Deep Learning Based Surgical Instrument Detection in a Synthetic Endonasal Surgical Simulator

Anthony M Asher, MD[1]; Kaan Duman, PhD[2]; Margaux Masson-Forsythe, MS[2]; Kyle K Van Koevering, MD[3]; Daniel Prevedello, MD[3]; Daniel Donoho, MD[2]; [1]Barrow Neurological Institute; [2]Surgical Data Science Collective; [3]The Ohio State University

Introduction

Computer vision models have demonstrated the ability to identify and label surgical instruments in live and cadaveric endonasal procedures, but these techniques remain underexplored in synthetic surgical simulators. This study aims to develop a dataset and train instrument detection models for synthetic surgical simulators using video recorded at AANS 2024 in Chicago.

Methods and Materials

At AANS 2024, surgeons and trainees were tasked with controlling a simulated carotid artery injury using the UpSurgeOn transnasal surgery model. Surgical video was recorded at 30fps with a Karl Storz endoscope. Surgical Instruments including suction, pituitary grasper, and cotton patties were manually labeled with bounding boxes frame by frame using Encord. Frames were divided into training, validation, and test sets using two methods:

Random split: Frames were randomly assigned (80/10/10) irrespective of individual video.

Video split: One video was reserved for testing, and frames from the remaining five were split into training/validation (90/10).

The YOLOv8s deep-learning framework was used to train two separate instrument detection models.

Results

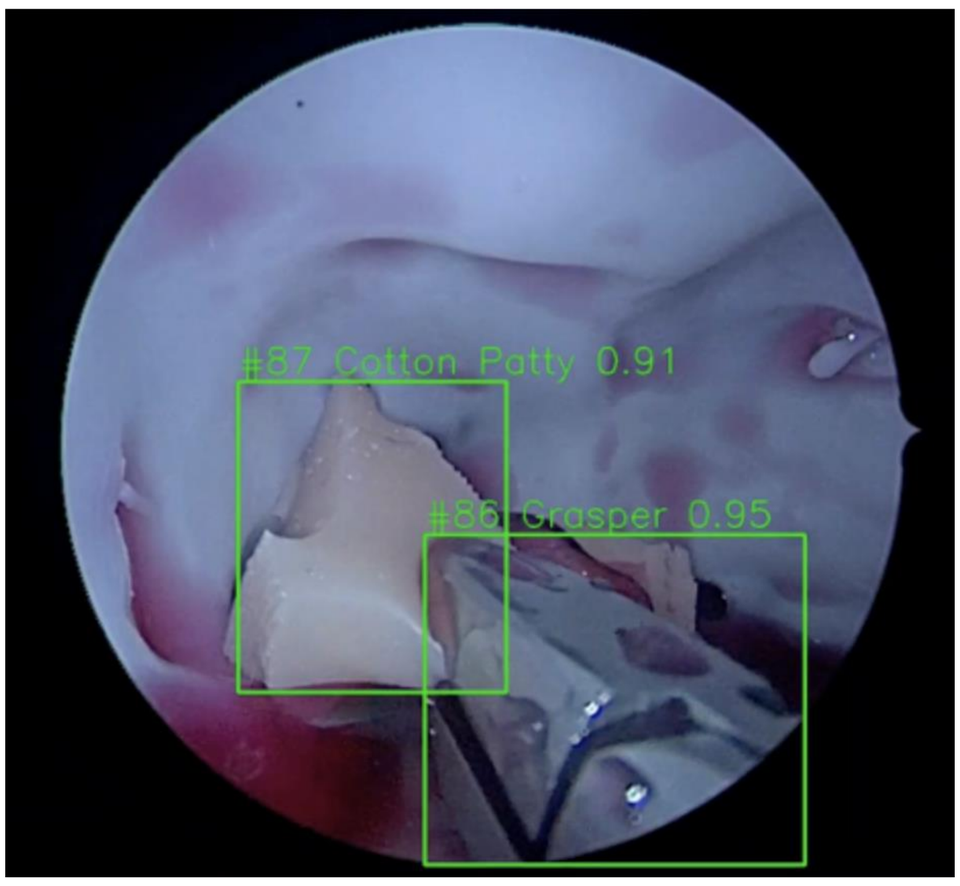
Six carotid artery injury videos, totaling 6994 frames, were labeled. The pituitary grasper, suction, and cotton patties were labeled 1440, 6623, and 3286 times, respectively. Performance metrics for the random and video split models, respectively, are as follow: weighted precision 0.987 vs 0.925, recall 0.984 vs 0.924, F1-score 0.986 vs 0.925, mAP50 0.988 vs 0.954, mAP50-95 0.941 vs 0.847.

Conclusions

Surgical video obtained using a synthetic carotid artery injury simulator can be successfully used to train and validate a deep learning-based instrument detection model. The random split model performed significantly better across all metrics, but its superior performance is likely due to overfitting. The video split model, which better represents real-world deployment, still performed at a high level. Ultimately, synthetic surgical simulation platforms provide a robust arena for the development of AI models for instrument detection and surgical education more broadly.

Metric	Random Split	Video Split
Precision	0.800	0.790
Recall	0.356	0.856
F1-score	0.228	0.134
mAP50	0.954	0.875
mAP50-95	0.324	0.325

Table 1. Performance metrics of 3-class instrument detection models trained on random split and video split datasets.



Contact

Anthony M. Asher, MD
Barrow Neurological Institute, Department of Neurosurgery
Anthony.asher@commonspirit.org