HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2015 April 22.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2015 March 20; 9413: . doi:10.1117/12.2081012.

Efficient Abdominal Segmentation on Clinically Acquired CT with SIMPLE Context Learning

Zhoubing Xu[*,a], **Ryan P. Burke**[b], **Christopher P. Lee**[c], **Rebeccah B. Baucom**[d], **Benjamin K. Poulose**[d], **Richard G. Abramson**[e], and **Bennett A. Landman**[a,b,c,e]

[a] Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235

[b] Biomedical Engineering, Vanderbilt University, Nashville, TN, USA 37235

[c] Computer Science, Vanderbilt University, Nashville, TN, USA 37235

[d] General Surgery, Vanderbilt University, Nashville, TN, USA 37235

[e] Radiology and Radiological Science, Vanderbilt University, Nashville, TN, USA 37235

Abstract

Abdominal segmentation on clinically acquired computed tomography (CT) has been a challenging problem given the inter-subject variance of human abdomens and complex 3-D relationships among organs. Multi-atlas segmentation (MAS) provides a potentially robust solution by leveraging label atlases via image registration and statistical fusion. We posit that the efficiency of atlas selection requires further exploration in the context of substantial registration errors. The selective and iterative method for performance level estimation (SIMPLE) method is a MAS technique integrating atlas selection and label fusion that has proven effective for prostate radiotherapy planning. Herein, we revisit atlas selection and fusion techniques for segmenting 12 abdominal structures using clinically acquired CT. Using a re-derived SIMPLE algorithm, we show that performance on multi-organ classification can be improved by accounting for exogenous information through Bayesian priors (so called context learning). These innovations are integrated with the joint label fusion (JLF) approach to reduce the impact of correlated errors among selected atlases for each organ, and a graph cut technique is used to regularize the combined segmentation. In a study of 100 subjects, the proposed method outperformed other comparable MAS approaches, including majority vote, SIMPLE, JLF, and the Wolz locally weighted vote technique. The proposed technique provides consistent improvement over state-of-the-art approaches (median improvement of 7.0% and 16.2% in DSC over JLF and Wolz, respectively) and moves toward efficient segmentation of large-scale clinically acquired CT data for biomarker screening, surgical navigation, and data mining.

Keywords

Multi-Atlas Segmentation; SIMPLE; Context Learning; Liver; Spleen; Kidneys

[*]zhoubing.xu@vanderbilt.edu; http://masi.vuse.vanderbilt.edu; Medical-image Analysis and Statistical Interpretation Laboratory, Department of Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235.

1. INTRODUCTION

The human abdomen is an essential, yet complex body space. Computed tomography (CT) scans are routinely obtained for the diagnosis and prognosis of abdomen-related disease. Automated segmentation of abdominal anatomy may improve patient care by decreasing or eliminating the subjectivity inherent in traditional qualitative assessment. In large-scale clinical studies, efficient segmentation of multiple abdominal organs can also be used for biomarker screening, surgical navigation, and data mining.

Atlas-based segmentation provides a general-purpose approach to segment target images by transferring information from canonical atlases via registration. When adapting to abdomen, the variable abdominal anatomy between (e.g., weight, stature, age, disease status) and within individuals (e.g., pose, respiratory cycle, clothing) can lead to substantial registration errors (Figure 1). Previous abdominal segmentation approaches have used single probabilistic atlases constructed by co-registering atlases to characterize the spatial variations of abdominal organs (e.g., [1, 2]). Multi-atlas segmentation (MAS), on the other hand, is a technique that has been proven effective and robust in neuroimaging by registering multiple atlases to the target image separately, and combining voxel-wise observations among the registered labels through label fusion [3]. Recently, Wolz et al. applied MAS to the abdomen using locally weighted subject-specific atlas [4]; yet the segmentation accuracies were inconsistent. We posit that the efficiency of atlas selection for abdominal MAS requires further exploration in the context of substantial registration errors, especially on clinically acquired CT.

The selective and iterative method for performance level estimation [5] (SIMPLE) algorithm raised effective atlas selection criteria based on the Dice similarity coefficient [6] overlap with intermediate voting-based fusion result, and addressed extensive variation in prostate anatomy to reduce the impact of outlier atlases. In [7], we generalized a SIMPLE theoretical framework to account for exogenous information through Bayesian priors – referred to as context learning; the newly presented model selected atlases more effectively for segmenting spleens in metastatic liver cancer patients. A further integration with joint label fusion (JLF) [8] addressed the label determination by reducing the correlated errors among the selected atlases, and yielded a median DSC of 0.93 for spleen segmentation.

