Deep residual networks for automatic segmentation of laparoscopic videos of the liver

Eli Gibson[ab*], Maria R. Robu[a], Stephen Thompson[a], P. 'Eddie' Edwards[a], Crispin Schneider[c], Kurinchi Gurusamy[c], Brian Davidson[c], David J. Hawkes[a], Dean C. Barratt[a], Matthew J. Clarkson[a]
[a] UCL Centre for Medical Imaging Computing, Department of Medical Physics and Biomedical Engineering, University College London (UCL), London, UK; [b] Diagnostic Image Analysis Group, Radboud University Medical Centre, Nijmegen, The Netherlands; [c] Division of Surgery and Interventional Sciences, Royal Free Hospital, London, UK;

ABSTRACT

Motivation: For primary and metastatic liver cancer patients undergoing liver resection, a laparoscopic approach can reduce recovery times and morbidity while offering equivalent curative results; however, only about 10% of tumours reside in anatomical locations that are currently accessible for laparoscopic resection. Augmenting laparoscopic video with registered vascular anatomical models from pre-procedure imaging could support using laparoscopy in a wider population. Segmentation of liver tissue on laparoscopic video supports the robust registration of anatomical liver models by filtering out false anatomical correspondences between pre-procedure and intra-procedure images. In this paper, we present a convolutional neural network (CNN) approach to liver segmentation in laparoscopic liver procedure videos.

Method: We defined a CNN architecture comprising fully-convolutional deep residual networks with multi-resolution loss functions. The CNN was trained in a leave-one-patient-out cross-validation on 2050 video frames from 6 liver resections and 7 laparoscopic staging procedures, and evaluated using the Dice score.

Results: The CNN yielded segmentations with Dice scores ≥0.95 for the majority of images; however, the inter-patient variability in median Dice score was substantial. Four failure modes were identified from low scoring segmentations: minimal visible liver tissue, inter-patient variability in liver appearance, automatic exposure correction, and pathological liver tissue that mimics non-liver tissue appearance.

Conclusion: CNNs offer a feasible approach for accurately segmenting liver from other anatomy on laparoscopic video, but additional data or computational advances are necessary to address challenges due to the high inter-patient variability in liver appearance.

Keywords: Segmentation, deep learning, laparoscopic video, liver resection, liver staging, minimally invasive surgery, image-guided interventions

1. INTRODUCTION

Liver resection is the main curative treatment for eligible patients with primary liver cancer and liver-only colorectal metastases, with over 2000 resections performed annually in the United Kingdom. Minimally invasive laparoscopic liver resections can result in shorter recovery times and lower postoperative morbidity compared to open procedures with equivalent curative results[1]. There has been a widespread uptake of the laparoscopic approach for minor and non-complex liver resections. However, only a few centres have reported significant numbers of major or complex laparoscopic liver resections which implies that only a relatively small number of patients with more extensive disease have access to the potential advantages offered by a laparoscopic liver resection[2,3]. This discrepancy is thought to be due, in part, to concerns about identifying major vascular structures and tumour margins, which is an essential step to avoid vascular injury and incomplete tumour resection.

Preliminary evidence from other anatomical locations (e.g. neurosurgery[4]) suggests that combining anatomical models from pre-procedure imaging with intra-procedural imaging supports the localization of vasculature and tumors. This suggests that accurate registration of patient-specific vascular models from pre-procedure images[5,6] to intra-procedural laparoscopic video may enable surgeons to avoid these structures and thus support the use of laparoscopic resection in a wider patient population.

* eli.gibson@ucl.ac.uk

Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling, edited by
Robert J. Webster III, Baowei Fei, Proc. of SPIE Vol. 10135, 101351M · © 2017 SPIE
CCC code: 1605-7422/17/$18 · doi: 10.1117/12.2255975

Registration of a liver model (extracted from pre-procedure CT) to stereoscopic laparoscopic video by fitting the model to liver surface patches extracted using dense stereoscopic point correspondence has been demonstrated in pre-clinical porcine models[6]. However, preliminary evaluation on video from human laparoscopic procedures suggested that surface patches extracted from non-liver anatomy limited the robustness of the registration. Recent algorithms for liver segmentation on laparoscopic video[7] hold the potential to eliminate non-liver patches, but the sensitivity to parameter tuning was identified as a major weakness. Recent advances in convolutional neural networks (CNNs), in contrast, enable parameters to be learnt from image data *directly*. In this work, we present our preliminary findings on the application of CNNs to liver segmentation on laparoscopic video.

