Capsulorrhexis Trajectories for Automated Surgical Training Feedback

Ben Ehlert, David Kuo, Ben Viggiano Department of Biomedical Data Science, Stanford University

Introduction

There is great demand among surgical trainees for any resource that can help maximize the educational value of each surgical case given the inherent variability in volume and quality of surgical cases during training.

Machine-learning based computer vision methods have been applied to evaluate surgical skill in laparoscopic and open surgeries from surgical videos.

We extend these methods to the capsulorrhexis step of cataract surgery and develop a keypoint detection model tracking instrument tips in order to generate instrument trajectories for surgical feedback.

Problem Statement

To develop a keypoint detection model for frame by frame detection of utrada forceps instrument tips during the capsulorrhexis step of cataract surgery.

Inputs: surgical video frames in JPEG format.
Outputs: bounding box and keypoint predictions for the utrada forceps and utrada forceps tips.
Evaluation metrics: mean average precision (mAP) of keypoint detection on the validation set

Methods

We fine-tune Keypoint R-CNN-FPN with backbones of different architectures and depths (ResNet-50-FPN, ResNet-101-FPN, ResNeXt-101-FPN) pretrained on the COCO dataset, and modified for our task

Keypoint R-CNN is a modified Mask R-CNN that detects keypoints by treating each keypoint as a separate one-hot binary mask with only 1 pixel labeled.

A Feature Pyramid Network (FPN) generates feature maps of different spatial resolutions to improve region proposal, and detection of objects of multiple scales.

Experiments

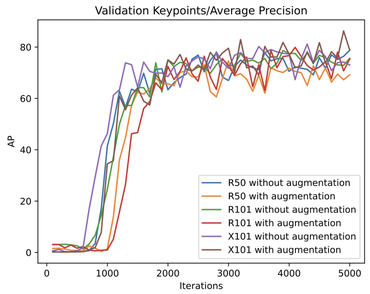


FIGURE 1: Data augmentation improves model performance

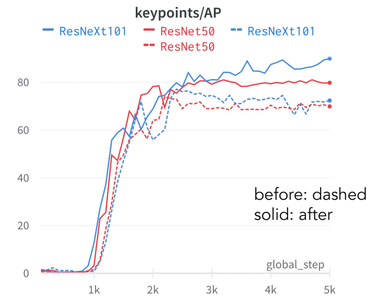


FIGURE 2: bbox relaxation improves model performance

Dataset Size	bbox mAP	keypoint mAP
400 images	70.83	77.41
1000 images	79.21	80.96

TABLE 1: Increasing data set size improves performance

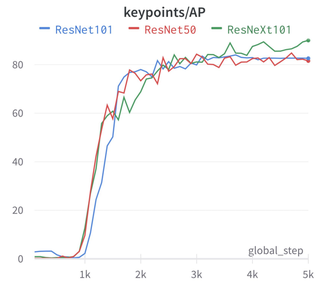


FIGURE 3: ResNeXt101 outperforms ResNet50/101

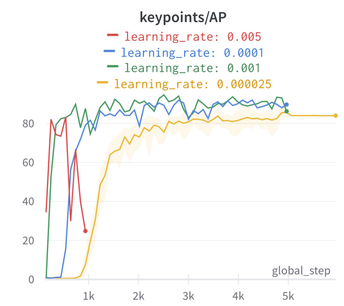
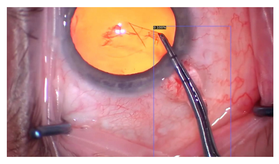
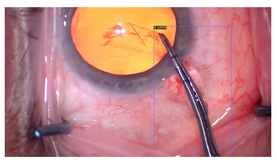
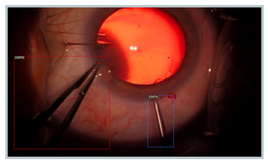


FIGURE 4: LR 0.001 was the best learning rate

Best Model

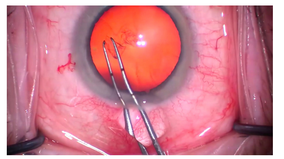
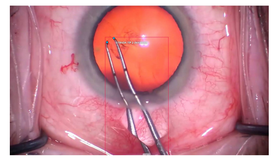
Keypoint R-CNN-FPN with ResNeXt101 backbone trained with data augmentation and LR 0.001 on 1000 images, achieved 94.65 keypoints/mAP and 80.388 bbox/mAP.

Dataset

Two ophthalmologists curated 200 cataract surgery videos performed by junior resident, senior resident, and attending surgeons

1000 random video frames from 40 random videos were annotated with bounding boxes and keypoints and divided 800/100/100 into training, validation, and test sets.

Conclusions

We successfully trained a keypoint detection model to detect utrada forceps tips in cataract surgery videos with high mean average precision and qualitatively accurate keypoint predictions.

Future Work

1. Extend our dataset to more instruments and anatomic landmarks for better feedback.
2. Explore weakly-supervised or self-supervised methods for our unlabeled data.
3. Leverage the temporal component of our videos with optical flow, 3D CNNs, RNNs or transformers

Acknowledgements

1. Riya Sinha (CS229 student): for contributions to labeling and running experiments for hyperparameter tuning
2. Susan Qi, MD, MS (ophthalmologist): for contributions in curating and annotating data
3. Zhuoyi Huang (TA mentor): for feedback and guidance throughout the project

https://github.com/davidekuo/rhexis-trajectory