Herein, we propose an efficient approach for segmenting 12 abdominal organs of interest (Figure 1) in 75 metastatic liver cancer patients and 25 ventral hernia patients on clinically acquired CT. Based on the re-derived SIMPLE framework [7], we construct context priors, select atlases, and fuse estimated segmentation using JLF for each organ individually, and combine the fusion estimates of all organs into a regularized multi-organ segmentation using graph cut [9] (Figure 2). The segmentation performances are validated with other MAS approaches, including majority vote (MV), SIMPLE, JLF, and the Wolz approach.

2. THEORY

We re-formulate the SIMPLE algorithm from the perspective of Expectation-Maximization (EM) while focusing on the atlas selection step. In this principled likelihood model, the

Bayesian prior learning from context features (e.g., intensity, gradient) is considered as exogenous information to regularize the atlas selection.

2.1 Statistical SIMPLE Model

Consider the latent true segmentation as a vector, $\boldsymbol{T} \in \boldsymbol{L}^{N\times 1}$, where $\boldsymbol{L} = \{0, ..., L-1\}$ is the set of possible label types. Consider a collection of R registered atlases with label decisions, $\boldsymbol{D} \in \boldsymbol{L}^{N\times R}$. Let $\boldsymbol{c} \in \boldsymbol{S}^{R}$, where $\boldsymbol{S} = \{0,1\}$ indicates the atlas selection decision, i.e., 0 – ignored, and 1 – selected. Let i be the index of voxels, and j of registered atlases. We propose a non-linear rater model, $\theta \in \mathbb{R}^{R\times 2\times L\times L}$, that considers the two atlas selection decisions. Let the ignored atlases be no better than random chance, and the selected atlases be slightly inaccurate with error factors $\boldsymbol{\varepsilon} \in \mathbf{E}^{\mathrm{R}\times 1}$, where $\mathbf{E} \in \left(0, \frac{L-1}{L}\right)$. Thus

$$\theta_{j0s's} = \frac{1}{L}, \quad \forall s'; \theta_{j1s's} = \begin{cases} 1-\epsilon_j, & s' = s \\ \frac{\epsilon_j}{L-1}, & s' \neq s \end{cases} \quad (1)$$

where each element $\theta_{jns's}$ represents the probability that the registered atlas j observes label s' given the true label is s and the atlas selection decision is n with an error factor ϵ_j if selected, – i.e., $\theta_{jns's} \equiv f(D_{ij} = s' | T_i = s, c_j = n, \varepsilon_j)$.

Following [10], let $\boldsymbol{W} \in \mathbb{R}^{L\times N}$, where $W_{si}^{(k)}$ represents the probability that the true label associated with voxel i is label s at the k[th] iteration. Using Bayesian expansion and conditional inter-atlas independence, the E-step can be derived as

$$W_{si}^{(k)} = \frac{f(T_i = s)\,\Pi_j f\left(D_{ij} | T_i = s, c_j^{(k)} = n, \epsilon_j^{(k)}\right)}{\Sigma_{s'} f(T_i = s')\,\Pi_j f\left(D_{ij} | T_i = s', c_j^{(k)} = n, \epsilon_j^{(k)}\right)} \quad (2)$$

where $f(T_i = s)$ is a voxel-wise *a priori* distribution of the underlying segmentation. Note that the selected atlases contribute to $\boldsymbol{W}$ in a similar way as globally weighted vote given the symmetric form of $\theta_{j1s's}$ as in the original SIMPLE. In the M-step, the estimation of the parameters is obtained by maximizing the expected value of the conditional log likelihood function found in Eq. 2. For the error factor,

$$\begin{aligned} \epsilon_j^{(k+1)} &= arg \max_{\epsilon_j} \Sigma_i E\left[ln f\left(D_{ij} | T_i, c_j^{(k)}, \epsilon_j\right) | \boldsymbol{D}, c_j{}^{(k)}, \epsilon_j^{(k)}\right] \\ &= arg \max_{\epsilon_j} \Sigma_{s'} \Sigma_{i:D_{ij}=s'} \Sigma_s W_{si}^{(k)} ln\theta_{j_{c_j^{(k)}} s's} \equiv L_{\epsilon_j} \end{aligned} \quad (3)$$

