Deep Learning-Based Automated Tool Detection and Analysis of Surgical Videos to Assess Operative Skill

Jin, Amy

Annually, 7 million patients suffer surgical complications, many of which are linked to inadequate surgical training and a lack of individualized feedback. To improve operative skill, it is essential to review surgeon performance. However, this manual process is subjective and time consuming. Thus, we created a deep learning model to detect and track surgical instruments and to analyze their movements in real-time, enabling automated assessment of surgical performance. Since this study was the first to perform this task, we had to assemble our own dataset. We labeled 2200 video frames across 15 cholecystectomy videos with the coordinates of spatial bounding boxes around instruments. This dataset has been publicly released. We then applied it to train a deep learning model to perform tool detection and localization, leveraging region-based convolutional neural networks. Input video frames were passed through the VGG-16 convolutional neural network and then through a region proposal network, and the model outputted the coordinates of bounding boxes around surgical instruments. In comparison with state-of-the-art approaches for automated tool detection, we outperformed existing methods by 23%, improving mean average precision from 63.7 to 78.2, and achieved a real-time frame rate of 5 fps. With the model's output, we assessed surgeon performance by extracting key metrics that reflect surgical skill level, such as tool usage patterns, motion economy, and path length. Surgeons from Stanford Health Care independently reviewed the videos and validated our findings. This system paves the way for AI technologies that assist and train surgeons by monitoring procedures and by pinpointing improper techniques to improve surgical outcomes.

Awards Won:
Second Award of $2,000
Samvid Education Foundation: Geno Award of $1000 honoring the literary work of Tamil novelist, Sujatha.