# Content Recognition in Surgical Videos

Madhusudan Hegde
mrhegde@stanford.edu
https://youtu.be/4nySlL_CubU

Nithin Akkati
nakkati@stanford.edu
https://youtu.be/N6zkqqxJ3mg

Senthil Vel Gunasekaran
sguna@stanford.edu
https://youtu.be/x6o51Kbj_Q4

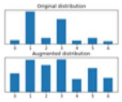
## Overview

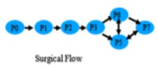
- Advances in Computer Vision promise great improvements in surgical work-flow detection.
- We built a CNN + LSTM model to to detect work flow from surgical videos.
- Prior knowledge of surgical flow applied to achieve accuracy of 88%.
- We built tool detection model to evaluate tool importance in phase detection

## Data

- 80 videos of cholecystectomy are split into 50/25/25 train/dev/test ratio
- Video converted to 224 x 224 RGB images at 5 FPS
- The labeled data was unbalanced, and data augmentation was used
- Surgical flow shows Markov chain as below and used for optimization

| ID | Surgical Phase | Duration(min) |
|---|---|---|
| P0 | Preparation | 2.2 |
| P1 | Calot Dissection | 10.4 |
| P2 | Clipping and Cutting | 2.8 |
| P3 | Gallbladder Dissection | 7.5 |
| P4 | Gallbladder Packaging | 1.8 |
| P5 | Cleaning Coagulation | 1.9 |
| P6 | Gallbladder Retraction | 4.5 |




Surgical Flow

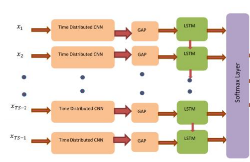
- Time step of 25 used for LSTM processing.
- The images showed tool rotations for the same activity (below) and random horizontal flips added to augment the data
- Tool detection used 1250 images with bounding boxes for train set. The train/dev/test ratio of 70/15/15 is used.

## CNN+LSTM Training

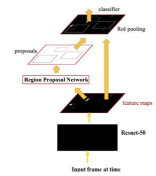
| Training Method | Advantages | Disadvantages |
|---|---|---|
| **Train CNN and LSTM Separately** | • Hyper-parameters can be chosen independently as separate loss functions are used<br>• Computationally less expensive | • Sub-optimal due to significant intra-class variance and limited interclass variance of visual features |
| **End to End CNN + LSTM training** | • The LSTM error is backpropagated to CNN and CNN training is helped by sequence discrimination of LSTM | • Restriction on hyper parameter selection as both the models share same optimizer/loss function in Keras |
| **Stateful LSTM** | • Captures correlations across time steps<br>• Useful for small batches as Keras resets the LSTM internal states every batch size | • Care must be taken to maintain temporal information in the input data (no shuffling). Otherwise will degrade the performance |

## Model

- Time Distributed CNN with Global Average Pool Layer (GAP) feeds LSTM
- LSTM layer of size 2048 with a Dense layer followed by Softmax
- Nadam optimizer with learning rate of 0.00001 and rate decay of 0.004
- Batchsize = 8, Dropouts (0.5) and L2 regularization (0.01) used for LSTM




Phase detection

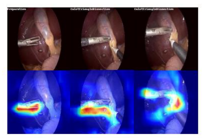


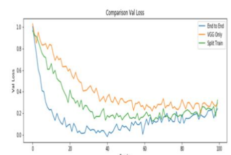

Tool detection

- Model takes frames with bounding boxes as input. Outputs tool label, confidence and localization information.
- Faster R-CNN has 2 networks, region proposal network(RPN) for generating proposals and detector. We use Resnet50 as our CNN.
- We tuned learning rate, bbox threshold, type of CNN's to get to 80% accuracy.

## Error Analysis

- Activation map used to understand mis-detections of CNN output
- Loss function used to analyze overall convergence and performance
- Early stopping is used to prevent overfitting

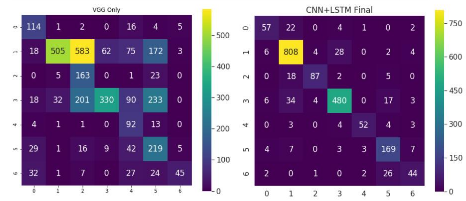





- Tool detection model got accuracy of 80%, on-par with state-of-the-art model. Hand tuned weights and tool predictions are combined with phase detector model.
- This approach showed improvement in F1 scores.
- Frames with gas were mis-classified as bag in many examples

## Results

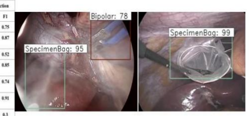
| CNN+LSTM Final | Precision | recall | f1-score | support |
|---|---|---|---|---|
| (0) Preparation | 0.76 | 0.66 | 0.71 | 86 |
| (1) CalotTriangleDissection | 0.91 | 0.95 | 0.93 | 852 |
| (2) ClippingCutting | 0.91 | 0.78 | 0.84 | 112 |
| (3) GallbladderDissection | 0.92 | 0.88 | 0.9 | 544 |
| (4) GallbladderPackaging | 0.9 | 0.79 | 0.84 | 66 |
| (5) CleaningCoagulation | 0.76 | 0.88 | 0.81 | 193 |
| (6) GallbladderRetraction | 0.7 | 0.59 | 0.64 | 75 |
| Overall | 0.88 | 0.88 | 0.88 | 1,928 |




Confusion matrix for model predictions.

## Tool Detection Results

| | Baseline CNN+LSTM | | | | Baseline +Tool Detection | | |
|---|---|---|---|---|---|---|---|
| | Count | Precision | Recall | F1 | Precision | Recall | F1 |
| Preparation | 128 | 0.90 | 0.59 | 0.72 | 0.66 | 0.88 | 0.75 |
| Calot Training Dissection | 500 | 0.92 | 0.82 | 0.87 | 0.94 | 0.82 | 0.87 |
| Clipping Cutting | 68 | 0.54 | 0.44 | 0.48 | 0.62 | 0.44 | 0.52 |
| Gallblader Dissection | 370 | 0.77 | 0.93 | 0.85 | 0.79 | 0.93 | 0.85 |
| Gallblader packaging | 80 | 0.62 | 0.86 | 0.72 | 0.83 | 0.67 | 0.74 |
| Cleaning Coagulation | 58 | 0.87 | 0.96 | 0.91 | 0.89 | 0.93 | 0.91 |
| Gallblader Retraction | 25 | 0.40 | 0.40 | 0.40 | 0.33 | 0.30 | 0.3 |
| Overall | | | | 0.81 | | | 0.82 |




## Conclusion/Future Work

- End to end training of CNN+LSTM provides best results of 88% accuracy
- Prediction accuracy can be improved with prior knowledge of surgical flow
- We achieved 80% accuracy in tool detection and this information can help surgical phase detection.
- We plan to explore stacked LSTM and stateless LSTM to improve the accuracy
- Use tool detection to improve the confidence of surgical phase detection

## Reference

[1] Y. Jin et al.,"SV-RCNet: Workflow Recognition from Surgical Videos Using Recurrent Convolutional Network," in IEEE Transactions on Medical Imaging, vol. 37, no. 5, pp. 11141126, May 2018.
[2] O. Dergachyova, D. Bouget, A. Huaulmé, X. Morandi, and P. Jannin, "Automatic data-driven real-time segmentation and recognition of surgical workflow," Int. J. Comput. Assist. Radiol. Surgery, vol. 11, no. 6,
[14] Amy Jin, Serena Yeung, Jeffrey Jopling, Jonathan Krause, Dan Azagury, Arnold Milstein, and Li Fei-Fei(2017). Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks Stanford University.