Consider the binary segmentation for simplicity, let $M_{TP} = \Sigma_{i:D_{ij}=1} W_{1i}^{(k)}$, $M_{FP} = \Sigma_{i:D_{ij}=1} W_{0i}^{(k)}$, $M_{FN} = \Sigma_{i:D_{ij}=0} W_{1i}^{(k)}$, $M_{TN} = \Sigma_{i:D_{ij}=0} W_{0i}^{(k)}$, and $M_T = M_{TP} + M_{TN}$, $M_F = M_{FP} + M_{FN}$. After taking partial derivative of $L_{\epsilon j}$,

$$\epsilon_j^{(k+1)} = \frac{M_F}{M_T + M_F}, i.e., 1 - \epsilon_j^{(k+1)} = \frac{M_T}{M_T + M_F} \quad (4)$$

Then for the atlas selection decision

$$\begin{aligned} c_j^{(k+1)} &= arg \max_{c_j} \Sigma_i E\left[lnf\left(D_{ij} | T_i, c_j, \epsilon_j^{(k+1)} \right) | \boldsymbol{D}, c_j^{(k+1)}, \epsilon_j^{(k+1)} \right] \\ &= arg \max_{c_j} \Sigma_{s'} \Sigma_{i:D_{ij}=s'} \Sigma_s W_{si}^{(k)} ln\theta_{j_{c_j} s' s}. \end{aligned} \quad (5)$$

Given the intermediate truth estimate $W_{si}^{(k)}$, $c_j^{(k+1)}$ can be maximized by evaluating each 0/1 atlas selection separately. Note the selecting/ignoring behavior in Eq. 5 is parameterized with the error factor ε_j, and thus affected by the four summed values of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) as in Eq. 4. This differs from DSC using in the original SIMPLE in that DSC does not factor the impacts of TN. Typical practice for a fusion approach might use the prior probability, $f(T_i = s)$, to weight by expected volume of structure. With outlier atlases, one could reasonably expect a much larger region of confusion (i.e., non "consensus"[11]) than true anatomical volume. Hence, an informed prior would greatly deemphasize the TN and yield a metric similar to DSC. Therefore, we argue that SIMPLE is legitimately viewed as a statistical fusion algorithm that is approximately optimal for the non-linear rater model proposed in Eq. 1.

2.2 Context Learning

Different classes of tissues in CT images can be characterized with multi-dimensional Gaussian mixture models using intensity and spatial "context" features. On a voxel-wise basis, let $\nu \in \mathbb{R}^{d \times 1}$ represent a d dimensional feature vector, $m \in \boldsymbol{M}$ indicate the tissue membership, where $\boldsymbol{M} = \{1, ..., M\}$ is the set of possible tissues, and typically, a superset of the label types, i.e., $\boldsymbol{M} \supseteq \boldsymbol{L}$. The probability of the observed features given the tissue type is t can be represented with the mixture of N_G Gaussian distributions,

$$f(\nu | m=t) = \sum_{k=1}^{N_G} \frac{\alpha_{kt}}{(2\pi)^{\frac{d}{2}} |\boldsymbol{C}_{kt}|^{\frac{1}{2}}} exp\left[-\frac{1}{2} (\boldsymbol{v} - \mu_{kt})^T \boldsymbol{C}_{kt}^{-1} (\nu - \mu_{kt}) \right] \quad (6)$$

where $\alpha_{kt} \in \mathbb{R}^{1 \times 1}, \mu_{kt} \in \mathbb{R}^{d \times 1}, \quad and \quad \boldsymbol{C}_{kt} \in \mathbb{R}^{d \times d}$ are the unknown mixture probability, mean, and covariance matrix to estimate for each Gaussian mixture component k of each tissue type t by the EM algorithm following [12]. This context model can be trained from datasets with known tissue separations.

The tissue likelihoods on an unknown dataset can be inferred by Bayesian expansion and can use a flat tissue membership probability from extracted feature vectors.

$$f(m=t|\nu) = \frac{f(\nu|m=t) f(m=t)}{\Sigma_{t'} f(\nu|m=t') f(m=t')} = \frac{f(\nu|m=t)}{\Sigma_{t'} f(\nu|m=t')} \quad (7)$$

Consider a desired label s as one tissue type t, and thus $f(T_i = s) = s) \equiv f(m = t|v)$, the Bayesian prior learning from context features serves to regularize the intermediate fusion estimate in Eq. 3, and hence the atlas selection.