2. METHODS

Patients

We analysed laparoscopic video from patients undergoing liver resection (N=6) and staging laparoscopy (N=7) with informed consent and the approval of our institutional research ethics board.

Imaging

Laparoscopic video images were captured from a Storz TIPCAM 3D stereo laparoscope. Images were output in DVI format as a 1920×1080 pixel image representing left and right channels interleaved (effective 1920×540 pixels per channel). They were converted to SDI using an AJA ROI-DVI mini-converter, and grabbed using the NVIDIA Quadro SDI capture card. We simultaneously recorded video at 29.9 Hz, and tracking data at 40Hz using the NifTK software platform[8]. The data were timestamped, recorded to a hard-disk during each procedure, and processed offline.

Reference standard

Two thousand and fifty laparoscopic video frames were segmented manually from the 13 intra-procedural videos. Because adjacent video frames can be highly correlated (decreasing their value for machine learning), video frames were selected at 50 or 100 frame intervals, manually excluding frames that showed highly similar views of the liver and highly similar textures, colours and shapes. Manual contouring was performed in NiftyIGI by a clinical research associate in General Surgery (C.S).

Deep learning architecture

The segmentation used a fully-convolutional neural network (illustrated in Figure 1), implemented using the Caffe deep learning framework[9]. The network comprised a convolutional feature layer, four deep residual learning units[10], three parallel intermediate segmentation units with inputs at 3 different resolutions from the 2nd, 3rd, and 4th residual units; and a fusion layer to combine the intermediate segmentations.

The convolutional feature unit used 32 outputs and a 3×3 kernel. Each deep residual learning unit comprised a convolutional path (2 sets of pre-activation[11] and 3×3 convolution layers, with the feature count doubled and the resolution halved by the first set) and a shortcut path (using identity shortcuts with 3×3 average pooling to preserve information from the previous layer and 3×3 projection shortcuts to increase the feature count). Each intermediate segmentation units comprised 2 sets of pre-activation[11] and 3×3 convolution layers. The fusion layer upsampled the 2 lowest resolution segmentations and combined intermediate segmentations into a composite 81×21 pixel segmentation.

Each segmentation layer outputted the probability v_i that each pixel was liver tissue. The objective function (i.e. network loss) was a weighted sum of the per-pixel logistic loss ($-0.4 * log(v_i)$ for liver voxels and $-0.6 * log(1 - v_i)$ for non-liver voxels) for each intermediate segmentation output and for the composite segmentation. The neural network was trained for 10,000 iterations, comprising a stochastic gradient descent step computed from a 20-image 'mini-batch', with weight-decay=4e-4, momentum=0.9, and learning rates starting at 1e-4 and halving every 1000 iterations.

Experimental design

In order to maximize the size of the training data, the network was trained using a 13-fold cross-validation: for each fold, all data from one from patient was excluded from training; the network was trained on data from the 12 remaining patients; and the segmentation accuracy was measured on data from the excluded patient. Segmentation was evaluated using 2 metrics: the Dice score for final results and a per-voxel accuracy metric (the average of the true positive and false positive rates), measured during the training process.

