Liver segmentation in color images

Burton Ma[a], T. Peter Kingham[b], William R. Jarnagin[b], Michael I. Miga[c], and Amber L. Simpson[b]

[a]Dept. of Electrical Engineering and Computer Science, York University, Toronto, Canada
[b]Hepatopancreatobiliary Service, Memorial Sloan Kettering Cancer Center, NY, NY, USA
[c]Biomedical Engineering, Vanderbilt University, Nashville, TN, USA

ABSTRACT

We describe the use of a deep learning method for semantic segmentation of the liver from color images. Our intent is to eventually embed a semantic segmentation method into a stereo-vision based navigation system for open liver surgery. Semantic segmentation of the stereo images will allow us to reconstruct a point cloud containing the liver surfaces and excluding all other non-liver structures. We trained a deep learning algorithm using 136 images and 272 augmented images computed by rotating the original images. We tested the trained algorithm on 27 images that were not used for training purposes. The method achieves an 88% median pixel labeling accuracy over the test images.

Keywords: liver, segmentation, color image, deep learning

1. INTRODUCTION

Partial liver resection is the most effective and the only potentially curative therapy for many primary and secondary liver tumors[1,2] but is technically challenging due to the complex vasculature and the need maintain sufficient remnant volume to ensure intact vascular inflow and outflow and biliary drainage. In order to achieve an adequate tumor resection or ablation, with good oncologic outcomes and without injury to and needless sacrifice of adjacent normal hepatic parenchyma, the surgeon must have precise knowledge of the tumor location relative to the surrounding major vascular and biliary structures. With these factors in mind, image-guided liver surgery aims to enhance the precision of resection and ablation, while preserving uninvolved parenchyma and vascular structures. Current navigational technologies provide intraoperative guidance for a single reference point in time; once the liver is disturbed, either through patient respiration or surgical manipulation, accuracy is lost. Our goal is to overcome the primary limitation of current navigation systems – inaccuracy due to the inability to account for intraoperative organ changes during resection. We aim to move from the static paradigm of current navigation systems to real-time continuous imaging and feedback as the surgical intervention proceeds. Toward this end, we propose a method to automatically label anatomical regions of interest in the liver in video.

2. METHODS

We retrospectively analyzed video recordings of approximately 100 open liver surgeries to search for images of an exposed anterior aspect of the liver. These videos were recordings of conventional liver surgeries, and there was wide variability in the camera type, orientation, view point, aspect ratio, exposure, lighting conditions, and zoom. From these videos, we found 136 images of the anterior aspect of the liver, although most of these images were not entirely suitable for registration purposes for a variety of reasons (such as insufficient exposure, only one lobe visible, presence of instrumentation, manipulation of the liver). We acquired an additional 26 images from a study involving a laser-range scanner with a co-registered camera for navigated liver surgery.[3] These images were more consistent in appearance because the same camera system was used to capture all of the images, but there were significant differences in the camera view point, orientation, and liver exposure. A total of $(136 + 26) = 162$ images were collected. Examples of the training images are shown in Fig. 1.

Further author information: (Send correspondence to B.M.)
B.M.: E-mail: burton@cse.yorku.ca
A.L.S.: E-mail: simpsonl@mskcc.org