3. METHODS

3.1 Data and Experimental Setup

Under Institutional Review Board (IRB) supervision, the first-session of abdomen CT scans of 75 metastatic liver cancer patients were randomly selected from an ongoing colorectal cancer chemotherapy trial, and an additional 25 retrospective scans were acquired clinically from post-operative patients with suspected ventral hernias. The 100 scans were captured with variable field of views (approx. 300 × 300 × 250 mm ~ 500 × 500 × 700 mm) and resolutions (approx. 0.5 × 0.5 × 1.5 mm ~ 1.0 × 1.0 × 7.0 mm). Twelve abdominal organs were manually labeled by two experienced undergraduate students, and verified by a radiologist on a volumetric basis using the MIPAV software (NIH, Bethesda, MD [13]). All images and labels were cropped along the cranio-caudal axis with a tight border without excluding liver, spleen, and kidneys before any processing (following [4]).

Forty of the 100 subjects were randomly selected as the atlases for validating five MAS approaches, including MV, SIMPLE, JLF, the Wolz approach, and our proposed method. The five approaches shared a common multi-stage registration procedure for each target image, where all atlases (excluding the target if in the set of atlases) were aligned to the target in the order of rigid, affine and a multi-level non-rigid registration using free-form deformations with B-spline control point spacings of 20, 10, and 5mm [14]. Ten of the 100 subjects were randomly selected as training datasets for context learning, thus we performed validation on the segmentations of the rest 90 subjects against the manual labels using DSC, mean surface distance (MSD), and Hausdorff distance (HD).

3.2 Implementations

We used 10 subjects to train context models for 15 tissue types, including twelve manually traced organs, and three automatically retrieved tissues (i.e., muscle, fat, and other) using intensity clustering and excluding the traced organ regions. Six context features were extracted, including intensity, gradient, and local variance, and three spatial coordinates with respect to a single landmark, which was loosely identified as the mid-frontal point of the lung at the plane with the largest cross-sectional lung area. We specified the number of components of Gaussian mixture model,N_G, = 3. For each organ, the foreground and background likelihoods were learned from the context models based on the context features on target images, and used as a two-fold spatial prior to regularize the organ-wise SIMPLE atlas selection. We constrained the number of selected atlases as no less than five and no larger than ten. When using JLF on the selected atlases for each organ, we specified the local search radii as 3 × 3 × 3, the local patch radii as 2 × 2 × 2, and set the intensity difference mapping parameter, and the regularization term as 2 and 0.1, respectively (i.e., default parameters). Following [4], we regularized the final segmentation with graph cut (GC), where the probabilistic fusion estimate from JLF and the intensity likelihoods using 1-D context learning were combined as the data term with a weight parameter of 3.3 over the

smoothness term. Note that we only applied GC smoothing to large organs (i.e., spleen, kidneys, liver, stomach), and kept the JLF results for the remaining organ structures.

For the direct JLF approach, the same parameters were used as above, except that it was conducted for all organs simultaneously. For the Wolz approach, we kept 30 atlases for the global atlas selection, adjusted the exponential decay for the organ level weighting to support 10 atlases, followed [4] for voxel-wise weighting by non-local means, and used the same GC scheme as applied to our proposed method.

4. RESULTS AND DISCUSSIONS

Compared to the other MAS approaches, the proposed method presents consistently improved segmentation in DSC on 11 of 12 organs of interest (Figure 3, Table 1). Based on the mean DSC of each organ, a median improvement of 7.0% and 16.2% were achieved over JLF and Wolz, respectively. The segmentations of spleen, gallbladder, esophagus, and aorta using the proposed method significantly outperformed those using the other approaches.

The serpentine labels of portal vein and splenic vein are barely captured by registration (0.06 in DSC by median), thus the intermediate voting-based fusion estimates had a good chance of missing the structure entirely (zero median in DSC for MV and SIMPLE). A MV fusion (instead of JLF) of the selected atlases by SIMPLE context learning identified this structure better (0.25 in DSC by median). While with limited atlases of catastrophic registration errors, our proposed method was outperformed by JLF with all available atlases.

