy varies for different structures. This could also be the case for different locations. A further improvement may be obtained by not only selecting similar atlases locally but also choosing the best number of atlases adaptively at each location. This can be done by modeling the segmentation errors as a function of K as proposed in (Awate and Whitaker, 2014). After the function is fitted, the best number of atlases can be estimated at each location. However, it should be mentioned that the process of estimating the best K at each location may increase the computational complexity of our proposed method.

In terms of computational time, patch based segmentation methods (Coupé et al., 2011) can gain some computational efficiency by avoiding the need for non-rigid registration. However, they still suffer from the high computational burden in the label fusion stage (Eskildsen et al., 2011; Wang et al., 2014), which becomes a significant problem for the large abdominal organs in high resolution images. This is why the multi-resolution framework was combined in our work to speed up the segmentation process. A very recent patch-based segmentation method using the patch match algorithm (Ta et al., 2014) allows speed ups of around 200 to 1000 fold in the label fusion stage without losing segmentation accuracy, providing a new potential way to gain further computational efficiency in our work. Overall, a segmentation method providing a high accuracy that can be implemented efficiently will be preferable.

Although the proposed method works well on the segmentation of abdominal organs in CT scans, it has several drawbacks. First, the proposed method still requires alignment between atlas images and the target image with a global affine registration. This process can still be a problematic step in images with a high degree of anatomical variance (Wang et al., 2013). Another direction for future work will be to investigate the extension of our proposed method without registration. Second, atlas selection is an essential step in the proposed method for achieving good segmentation performance. A subset of similar atlases are selected globally or locally from all the training atlases for the segmentation of each target image. However, the remaining "dissimilar" atlases could potentially provide valuable information to aid the segmentation. For example, similar patches could still be present in dissimilar atlases, which can provide additional information for labeling the target patches. In future work, the potential to perform segmentation without atlas selection will be investigated in order to take full advantage of the whole atlas dataset. Furthermore, the proposed method uses local patches for segmentation, which can only provide local intensity patterns, but neglects the global anatomical patterns. The global anatomical information, however, can be helpful for the segmentation work. For instance, the inter-organ relations has been demonstrated to be helpful for segmentation as shown in (Okada et al., 2013; Cerrolaza et al., 2014; Wang et al., 2014), which can also be integrated into the proposed method for a further improvement.

Appendix A. Energy terms of Graph Cuts

The data term $D(S_t)$ in Equation (5) is estimated from a spatial prior and a probabilistic model of the intensity in the target image. It is formulated as:

$$\begin{aligned} D(S_t) &= \alpha E_{intensity}\left(S_t\right) + (1-\alpha)\, E_{prior}\left(S_t\right) \\ &= -\alpha \ln P\left(I|S_t\right) - (1-\alpha) \ln P_{sprior}\left(S_t\right) \end{aligned} \tag{A.1}$$

Here P_{sprior} is the spatial prior, which is obtained using our DDLS method as described in section 2.4. $P\left(I|S_t\right)$ is the image likelihood and usually modeled by a Gaussian probability distribution. Specifically, the distribution of foreground intensities of a particular structure is modeled using a single Gaussian distribution, while the background distribution for a particular structure is estimated using a mixture of Gaussians (MOG) model. The details of modeling the intensity prior using Gaussian distributions can be found in (Wolz et al., 2009). The parameter α was set to 0.1 in all experiment as in (Wolz et al., 2013).

The second term in Equation (5) is a smooth term used to define the weights of edges connecting two neighboring voxels v_i and v_j, which is given by:

$$\begin{aligned} E_{v_i,v_j}(S_t(v_i), S_t(v_j)) &= c\left(1 + \ln\left(1 + \frac{1}{2}\left(\frac{|I(v_i) - I(v_j)|}{\sigma}\right)^2\right)\right)^{-1} \\ &+ (1-c)\left(1 - \max_{x \in M_{v_i,v_j}}(B_x)\right) \end{aligned} \tag{A.2}$$

where B is the intervening contour probabilistic map derived from the gradient image, M_{v_i,v_j} is a line joining v_i and v_j, and σ is the robust scale of image (Wolz et al., 2009). The parameter c is empirically set to 0.5 as in Wolz et al. (2009).