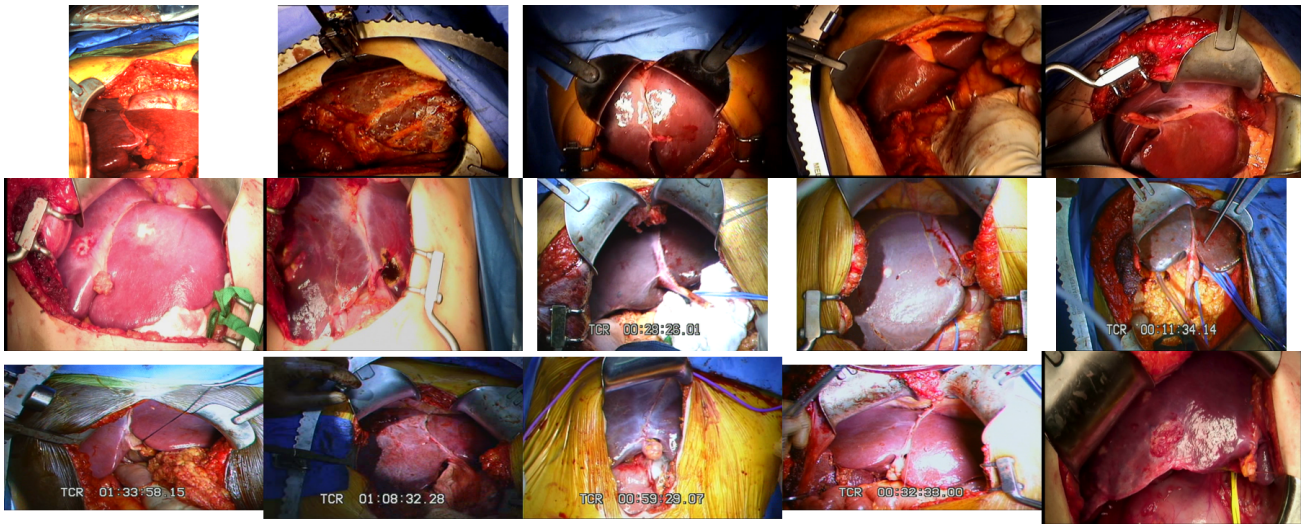

Figure 1. A sample of 15 (of 162) training images. The images have been scaled to fit in the figure, but the aspect ratios are unchanged.

The images were downsampled by a factor of two so that the largest image dimension was no more than 500 pixels to ensure that training of the segmentation algorithm could be performed within the limits of our video card RAM (see next paragraph for details). The images were manually segmented by a single observer (B.M.). For the purposes of this study, we were primarily concerned with segmenting the whole liver rather than the parts of the liver; however, we chose to manually segment the parts of the liver noting that it would be easy to merge the segmented parts into a whole liver segmentation. The manual segmentation produced a a ground truth labeling of four classes: the left lobe, right lobe, falciform ligament, and instrumentation. 27 images were selected as test images with the main criteria being the image had to be suitable for registration purposes. We augmented the training set by rotating the training images and their segmentations by $\pm 30^\circ$ and added the rotated images to the training set to produce a total of 408 training images.

We trained a semantic segmentation algorithm (CRFasRNN) that utilizes deep learning techniques and probabilistic graphical models.[4] The CRFasRNN network is the FCN8s network[5] followed by a layer that implements mean-field conditional random field inference using a recurrent neural network. CRFasRNN is implemented using the Caffe deep learning framework.[6] We use a GPU accelerated version of the mean-field layer.[7] Training was primarily GPU bound and required approximately 9 GB of video RAM. We used an NVIDIA Titan X video card having 12 GB of video RAM.

We first trained the FCN8s network using the four object classes along with the twenty classes from the PASCAL VOC 2012 Challenge[8] (including the additional training data described in Ref. 9) following the methods described in Ref. 5, 10. We used the PASCAL object classes to verify that we were performing the training correctly and to mitigate overfitting caused by our small amount of liver training data. We then trained the CRFasRNN end-to-end following the methods described in Ref. 4, 11.

We tested the performance of the trained network on the test images and compared the resulting segmentations to the ground truth labelings. We computed pixel labeling accuracy rate, false positive rate, and the false negative rate, all expressed as ratios relative to the number of ground truth liver pixels:

$$\text{accuracy} = \frac{\text{number of pixels correctly labeled as liver}}{\text{number of ground truth liver pixels}} \quad (1)$$

$$\text{false positive rate} = \frac{\text{number of pixels incorrectly labeled as liver}}{\text{number of ground truth liver pixels}} \quad (2)$$

$$\text{false negative rate} = \frac{\text{number of pixels incorrectly labeled as not liver}}{\text{number of ground truth liver pixels}} \quad (3)$$

