BIOMEDICAL ENGINEERING

Surgical Action Recognition Using TD-CNN-LSTM

Team Members: Nanthini Narayanan, Sam Capocyan
Instructor: Dr. Adam Charles | TA : Jayanta Dey
Department of Biomedical Engineering | Johns Hopkins University | Whiting School of Engineering | Baltimore, MD

JOHNS HOPKINS WHITING SCHOOL of ENGINEERING

Introduction

Action recognition is one of the most essential and increasingly explored task in computer vision. The development of a precise action recognition method on a surgical dataset is particularly challenging but beneficial since it could contribute to the guidance of a surgical robot and surgical education.[2]

We propose a Time Distributed-Convolutional Neural Network-Long Short-Term Memory (TD-CNN-LSTM) model that improves upon the Endo3D CNN-LSTM baseline model via a time-distributed framework.[1] Our method is end-to-end and has the potential to outperform state-of-the-art models.

Objectives

- Classification of 8 different actions from robot-assisted radical prostatectomies
- Outperform current models used for action detection during the EndoVis MICCAI 2022 Challenge

Method

The training set of SAR-RARP50 contains 40 continuously annotated video segments which were cut into video clips of 16 frames each.[3] The dataset is highly imbalanced as can be observed in the figure below.

Action ID	Action
0	Other
1	Picking up the needle
2	Positioning the needle tip
3	Pushing the needle through the tissue
4	Pulling the needle out of the tissue
5	Tying a knot
6	Cutting the suture
7	Returning/dropping the needle

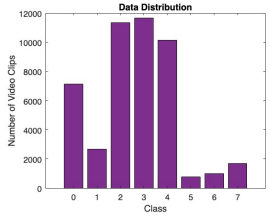
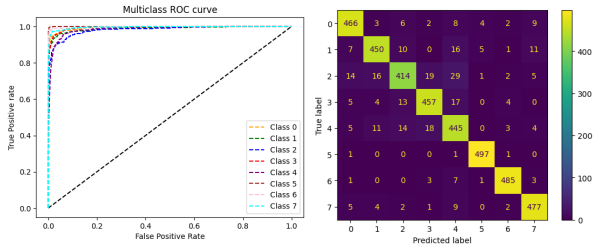


Figure 1—Data labels and distribution

To investigate the effect of imbalanced dataset on the model's performance, 8-class classification was implemented with scaled down imbalanced and balanced datasets.

5-fold cross-validation was implemented to evaluate the generalizability of the TD-CNN-LSTM model.

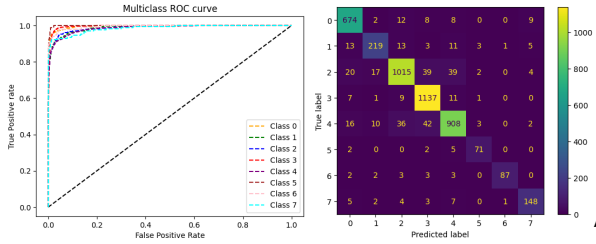
Results



Class	F1 Score
0	93%
1	90%
2	83%
3	91%
4	89%
5	99%
6	97%
7	95%

Accuracy: 92.32%

Figure 2—8-Class Balanced Results



Class	F1 Score
0	93%
1	84%
2	91%
3	95%
4	90%
5	89%
6	92%
7	88%

Accuracy: 91.70%

Figure 3—8-Class Imbalanced Results

Model Architecture

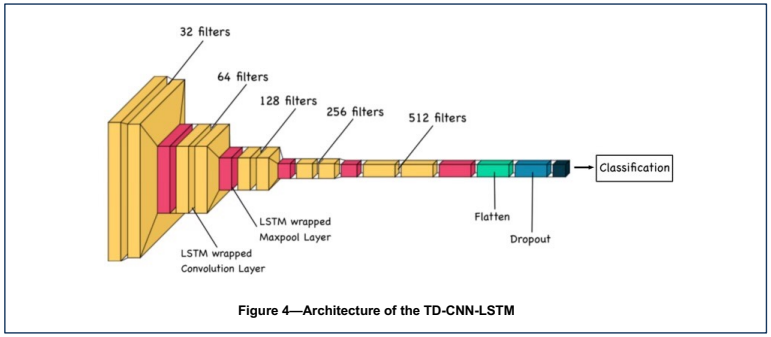


Figure 4—Architecture of the TD-CNN-LSTM

Discussion

Our proposed model captures fine-level and higher-level temporal information with the CNN-LSTM architecture and is an end-to-end approach.

The preliminary results are highly promising. While the accuracy of the model with a balanced data subset is slightly higher than with the imbalanced data subset, the overall model performance is good in both cases. The confusion matrices, ROC curves, and high metrics indicate the same. Thus, the class imbalance does not seem to affect the performance of the model significantly and moving forward should not be a major pitfall when training the model with the entire dataset. Based on the projection of our preliminary results, our proposed model should be effective for surgical activity recognition with performance comparable to other benchmark models when the entire dataset is used.

Next Steps

- Train the model with the entire training set and evaluate its performance
- Compare performance of the TD-CNN-LSTM model with 3D-CNN on the SAR-RARP50 dataset to demonstrate the effectiveness of our approach

References

[1]Chen, Weixiang, et al. "Endo3d: online workflow analysis for endoscopic surgeries based on 3d cnn and lstm." OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis: First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5. Springer International Publishing, 2018.

[2]Bawa, Vivek Singh, et al. "The SARAS endoscopic surgeon action detection (ESAD) dataset: challenges and methods." arXiv preprint arXiv:2104.03178 (2021).

[3]The SAR-RARP50 paper has not been published yet. The challenge description is available at: https://www.synapse.org/Synapse:syn27618412/wiki/