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.

13D

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

13D + LSTM

Using 13D for visual modelling and short-term temporal modeling, and an LSTM for longer term temporal modeling.

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 dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D CNN + LSTM</td>
<td>45.5%</td>
<td>39.3%</td>
</tr>
<tr>
<td>13D CNN</td>
<td>59.1%</td>
<td>54.2%</td>
</tr>
<tr>
<td>13D CNN + LSTM</td>
<td>34.5%</td>
<td>18.1%</td>
</tr>
</tbody>
</table>

Table 1: Performance of the three methods

Confusion Matrix for 13D CNN

Conclusions

From these results we believe that the 13D 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 the more class balanced dataset that 13D could serve as a good base for the visual analysis of this dataset.

Future Directions

Apart from improving the 13D 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.