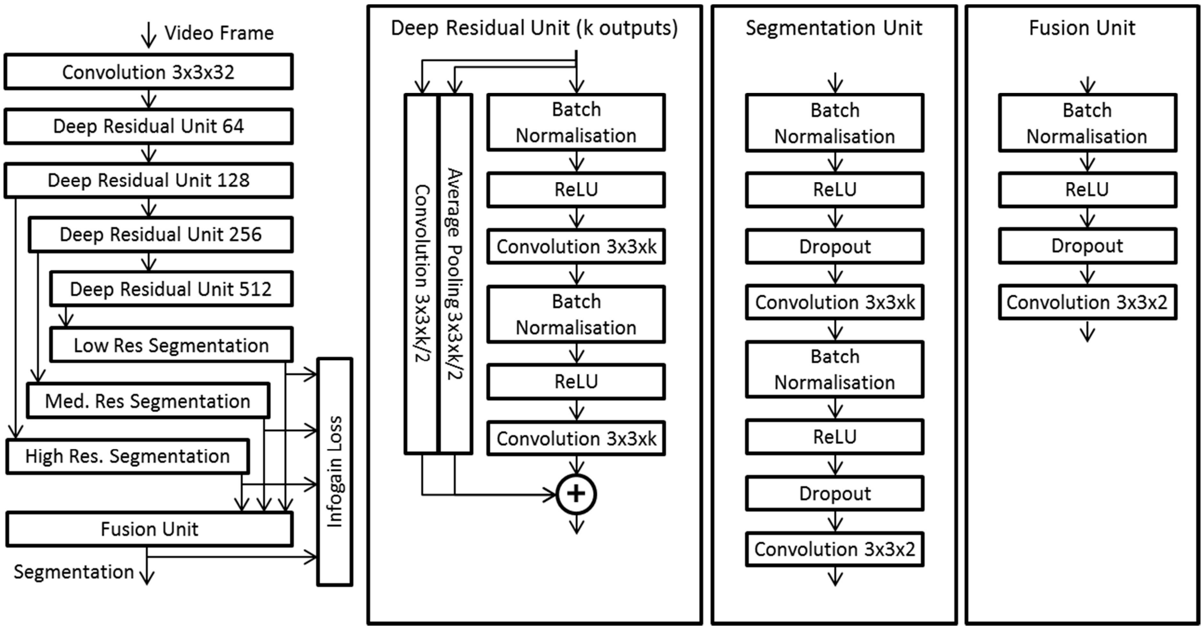

Figure 1: Neural network architecture

3. RESULTS

Segmentation performance

The median Dice score was 0.95; two illustrative examples with this median Dice score are shown in Figure 2. The median Dice score within each patient was ≥0.95 for 9/13 patients; however, it varied considerably between patients with scores of 0.78, 0.89, 0.90, 0.94, 0.95, 0.95, 0.97, 0.97, 0.97, 0.97, 0.97, 0.97, and 0.98 calculated for each patient.

