Workflow Recognition in Cholecystectomy Videos

Suraj Menon, Ruben Mayer, Wei Cheung, Hossein Mohamadipanah, PhD[a]

Introduction

Workflow recognition in surgery videos is a new field that is aiming to help healthcare providers become more rigorous in improving surgical procedures. Our work focuses on classifying phases from Cholecystectomy (gallbladder extractions) procedures. The goal is to be able to output a frame by frame label of the phase of the surgery from the video feed.

Methods

Conv + LSTM

2-D convolutions with a LSTM placed at the end for sequence modeling.

I3D

A DeepMind network that uses 3-D convolutions over a sequence of images to do temporal modeling.

I3D + LSTM

Using I3D for visual modelling and short-term temporal modeling, and an LSTM for longer term temporal modeling.

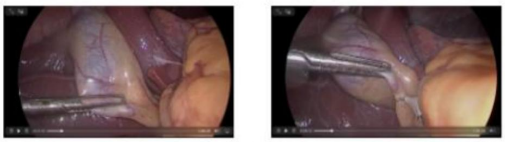


Figure 2: Left: 'Preparation' phase mislabelled as 'CalotTriangleDissection'— Right: 'Correct labelling of 'CalotTriangleDissection'

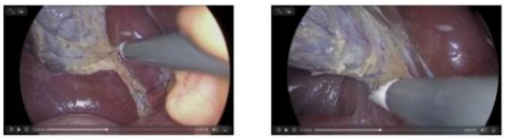


Figure 3: Images with similar visual features confuse the network due to lack of inter-phase difference

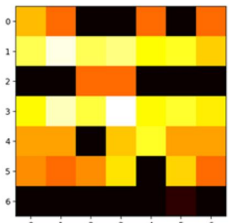
Results

The Inflated 3D CNN reached 59% accuracy. This is lower than the state of the art, but on par with other research on the same data set.

Our Inflated 3D CNN + LSTM underperformed the baseline

Method	Accuracy	F1-Score
2D CNN + LSTM	45.6%	39.8%
I 3D CNN	59.1%	54.2%
I 3D CNN + LSTM	34.5	18.1

Table 1: Performance of the three methods



Confusion Matrix for I 3D CNN

Conclusions

From these results we believe that the I3D network is a viable solution for workflow analysis for Cholecystectomy surgeries and potentially others. The accuracy achieved shows promise that if trained with more unfrozen layers and for a longer period of time on the more class balanced dataset that I3D could serve as a good base for the visual analysis of this dataset.

Future Directions

Apart from improving the I3D implementation, other work we would like to continue is to find a better method of using a sequence model to tie features together one from 10s clip to another. We are hoping that a technique like this may be able to better capture the transitions from phase to phase.