On the other hand, in the context of reasonably substantial registration errors for other organs, our proposed method yields segmentation with better performances in not only accuracies, but also efficiencies. With much fewer atlases (while more target-alike than average) included for label fusion, our method relieved massive computational time and memory required by JLF and Wolz approach, and thus provides more efficient abdominal segmentations. As found in our previous study, the MV fusion of the registered atlases with the top five DSC achieves a median DSC of 0.9 for spleen [7]. Therefore, we considered the global non-linear selection of the atlases as a necessary procedure in addition to the locally weighted label determination for MAS in abdomen.

With a closer look, our proposed method yielded the segmentation with at least 0.89 in DSC and less than 3.3 mm in MSD for the major organs of interest, i.e., spleen, kidneys, and liver. For other structures, the proposed method also provided successful identification over half of the subjects, even those that empirically considered difficult to capture, e.g., adrenal glands (Table 2). Qualitatively, the segmentation on a subject with median accuracy captures the organs from the perspective of both 3-D rendering and 2-D coronal slices (Figure 4).

In conclusion, the proposed method provides a fully automated approach to segment twelve abdominal organs on clinically acquired CT. The SIMPLE context learning reduces the impact of the vastly problematic registrations with appropriate atlas selection considering exogenous contexts in addition to intermediate fusion estimate, and thus enables more

efficient abdominal segmentations. The estimated segmentation could be used in large-scale trials to provide abdominal surgical navigation, organ-wise biomarker derivation, or volumetric screening. The method also enables explorative studies on the correlation the structural organ metrics with surgical/physiological conditions. Further refinement of the proposed method may well consider deeper integration between SIMPLE context learning and JLF.

ACKNOWLEDGMENTS

This research was supported by NIH 1R03EB012461, NIH 2R01EB006136, NIH R01EB006193, ViSE/VICTR VR3029, NIH UL1 RR024975-01, NIH UL1 TR000445-06, NIH P30 CA068485, and AUR GE Radiology Research Academic Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

[1]. Linguraru MG, Pura JA, Pamulapati V, et al. Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. Medical image analysis. 2012; 16(4):904–914. [PubMed: 22377657]

[2]. Shimizu A, Ohno R, Ikegami T, et al. Segmentation of multiple organs in non-contrast 3D abdominal CT images. International journal of computer assisted radiology and surgery. 2007; 2(3-4):135–142.

[3]. Sabuncu MR, Yeo BT, Van Leemput K, et al. A generative model for image segmentation based on label fusion. IEEE transactions on medical imaging. 2010; 29(10):1714–29. [PubMed: 20562040]

[4]. Wolz R, Chu C, Misawa K, et al. Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE transactions on medical imaging. 2013; 32(9):1723–30. [PubMed: 23744670]

[5]. Langerak TR, van der Heide UA, Kotte AN, et al. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). Medical Imaging, IEEE Transactions on. 2010; 29(12):2000–2008.

[6]. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945; 26(3): 297–302.

[7]. Xu, Z.; Asman, AJ.; Shanahan, PL., et al. SIMPLE IS a Good Idea (and Better with Context Learning). Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI); 2014. in press

[8]. Wang H, Suh JW, Das SR, et al. Multi-Atlas Segmentation with Joint Label Fusion. IEEE transactions on pattern analysis and machine intelligence. 2012

[9]. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2001; 23(11):1222–1239.

[10]. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004; 23(7):903–21. [PubMed: 15250643]

[11]. Asman AJ, Landman BA. Robust statistical label fusion through COnsensus Level, Labeler Accuracy, and Truth Estimation (COLLATE). IEEE Trans Med Imaging. 2011; 30(10):1779–94. [PubMed: 21536519]

[12]. Van Leemput K, Maes F, Vandermeulen D, et al. Automated model-based bias field correction of MR images of the brain. IEEE transactions on medical imaging. 1999; 18(10):885–896. [PubMed: 10628948]

[13]. McAuliffe, MJ.; Lalonde, FM.; McGarry, D., et al. Medical image processing, analysis and visualization in clinical research. Computer-Based Medical Systems, 2001; CBMS 2001. Proceedings. 14th IEEE Symposium on; 2001. p. 381-386.