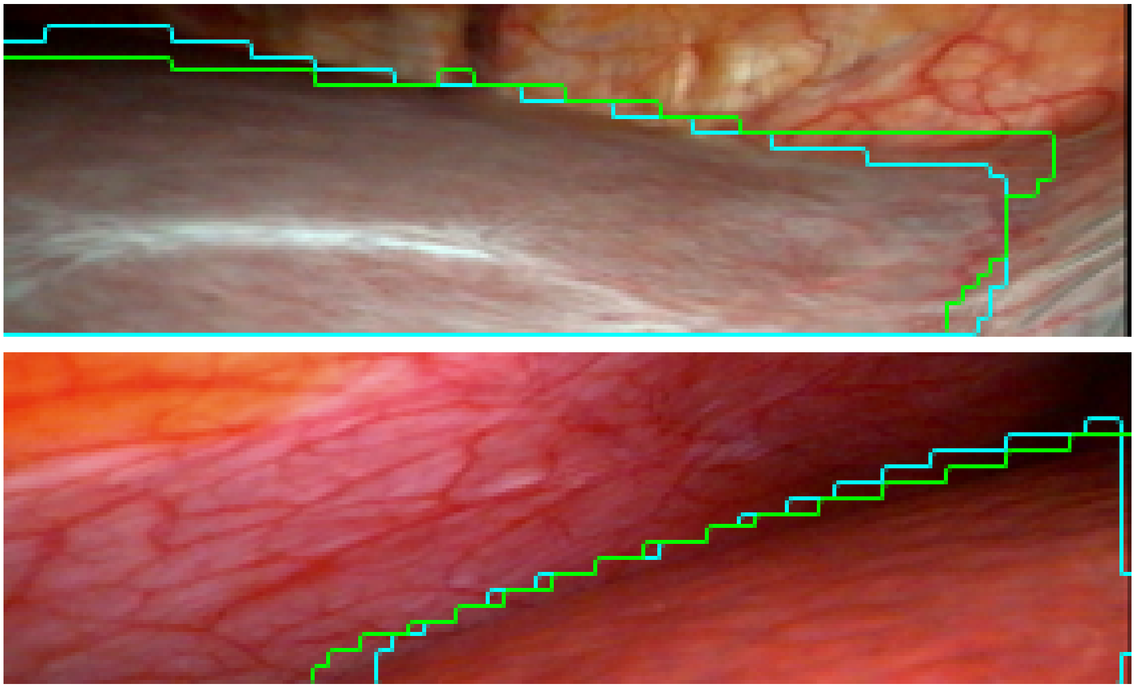

Figure 2: Two frames with the CNN's median Dice score (0.95), with the reference standard (green) and the automatic segmentation (cyan) overlaid. The CNN distinguishes between liver and non-liver tissues with similar colours, suggesting the CNN is encoding textural information as well.

Qualitative analysis of the segmentations identified four common failure modes, illustrated in Figure 3: inter-procedure variability in liver appearance, automatic exposure correction, and pathological tissue which mimics the appearance of non-liver tissue. Inter-procedure variability in liver appearance may be caused by pathology (e.g., bile duct obstruction, fatty liver [steatosis] and liver cirrhosis), lighting, or camera calibration differences.

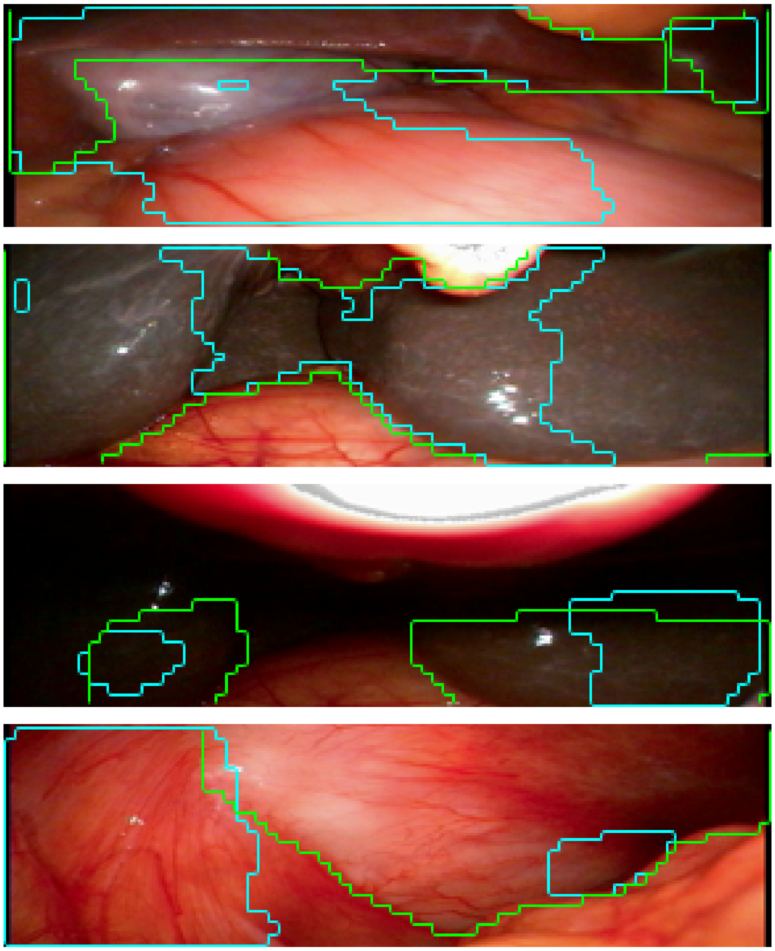

Figure 3: Four illustrative video frames, with the reference standard (green) and the automatic segmentation (cyan) overlaid, showing failure modes of the segmentation: minimal visible liver tissue (top left), inter-procedure variability in liver appearance (top right), automatic exposure correction (bottom left), and pathological tissue whose appearance mimics non-liver tissue (bottom right).