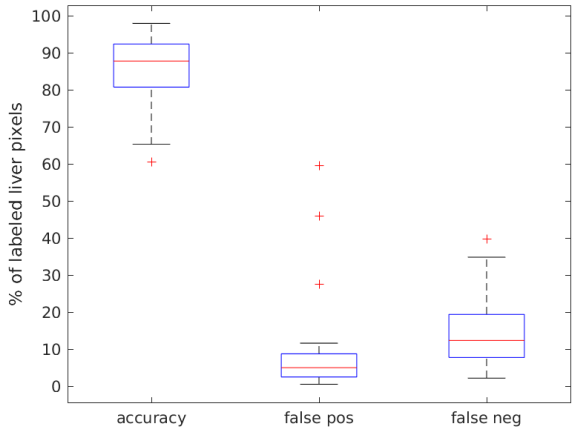

Figure 2. Box plot of accuracy, false positive, and false negative rates showing median (red line), interquartile range (blue box), range of non-outlying results (whiskers), and outlying results (red crosses).

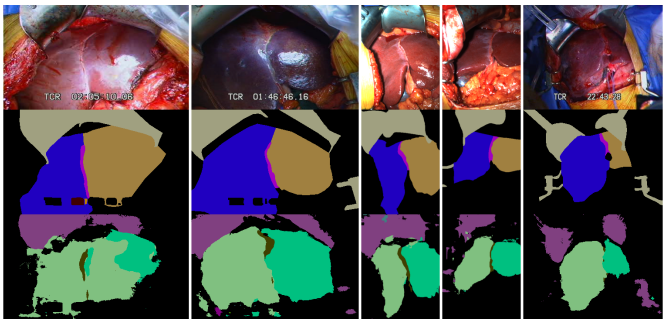

Figure 3. Five example segmentations with near-median accuracy. The top row shows the test images scaled to fit in the figure. The middle row shows the ground truth segmentations (blue = right lobe, magenta = falciform ligament, orange = left lobe, gray = instrumentation). The bottom row shows the computed segmentations (light green = right lobe, dark gray = falciform ligament, green = left lobe, purple = instrumentation). Pixel labeling accuracies for the whole liver (not the individual parts) are 80%, 87%, 91%, 92%, and 92% from left to right.

Note that we do not consider the part labelling accuracy or errors; our interest in this work is labelling of the whole liver.

3. RESULTS

A box plot of the accuracy, false positive, and false negative rates is shown in Fig. 2. The median accuracy was 88%, the median false positive rate was 5%, and the median false negative rate was 12%. Example segmentations with near-median accuracy are shown in Fig. 3. Example segmentations with high false positive and false negative rates are shown in Fig. 4.

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

We are planning several new additions to this work that will be presented in the final manuscript. We are updating the manual segmentations to include labeling of the diaphragm with the goal of training the network to distinguish between liver and diaphragm. We are also continually collecting additional images from clinical cases and we plan on updating the training and test image sets. We also plan on experimenting with other deep learning semantic segmentation algorithms. Finally, we will expand on the discussion section in the final manuscript.