[14]. Rueckert D, Sonoda LI, Hayes C, et al. Nonrigid registration using free-form deformations: Application to breast MR images. IEEE transactions on medical imaging. 1999; 18(8):712–721. [PubMed: 10534053]

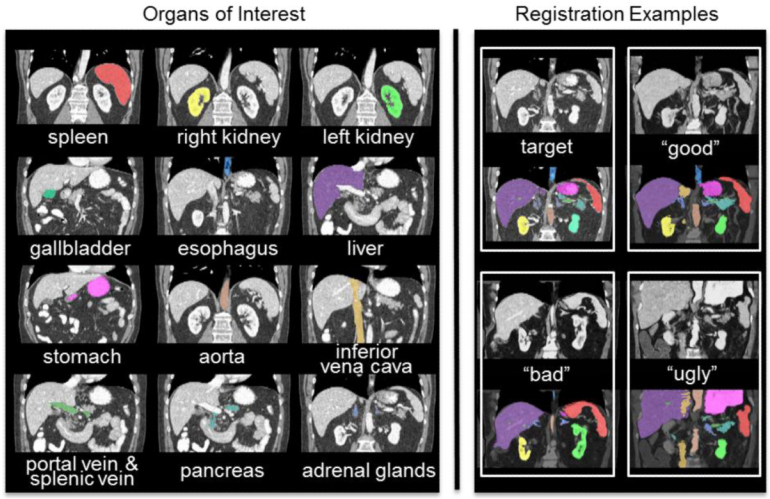

Figure 1.
Twelve organs of interest (left) and registration examples of variable qualities for one target image (right).

Figure 2.

Flowchart of the proposed method. Given registered atlases with variable qualities, atlas selection and statistical fusion are considered as two necessary steps to obtain a reasonable fusion estimate of the target segmentation. The SIMPLE algorithm implicitly combines these two steps to fusion selected atlases; however, more information can be incorporated to improve the atlas segmentation, and a more advanced fusion technique can be used after the atlases are selected. We propose to (1) extract a probabilistic prior of the target segmentation by context learning to regularize the atlas selection in SIMPLE for each organ, (2) use Joint Label Fusion to obtain the probabilistic fusion estimate while characterizing of the correlated errors from among the selected organ-specific atlases, and render the final segmentation for all organs via graph cut.

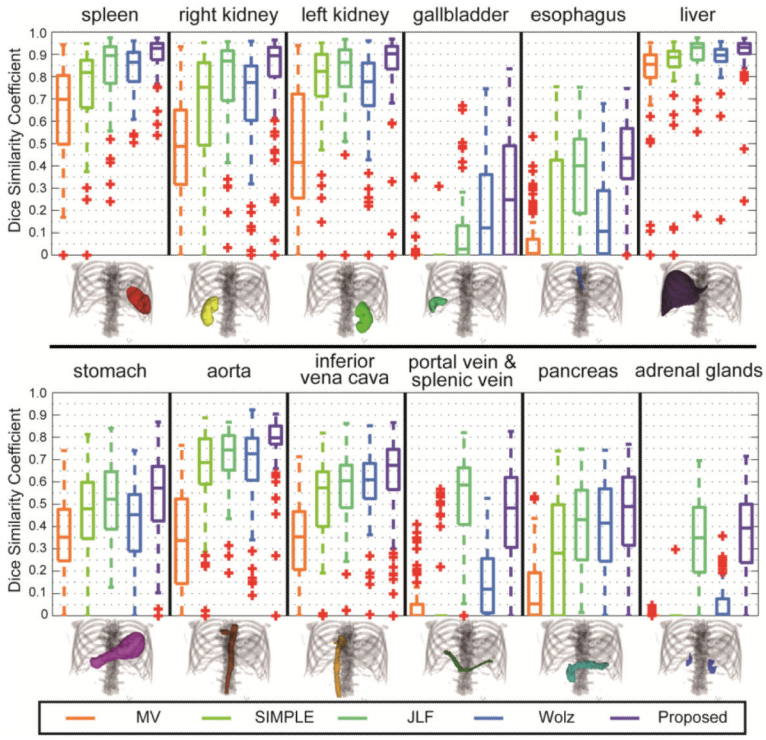

Figure 3.
Boxplot comparison among five tested methods for 12 organs.