Training

Figure 4 shows the network loss and the accuracy of segmentation on the training and testing sets over the course of training; the testing accuracy showed higher final variability than the network loss and training accuracy suggesting that the lower performance on some patients reflects poor generalization.

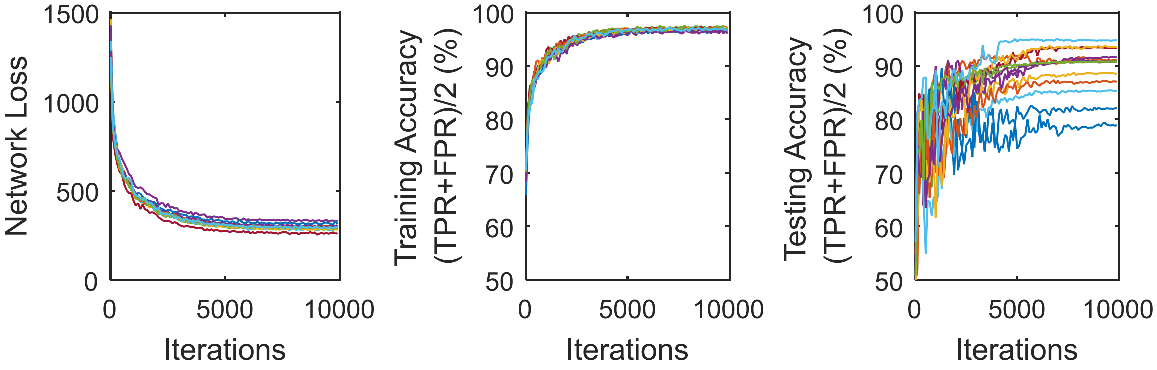

Figure 4: Learning curves from the leave-one-patient-out cross-validation. The network loss objective function and training accuracy converged to similar values for different patients, as expected in leave-one-out cross-validation. The accuracy on the testing set also converged; however, the final values had substantial inter-patient variability (see text).

4. DISCUSSION

This work demonstrated the feasibility of using convolutional neural networks to segment anatomy from laparoscopic video frames. Despite the small sample size and despite a data set chosen for its heterogeneity, the segmentation accuracy for the majority of frames was high. Further systems testing in the context of registering pre-procedural anatomic models to laparoscopic video is necessary to evaluate the system performance within the intraoperative setting. To support this clinical evaluation, the segmentation method and the Caffe deep learning framework[9] have been integrated into the NifTK software platform[8] and integration into a clinical laparoscopy system is ongoing.

The per-patient median Dice score was variable, in part due using 13 fold cross-validation yielding a less precise estimator than one using fewer folds; however, this enabled a larger training set which is critical for machine learning approaches that are sensitive to data size (e.g. deep learning). The lowest median Dice score (0.78) was for a patient with minimal visible liver tissue, partly because the Dice score is sensitive to small segmentation errors when reference areas are small. However, such errors have little clinical impact since good views of the liver are typically obtained before any resection activity. The next lowest Dice score (0.89) was for a staging laparoscopy on the only patient with liver congestion (due to biliary obstruction) resulting in an appearance not seen in the training set. Liver resections are typically performed on patients with healthy liver parenchyma on the surface of the organ, limiting the clinical impact of these errors.

The failure modes of the CNN suggest avenues for future research. In particular, frames with minimal visible liver tissue could be addressed by using image mosaics to expand the field of view[12]. Pathological tissues might be addressed by expanding the training data set, but patient-specific calibration of CNNs (e.g. via Siamese networks[13] or one-shot learners[14]) may also improve generalization. Automatic exposure correction is helpful for clinicians but challenging for the CNN. This could be addressed by normalization in pre-processing or application-specific data augmentation.

5. CONCLUSIONS

Convolutional neural networks are a feasible approach for accurately segmenting liver from other anatomy on laparoscopic video, yielding median Dice scores ≥0.95. Additional data or computational advances are necessary to address the high inter-patient variability in liver appearance.

