Leveraging Surgical Workflow Recognition for Skill Assessment in Simulated Cataract Surgery

Bining Long[1], Rebecca Hisey[1], Joshua Bierbrier[1], Christine Law[1], Gabor Fichtinger[1], Matthew Holden[2]
[1]*Queen's University, Kingston, Ontario, Canada* [2]*Carleton University, Ottawa, Ontario, Canada*

Introduction. Cataracts are a major cause of blindness affecting millions of individuals worldwide. Cataract surgery, which involves replacing the eye's cloudy lens with an artificial one to restore vision, is a precise procedure requiring skilled training. Automated surgical workflow recognition analyzes recorded surgical videos using deep learning to classify the major procedural steps. This approach has the potential to improve surgical training and evaluation by providing objective metrics that could facilitate task-specific skill assessment. Accurate workflow recognition relies on capturing spatial features and modeling temporal progression. While temporal convolutional networks (TCNs) have shown promise in video-based action recognition [1], their application to microscope video data, particularly in cataract surgery, remains underexplored. In this work, we deploy TCNs to recognize surgical workflow and demonstrate the potential of using the measured task durations derived from the recognized workflow as a measure for assessing surgical skill in cataract surgery.

Methods. A dataset of 25 monocular microscope videos of cataract surgeries on a Philips studio artificial eye was collected from Kingston Health Sciences Centre. The dataset includes videos of surgeries performed by one ophthalmologist and four residents, each completing five trials of four surgical steps: corneal entry, paracentesis and viscoelastic insertion, and two capsulorrhexis sub-steps: commencement of the flap and follow-through (Capsulorrhexis-A) and formation and circular completion (Capsulorrhexis-B). Videos range from 126 to 340 seconds (average 180 seconds) at 4 frames per second. Ground truths for surgical tasks were annotated by an ophthalmologist through video review. We employed a CNN-TCN approach for surgical workflow recognition. We extracted spatial features using a ResNet18 pretrained on ImageNet and fine-tuned for frame-wise tasks, and modeled long-term temporal dependencies using a TCN with dilated convolutions. Model performance was evaluated through five-fold leave-one-user-out cross-validation, with one participant for testing, one for validation, and the remaining three for training. We measured the model's performance using accuracy, defined as the proportion of video frames with correctly predicted tasks relative to the total frames. From the tasks recognized by the model, we calculated the time spent by the ophthalmologist and residents on each task for skill assessment. This was determined as the average number of predicted frames across their five videos, divided by the frame rate.

Results. The CNN-TCN model achieved a workflow recognition accuracy of 72.35%. Figure 1 presents whisker plots comparing predicted task duration variations between the ophthalmologist and residents, with ground truth durations as a reference. We observe that differences in predicted task durations generally align with ground truth patterns: residents consistently take less time than the ophthalmologist for corneal entry but more time for both capsulorrhexis sub-steps, as well as during periods of inactivity.

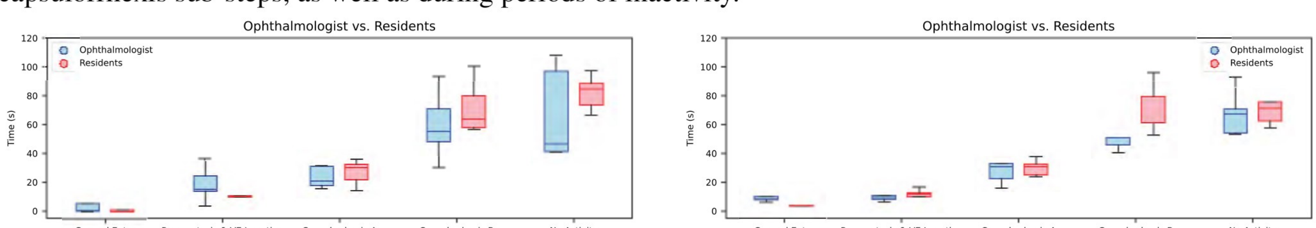

Figure 1: Whisker plots of the predicted (left) and ground truth (right) task durations for each surgical task.

Conclusions. In this work, we present a CNN-TCN approach for video-based surgical workflow recognition in simulated cataract surgery, and extend its application to calculate task durations for surgical skill assessment. While the model's task recognition performance leaves room for improvement, it was able to distinguish differences in task durations between the ophthalmologist and residents. Despite challenges posed by the small dataset size and the imbalance in surgeon expertise, with only one ophthalmologist included, this work demonstrates the potential of leveraging recognized workflow to provide objective, task-specific measures for surgical skill assessment.

References

[1] T. Czempiel, M. Paschali, M. Keicher, W. Simson, H. Feussner, S. T. Kim, and N. Navab, "Tecno: Surgical phase recognition with multi-stage temporal convolutional networks," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 343–352, Springer, 2020.