Deep learning for instrument detection and assessment of operative skill in surgical videos

Kyle Lam[1], Yujian An[1], Ara Darzi[1], Benny Lo[1], James M. Kinross[1], Sanjay Purkayastha[1]
[1]Department of Surgery and Cancer, Imperial College, London, W2 1NY, UK

Introduction

Surgical performance is directly related to patient outcomes[1]. There is significant variation in surgical performance and therefore a need to measure operative skill accurately and reliably. Despite this, current means of surgical performance assessment rely on expert observation which is labor-intensive, prone to rater bias and unreliable. Deep learning has the potential to process large amounts of surgical data and present rapid, objective, and automatic operative skill assessment.

We present a deep learning based approach to surgical performance assessment through the tracking of instruments in endoscopic video.

Methods

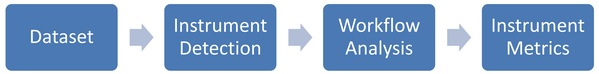


Figure 1. Framework of approach

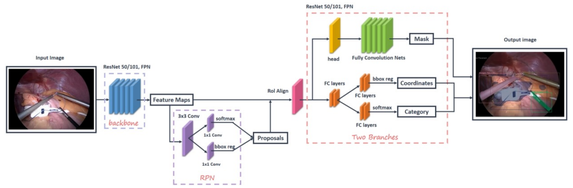


Figure 2. Architecture of object detection algorithm, Mask R-CNN

Dataset: Manual annotation of the spatial bounds of 2600 images from videos of laparoscopic gastric band insertion. Dataset split into 2100 images for training and 500 images for validation.

Instrument Detection: A deep learning instrument detection framework was based on Mask R-CNN[2], a state-of-the-art object detection algorithm. Videos or images are Inputted and the framework outputs a mask around the detected instrument, the classification of the object, and the coordinates of a bounding box.

Workflow Analysis*:* A Markov chain, based on the results of the instrument detection model, was modelled to divide the procedure into four stages: port insertion; formation of the retrogastric tunnel; gastric band placement; and port removal.

Instrument Metrics: 5 assessment indicators based on instrument detection results were designed: time spent in stage; instrument time; moving distance; smoothness of movement; and concentration of movement.

Results

A mean average precision of 83.9% for correct detection was achieved demonstrating excellent mask coverage ability. Classification recall for the grasper, gastric band and hook were 99%, 96% and 78% with precision of 82%, 91% and 100% respectively (Table 1).

Tool	Recall	Precision
Grasper	99.03%	82.14%
Hook	77.67%	99.79%
Retractor	57.94%	100.00%
Gastric Band	95.68%	91.00%
Curved Tips	64.50%	98.79%
Needle Holder	91.59%	99.42%
Scissors	50.57%	100.00%
Curved Grasper	38.60%	78.57%

Table 1. Instrument classification results

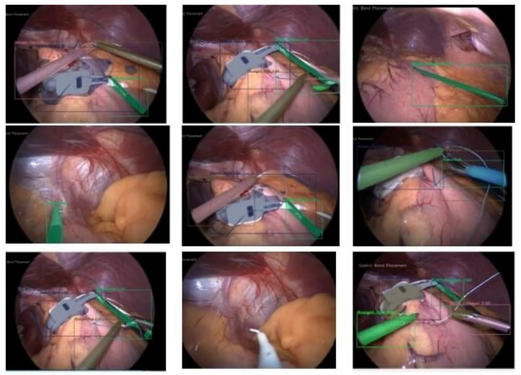


Figure 3. Example images of instrument detection results. Rows 1-2 show correct instrument detection results. Row 3 shows incorrect instrument detection results

Preliminary results of instrument metrics were able to differentiate expert and trainee surgeons. The trainee had a larger distance travelled for the grasper and hook in addition to a larger distance-time ratio (DTR) translating to inferior instrument movement efficiency. (Table 2).

Metric	Expert	Trainee
DTR of Grasper	8.98	9.12
DTR of Curved Tips	13.88	13.94
DTR of Needle Holder	11.54	13.83
Smoothness of Grasper	41.53	42.18
Smoothness of Curved Tips	43.28	44.47
Smoothness of Needle Holder	41.52	42.21
Smoothness of Scissors	27.03	52.06
Concentration of Curved Tips	137.73	158.49

Table 2. Instrument metrics for Stage 3 (Insertion of gastric band)

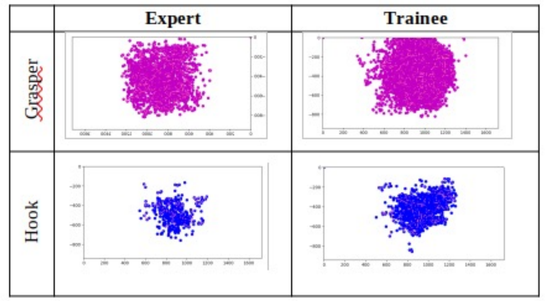


Figure 4. Trajectory map of instruments in Stage 2 (Formation of retrogastric tunnel)

Conclusions

We present a novel annotated dataset of 2600 surgical images, demonstrate a deep learning framework for instrument detection, and leverage this to generate instrument metrics for performance assessment. Future work will focus on further development and validation of these metrics.

Contact
Kyle Lam
k.lam@imperial.ac.uk



REFERENCES
1. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med. 2013;369(15):1434-1442.
2. K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," arXiv e-prints, p. arXiv:1703.06870, Mar 2017.