References

Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing 54 (11), 4311–4322.

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 46 (3), 726–738.

Awate, S., Whitaker, R., 2014. Multiatlas segmentation as nonparametric regression. IEEE Transactions on Medical Imaging PP (99), 1–14.

Bagci, U., Chen, X., Udupa, J. K., 2012. Hierarchical scale-based multiobject recognition of 3D anatomical structures. IEEE Transactions on Medical Imaging 31 (3), 777–789.

Cao, Y., Yuan, Y., Li, X., Turkbey, B., Choyke, P. L., Yan, P., 2011. Segmenting images by combining selected atlases on manifold. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011. Springer, pp. 272–279.

Cerrolaza, J. J., Villanueva, A., Reyes, M., Cabeza, R., Ballester, M. A. G., Linguraru, M. G., 2014. Generalized multiresolution hierarchical shape models via automatic landmark clusterization. In: Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 1–8.

Chen, X., Udupa, J. K., Bagci, U., Zhuge, Y., Yao, J., 2012. Medical image segmentation by combining graph cuts and oriented active appearance models. IEEE Transactions on Image Processing 21 (4), 2035–2046.

Chu, C., Oda, M., Kitasaka, T., Misawa, K., Fujiwara, M., Hayashi, Y., Nimura, Y., Rueckert, D., 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, pp. 165–172.

Cootes, T. F., Taylor, C. J., Cooper, D. H., Graham, J., 1995. Active shape models-their training and application. Computer vision and image understanding 61 (1), 38–59.

Coupé, P., Manjķn, J., Fonov, V., Pruessner, J., Robles, M., Collins, D., 2011. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage 54 (2), 940–954.

Eskildsen, S., Coupé, P., Fonov, V., Manjón, J., Leung, K., Guizard, N., Wassef, S., Østergaard, L., Collins, D., 2011. BEaST: Brain extraction based on nonlocal segmentation technique. NeuroImage 59 (3), 2362–2373.

Gao, L., Heath, D. G., Fishman, E. K., 1998. Abdominal image segmentation using three-dimensional deformable models. Investigative Radiology 33 (6), 348–355.

Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33 (1), 115–126.

Heimann, T., Van Ginneken, B., Styner, M. A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., et al., 2009. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Transactions on Medical Imaging 28 (8), 1251–1265.

Heimann, T., Wolf, I., Meinzer, H.-P., 2006. Active shape models for a fully automated 3D segmentation of the liver–an evaluation on clinical data. In: Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 41–48.

Linguraru, M. G., Pura, J. A., Pamulapati, V., Summers, R. M., 2012. Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. Medical image analysis 16 (4), 904–914.

Linguraru, M. G., Sandberg, J. K., Li, Z., Shah, F., Summers, R. M., 2010. Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation. Medical physics 37 (2), 771–783.

Liu, F., Zhao, B., Kijewski, P., Ginsberg, M. S., Wang, L., Schwartz, L. H., 2004. Automatic liver contour segmentation using GVF snake. In: SPIE Medical Imaging. International Society for Optics and Photonics, pp. 1466–1473.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2009. Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning. ACM, pp. 689–696.

Oda, M., Nakaoka, T., Kitasaka, T., Furukawa, K., Misawa, K., Fujiwara, M., Mori, K., 2012. Organ segmentation from 3D abdominal CT images based on atlas selection and graph cut. In: Abdominal Imaging. Computational and Clinical Applications. Springer, pp. 181–188.

Okada, T., Linguraru, M. G., Hori, M., Summers, R. M., Tomiyama, N., Sato, Y., 2013. Abdominal multi-organ CT segmentation using organ correlation graph and prediction-based shape and location priors. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, pp. 275–282.

Okada, T., Linguraru, M. G., Yoshida, Y., Hori, M., Summers, R. M., Chen, Y.-W., Tomiyama, N., 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, pp. 173–180.

Okada, T., Shimada, R., Hori, M., Nakamoto, M., Chen, Y.-W., Nakamura, H., Sato, Y., 2008a. Automated segmentation of the liver from 3D CT images using probabilistic atlas and multilevel statistical shape model. Academic radiology 15 (11), 1390–1403.

