Tool Annotation of Cataract Surgery Videos Using Augmented SqueezeNet (AUGSQZNT)

Chandan Panda

{Chandanpanda2006@yahoo.com}

Epsilon

1. Introduction

Data privacy laws impose constraints on usage of technological advances in Healthcare. For example, most hospitals would require extensive security protocols to be in place before they share medical images or video data outside of the organization.

While many advances have been made in usage of convolutional networks to process images, state of the art deep learning networks tend to have complex system architectures in order to maximize accuracy. Consequently, they require specialized computer systems in terms of memory and hardware (e.g. GPU) thus limiting applications to either a cloud based solution or to organization which currently have such systems in place.

This research work is focused on using the SQUEEZENET [1] network architecture to achieve near top leaderboard accuracy but with a smaller memory requirement

2. Preprocessing

The mp4 files were processed by resizing to 432 X 768 resolution and extracting random patches of size 360 X 640 resolution at 10 frames per second. The frames were normalized with mean RGB values and passed through a SQUEEZENET network excluding the final 6 layers, initialized with image net weights.

3. Training

A CNN network with a split architecture was used to train the model on the features extracted during preprocessing. The network consisted of 3 layers of convolutional blocks, each with three convolutional layers with 372 filters. At the end, the network split into 3 parts: one part for the Cannula set of labels, one part for the Forceps set and one part for the rest. The Forceps split of the network used a softmax activation while the other two splits of the network used sigmoid. The loss function was weighted by a 80:10:10 ratio and the binary cross entropy loss function was used. Adam optimizer was used with a learning schedule starting with the learning rate of 0.01 and subsequently reducing by multiplying with 0.1 after every 3 epochs with no improvement in validation loss.

4. Post Processing

5-fold test time augmentation was performed by taking the center, top left, top right, bottom right and bottom left patches from each frame in the test dataset. The predictions were averaged across the 5 patches for each frame.

5. References

[1] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Kurt Keutzer “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size”