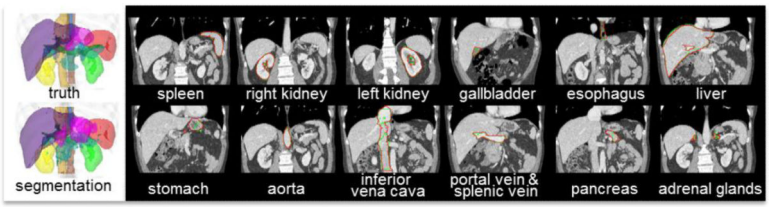

Figure 4.

Qualitative segmentation results on a subject with median DSC. On the left, the 3-D organ labels are rendered for the true segmentation, and the proposed segmentation. On the right, the truth (red) and the proposed segmentation (green) for each organ of interest are demonstrated on a representative coronal slice.

Table 1

Quantitative evaluation for five tested methods using dice similarity coefficient (mean ± std.).

	MV	SIMPLE	JLF	Wolz	Proposed
Spleen	0.63 ± 0.24	0.73 ± 0.22	0.84 ± 0.15	0.83 ± 0.10	**0.90 ± 0.08****
R. Kidney	0.47 ± 0.26	0.65 ± 0.27	0.79 ± 0.19	0.70 ± 0.24	**0.81 ± 0.20**
L. Kidney	0.46 ± 0.27	0.74 ± 0.25	0.81 ± 0.17	0.72 ± 0.21	**0.84 ± 0.20**
Gallbladder	0.01 ± 0.04	0.00 ± 0.03	0.09 ± 0.15	0.19 ± 0.21	**0.27 ± 0.26***
Esophagus	0.07 ± 0.11	0.20 ± 0.25	0.37 ± 0.21	0.18 ± 0.19	**0.43 ± 0.18***
Liver	0.79 ± 0.20	0.84 ± 0.18	0.89 ± 0.11	0.88 ± 0.09	**0.91 ± 0.09**
Stomach	0.34 ± 0.18	0.46 ± 0.19	0.51 ± 0.17	0.41 ± 0.19	**0.55 ± 0.18**
Aorta	0.34 ± 0.22	0.64 ± 0.22	0.72 ± 0.13	0.67 ± 0.18	**0.77 ± 0.13***
IVC	0.33 ± 0.18	0.50 ± 0.21	0.57 ± 0.15	0.58 ± 0.15	**0.62 ± 0.19**
PV & SV	0.05 ± 0.10	0.05 ± 0.15	**0.52 ± 0.20****	0.16 ± 0.16	0.45 ± 0.21
Pancreas	0.11 ± 0.13	0.27 ± 0.25	0.40 ± 0.19	0.40 ± 0.19	**0.45 ± 0.21**
A. Glands	0.00 ± 0.01	0.00 ± 0.03	0.34 ± 0.20	0.05 ± 0.08	**0.36 ± 0.19**

* indicates that the DSC value was significantly higher than the second best DSC across the methods for the organ segmentation as determined by a right-tail paired t-test with $p<0.05$.

** indicates a $p<0.01$.

Table 2

Quantitative metrics of the proposed segmentation method.

Metrics	***Dice Similarity Coefficient***	***Surface Distance (mm)***	
Organs	**Median [Min, Max]**	**Sym. MSD**	**Sym. HD**
Spleen	0.93 [0.54, 0.97]	1.75 ± 1.71	17.27 ± 8.42
R. Kidney	0.89 [0.00, 0.96]	2.99 ± 3.92	19.47 ± 11.37
L. Kidney	0.90 [0.00, 0.97]	2.00 ± 2.80	16.13 ± 8.05
Gallbladder	0.25 [0.00, 0.84]	14.36 ± 20.34	34.57 ± 22.87
Esophagus	0.43 [0.00, 0.75]	4.16 ± 2.05	17.97 ± 5.46
Liver	0.93 [0.24, 0.97]	3.22 ± 4.43	34.46 ± 15.03
Stomach	0.57 [0.00, 0.87]	10.26 ± 6.36	49.48 ± 18.91
Aorta	0.80 [0.00, 0.90]	3.02 ± 2.27	23.23 ± 10.98
IVC	0.67 [0.00, 0.87]	3.75 ± 1.84	19.89 ± 5.60
PV & SV	0.48 [0.00, 0.83]	5.92 ± 5.08	38.37 ± 17.18
Pancreas	0.49 [0.00, 0.77]	5.47 ± 3.51	31.34 ± 8.92
A. Glands	0.39 [0.00, 0.72]	4.06 ± 3.56	20.68 ± 8.68