Okada, T., Yokota, K., Hori, M., Nakamoto, M., Nakamura, H., Sato, Y., 2008b. Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images. In: Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 502–509.

Park, H., Bland, P. H., Meyer, C. R., 2003. Construction of an abdominal probabilistic atlas and its application in segmentation. IEEE Transactions on Medical Imaging 22 (4), 483–492.

Pluim, J. P., Maintz, J. A., Viergever, M. A., 2003. Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging 22 (8), 986–1004.

Rousseau, F., Habas, P., Studholme, C., 2011. A supervised patch-based approach for human brain labeling. IEEE Transactions on Medical Imaging 30 (10), 1852–1862.

Shi, F., Yap, P.-T., Fan, Y., Gilmore, J. H., Lin, W., Shen, D., 2010. Construction of multi-region-multi-reference atlases for neonatal brain mri segmentation. Neuroimage 51 (2), 684–693.

Shimizu, A., Kimoto, T., Kobatake, H., Nawano, S., Shinozaki, K., 2010. Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. International journal of computer assisted radiology and surgery 5 (1), 85–98.

Shimizu, A., Nakagomi, K., Narihira, T., Kobatake, H., Nawano, S., Shinozaki, K., Ishizu, K., Togashi, K., 2011. Auto-

mated segmentation of 3d CT images based on statistical atlas and graph cuts. In: Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging. Springer, pp. 214–223.

Shimizu, A., Ohno, R., Ikegami, T., Kobatake, H., Nawano, S., Smutek, D., 2007. Segmentation of multiple organs in non-contrast 3D abdominal CT images. International Journal of Computer Assisted Radiology and Surgery 2 (3-4), 135–142.

Spiegel, M., Hahn, D. A., Daum, V., Wasza, J., Hornegger, J., 2009. Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration. Computerized Medical Imaging and Graphics 33 (1), 29–39.

Ta, V.-T., Giraud, R., Coupé, P., Collins, D., 2014. Optimized patchmatch for near real time and accurate label fusion. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014. Springer, pp. 105–112.

Tong, T., Wolz, R., Coupé, P., Hajnal, J. V., Rueckert, D., 2013. Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. NeuroImage 76, 11–23.

van der Lijn, F., den Heijer, T., Breteler, M., Niessen, W. J., 2008. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. Neuroimage 43 (4), 708–720.

van Rikxoort, E. M., Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M. A., Pluim, J. P., van Ginneken, B., 2010. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. Medical Image Analysis 14 (1), 39–49.

Wang, H., Stout, D. B., Chatziioannou, A. F., 2012. Estimation of mouse organ locations through registration of a statistical mouse atlas with micro-CT images. IEEE Transactions on Medical Imaging 31 (1), 88–102.

Wang, L., Lekadir, K., Ismail, E.-H., Yacoub, M., Yang, G.-Z., 2010. Subject specific shape modeling with incremental mixture models. In: Medical Imaging and Augmented Reality. Springer, pp. 21–30.

Wang, Z., Bhatia, K. K., Glocker, B., Marvao, A., Dawes, T., Misawa, K., Mori, K., Rueckert, D., 2014. Geodesic patch-based segmentation. In: Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 666–673.

Wang, Z., Donoghue, C., Rueckert, D., 2013. Patch-based segmentation without registration: Application to knee MRI. In: Machine Learning in Medical Imaging. Springer, pp. 98–105.

Wimmer, A., Soza, G., Hornegger, J., 2009. A generic probabilistic active shape model for organ segmentation. In: Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 26–33.

Wolz, R., Aljabar, P., Hajnal, J., Hammers, A., Rueckert, D., et al., 2010. LEAP: learning embeddings for atlas propagation. NeuroImage 49 (2), 1316–1325.

Wolz, R., Aljabar, P., Rueckert, D., Heckemann, R. A., Hammers, A., 2009. Segmentation of subcortical structures and the hippocampus in brain MRI using graph-cuts and subject-specific a-priori information. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, pp. 470–473.

Wolz, R., Chu, C., Misawa, K., Fujiwara, M., Mori, K., Rueckert, D., 2013. Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE Transactions on Medical Imaging 32 (9), 1723–1730.