6. ACKNOWLEDGEMENTS

This publication presents independent research commissioned by the Health Innovation Challenge Fund (HICF-T4-317), a parallel funding partnership between the Wellcome Trust and the Department of Health. The views expressed in this publication are those of the authors and not necessarily those of the Wellcome Trust or the Department of Health. Eli Gibson was supported by a Canadian Institutes of Health Research Post-doctoral Fellowship and by the UK Medical Research Council. The authors also thank NVIDIA for the donation of the Quadro K5000 and SDI capture card.

REFERENCES

[1] Nguyen, K. T., “Comparative benefits of laparoscopic vs open hepatic resection,” Archives of Surgery 146(3), 348–356 (2011). https://doi.org/10.1001/archsurg.2010.248

[2] Goumard, C., Farges, O., Laurent, A., Cherqui, D., Soubrane, O., Gayet, B., Pessaux, P., Pruvot, F.-R., and Scatton, O., “An update on laparoscopic liver resection: The French Hepato-bilio-pancreatic Surgery Association statement,” Journal of Visceral Surgery 152, 107–112 (2015). https://doi.org/10.1016/j.jviscsurg.2015.02.003

[3] Ciria, R., Cherqui, D., Geller, D. A., Briceno, J., and Wakabayashi, G., “Comparative short-term benefits of laparoscopic liver resection,” Annals of Surgery 263(4), 761–777 (2016). https://doi.org/10.1097/SLA.0000000000001413

[4] Meola, A., Cutolo, F., Carbone, M., Cagnazzo, F., Ferrari, M., and Ferrari, V., “Augmented reality in neurosurgery: a systematic review,” Neurosurgical Review, 1–12, (2016). https://doi.org/10.1007/s10143-016-0732-9

[5] Kang, X., Azizian, M., Wilson, E., Wu, K., Martin, A. D., Kane, T. D., Peters, C. A., Cleary, K., and Shekhar, R., “Stereoscopic augmented reality for laparoscopic surgery,” Surgical Endoscopy 28(7), 2227–2235 (2014). https://doi.org/10.1007/s00464-014-3433-x

[6] Thompson, S., Totz, J., Song, Y., Johnsen, S., Stoyanov, D., Ourselin, S., Gurusamy, K., Schneider, C., Davidson, B., Hawkes, D., and et al., “Accuracy validation of an image guided laparoscopy system for liver resection,” Proc. SPIE Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling 941509-12 (2015). https://doi.org/10.1117/12.2080974

[7] Haouchine, N. and Cotin, S., “Segmentation and labelling of intra-operative laparoscopic images using structure from point cloud,” Proc. IEEE International Symposium on Biomedical Imaging, 115–118 (2016). https://doi.org/10.1109/ISBI.2016.7493224

[8] Clarkson, M. J., Zombori, G., Thompson, S., Totz, J., Song, Y., Espak, M., Johnsen, S., Hawkes, D., and Ourselin, S., “The NifTK software platform for image-guided interventions: platform overview and NiftyLink messaging,” International Journal of Computer Assisted Radiology and Surgery 10, 301–316 (2014). https://doi.org/10.1007/s11548-014-1124-7

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

[10] He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,” arXiv (2015). arXiv:1512.03385v1

[11] He, K., Zhang, X., Ren, S., and Sun, J., “Identity mappings in deep residual networks,” arXiv (2016). arXiv:1603.05027v2

[12] Mountney, P. and Yang, G.-Z., “Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping,” Proc. IEEE Engineering in Medicine and Biology Society, 1184–1187 (2009). https://doi.org/10.1109/IEMBS.2009.5333939

[13] Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., and Torr, P. H. S., “Fully-convolutional siamese networks for object tracking,” arXiv (2016). arXiv:1606.09549

[14] Bertinetto, L., Henriques, J. F., Valmadre, J., Torr, P. H. S., and Vedaldi, A., “Learning feed-forward one-shot learners,” arXiv (2016). arXiv:1606.05233