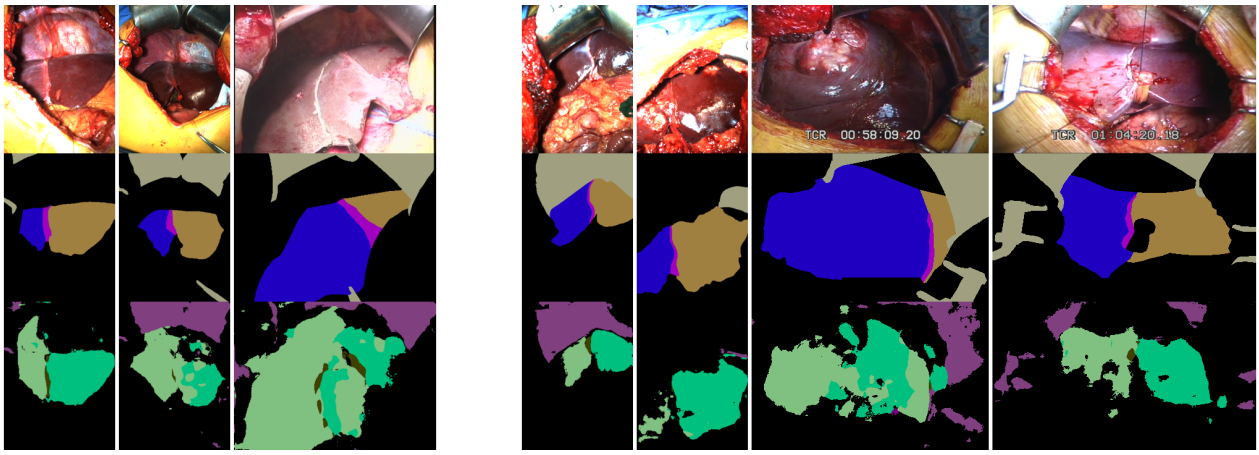

Figure 4. Three example segmentations with (left) high false positive rates and (right) high false negative rates. The top row shows the test images scaled to fit in the figure. The middle row shows the ground truth segmentations (blue = right lobe, magenta = falciform ligament, orange = left lobe, gray = instrumentation). The bottom row shows the computed segmentations (light green = right lobe, dark gray = falciform ligament, green = left lobe, purple = instrumentation). False positive rates for the whole liver (not the individual parts) are 28%, 46%, and 59% for the three leftmost columns. False negative rates for the whole liver (not the individual parts) are 30%, 30%, 34%, and 40% for the four rightmost columns.

5. DISCUSSION AND CONCLUSIONS

The high median correctness rate of 88% is encouraging, especially when considering the relatively small size of the training data set. We are collecting additional training and test data, and plan on experimenting with other deep learning semantic segmentation algorithms.

There were three cases where the false positive rate exceeded 10%. In all of these cases, the CRFasRNN mislabeled pixels corresponding to the diaphragm as being liver. We plan on relabeling the training data to include a diaphragm class and retraining the segmentation algorithm.

There were four cases where the false negative rate exceeded 20%. In these cases, the algorithm failed to label liver pixels that were in deep shadow, had strong specular highlights, corresponded to diseased tissue or blood. A typical stereo reconstruction algorithm would also fail to reconstruct surfaces in deep shadow or with strong highlights; thus, when considering the application of liver segmentation for the purposes of filtering a stereo point cloud, the failure to label such pixels may not be problematic.

In conclusion, we investigated the use of a deep learning semantic segmentation algorithm for segmenting the liver in color images. The eventual goal is to use liver segmentation in stereo images to retain stereo reconstructed points only from the liver surfaces. The CRFasRNN algorithm achieved a median pixel labeling correctness rate of 88% when tested on 27 images from liver surgeries.

This work has not been submitted for publication or presentation elsewhere.

ACKNOWLEDGMENTS

This work is funded in part by grant R01CA162477 and Cancer Center Support Grant P30 CA008748 from the National Cancer Institute.

REFERENCES

[1] Sardi, A., Akbarov, A., and Conaway, G., "Management of primary and metastatic tumors to the liver," *Oncology* **10**(6), 911–25; discussion 926, 929–30 (1996).

[2] Rosen, C. B., "Management of Hepatic Metastases," *Cancer Control* **5**(3 Suppl 1), 30–31 (1998).

[3] Pheiffer, T. S., Simpson, A. L., Lennon, B., Thompson, R. C., and Miga, M. I., "Design and evaluation of an optically-tracked single-ccd laser range scanner," *Medical physics* **39**(2), 636–642 (2012).

[4] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., and Torr, P. H., "Conditional random fields as recurrent neural networks," in [*Proceedings of the IEEE International Conference on Computer Vision*], 1529–1537 (2015).

[5] Long, J., Shelhamer, E., and Darrell, T., "Fully convolutional networks for semantic segmentation," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 3431–3440 (2015).

[6] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093* (2014).

[7] hyenal, "hyenal/crfasrnn." https://github.com/hyenal/crfasrnn.

[8] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A., "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision* **88**, 303–338 (June 2010).

[9] Hariharan, B., Arbeláez, P., Girshick, R., and Malik, J., "Simultaneous detection and segmentation," in [*European Conference on Computer Vision (ECCV)*], (2014).

[10] Long, J., Shelhamer, E., and Darrell, T., "shelhamer/fcn.berkeleyvision.org." https://github.com/shelhamer/fcn.berkeleyvision.org.

[11] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., and Torr, P. H., "torrvision/crfasrnn." https://github.com/torrvision/crfasrnn.