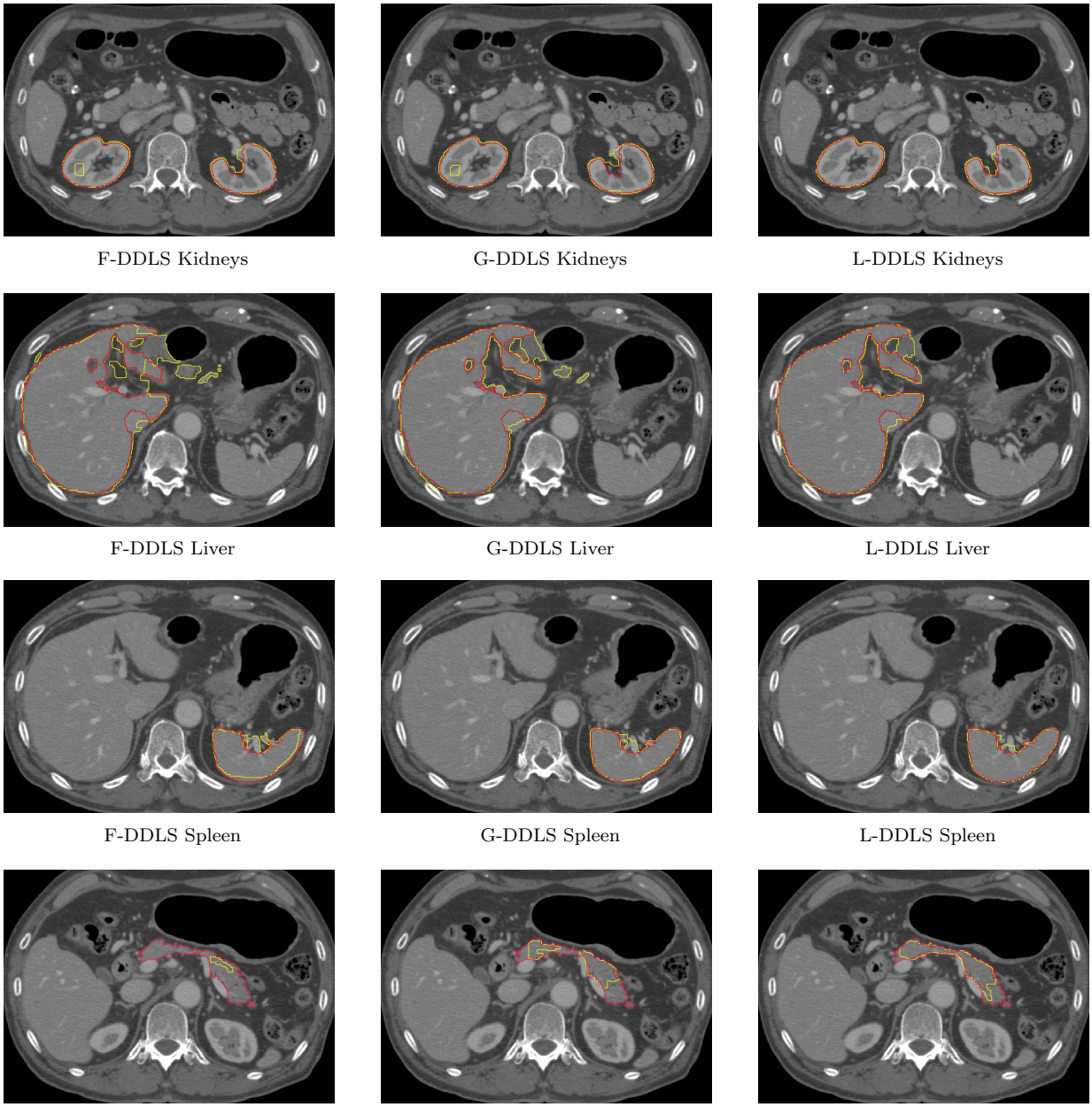

Figure A.7: Visual comparison of the segmentation results that were obtained by F-DDLS, G-DDLS and L-DDLS for liver, kidneys, pancreas and spleen of one subject. The automated segmentation is outlined in yellow and the manual segmentation is shown in red. The fourth row provides 3D renderings of different segmentation results.