Virtual ACS 2021 Surgeons and Engineers: A Dialogue on Surgical Simulation Meeting

Research In-Progress

Automated Surgical Video Analysis Using Multi-Task Learning

Babak Namazi, PhD; Ganesh Sankaranarayanan, PhD; and Venkat Devarajan, PhD.

University of Texas at Arlington, Arlington, TX; Baylor University Medical Center, Dallas, TX

Introduction: Recorded videos from laparoscopic procedures contain valuable information, which can be extremely useful in surgical education and training. To efficiently utilize these videos, important features such as the usage of surgical tools and different phases need to be extracted, which can be cumbersome to do manually and hence an automated system needs to be developed.

Methods: With the advent of deep learning, such tasks can be accomplished by training deep convolutional and recurrent neural networks (CNNs and RNNs) to learn the spatial and temporal visual features. We designed two Recurrent Convolutional Neural Networks (RCNNs) to identify the appearance of different surgical tool combinations and, the current phase of each frame of a laparoscopic video using the knowledge of five previous frames.

Preliminary Results: We tested our models on a dataset that contains 80 videos from laparoscopic cholecystectomy. We obtained frame-level accuracy of 79.97% for tool presence detection and 85.2% for surgical phase identification by separately training the RCNNs and further improved the performance by training RNNs at the post-processing step to 91.91% and 92.5% respectively.

Next Steps: The high correlation between the surgical tasks and the tools suggests an excellent potential of jointly training a deep learning model to simultaneously perform the two tasks. Our next step is to train our model with multiple tasks such as detecting the presence, location, and pose of surgical tools, surgical tasks or events, and anatomical structures in each frame of a surgical video. In order to accomplish this objective, the visual feature extractor will be shared among the various tasks, and multiple outputs will be taken to minimize multiple loss functions in the optimization algorithm. We believe using this technique can boost the performance of each task, while using a single deep learning system.