Development of a Computer Vision System for Surgical Instrument Analysis During Endoscopic Sinus and Skull Base Surgery

Corinne Stonebraker BA[1], Jaeho Cho[2], Katherine Liu MD[1], Lacy Brame DO[1], Raj Shrivastava MD[3], Alfred-Marc Iloreta MD[1]

1. Department of Otolaryngology–Head and Neck Surgery, Icahn School of Medicine at Mount Sinai
2. Albert Nerken School of Engineering, The Cooper Union
3. Department of Neurosurgery, Icahn School of Medicine at Mount Sinai

BACKGROUND

- Real-time instrument tracking can improve surgical workflow optimization and training[1]
- Manual video review is time-intensive, difficult to scale, and prone to inter-rater variability[2]
- Recent advances in deep learning-based computer vision, particularly object detection and image segmentation, enable automated, high-fidelity analysis of surgical video[3]
- Affordable, high-resolution action cameras allow unobtrusive capture of operative video, creating a practical platform for widespread implementation in the operating room[4]

METHODS

- **Data Collection**
  - 4 endoscopic sinus/skull base surgeries recorded
  - Insta360 GO camera (4K, 30 fps) placed on surface across operating table from surgeon
  - Total footage: 6h 21min 40s
- **Dataset**
  - Frames sampled every 12 seconds
  - 2,159 images: training (1515), validation (322), testing (322)
- **Annotation & Model**
  - Ground truth labeling performed in CVAT
  - YOLO11n object detection model (Ultralytics 8.3.203)
  - Trained using PyTorch 2.8.0 with CUDA acceleration

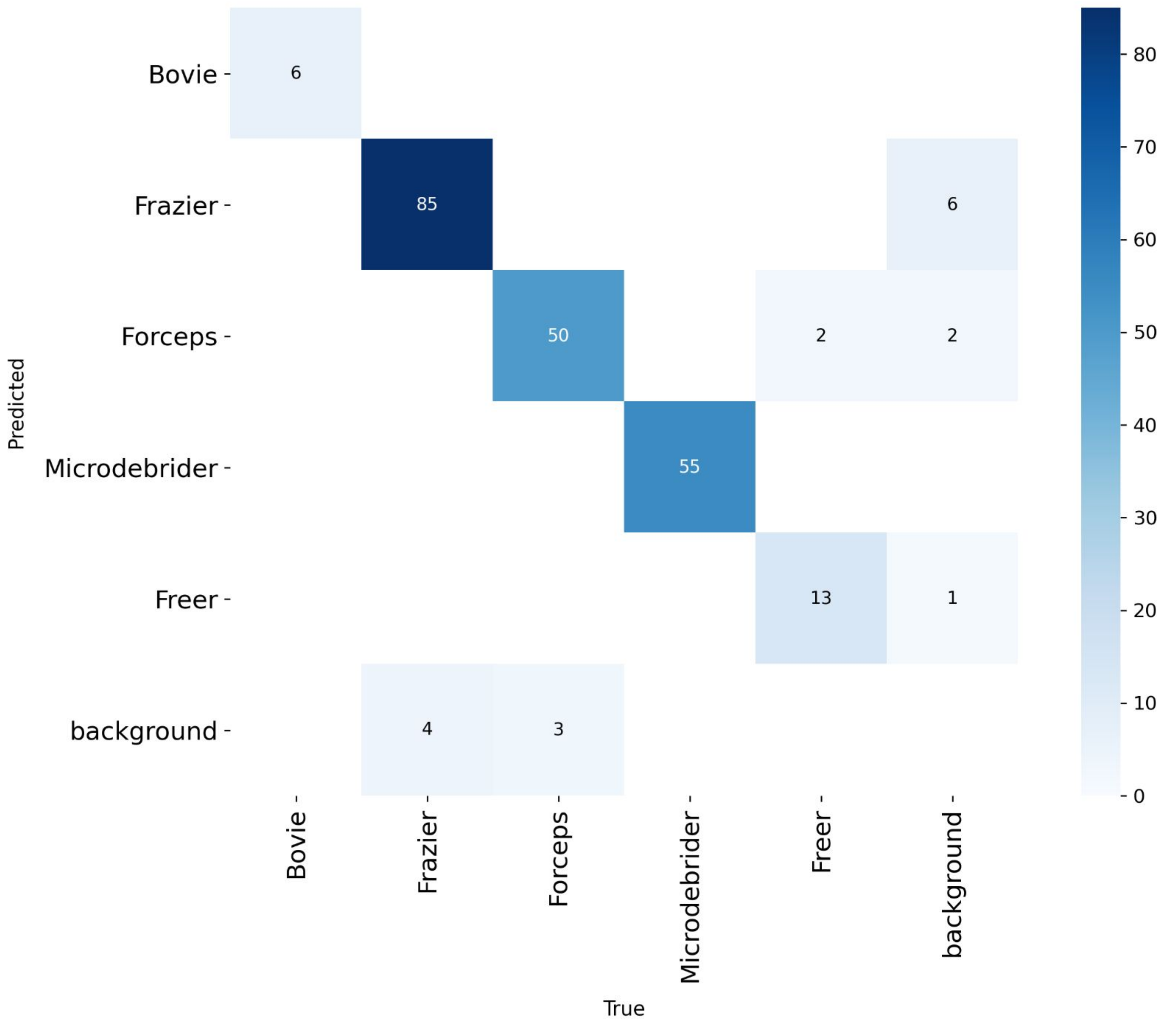

Figure 2: Confusion Matrix. Error analysis of the confusion matrix revealing frequency of misclassification

RESULTS

- **Overall Performance (Test Set)**
  - 322 images, 218 instances, inference time 4.1 ms/image
  - Precision (P): 96.4%; Recall (R): 94.8%; mAP50: 96.6%
- **Instrument-Level Performance (Fig. 1)**
  - Bovie - P: 100%; R: 100%; mAP50: 99.5%
  - Microdebrider - P: 100%; R: 99.9%; mAP50: 99.5%
  - Frazier - P: 95.4%; R: 93.0%; mAP50: 96.5%
  - Forceps - P: 93.9%; R: 94.3%; mAP50: 95.5%
  - Freer elevator - P: 92.9%; R: 86.7%; mAP50: 91.9%
- Background most commonly misclassified as Frazier (67%)
- **Confusion matrix (Fig. 2)** and sample predictions demonstrate accurate instrument detection with confidence scoring

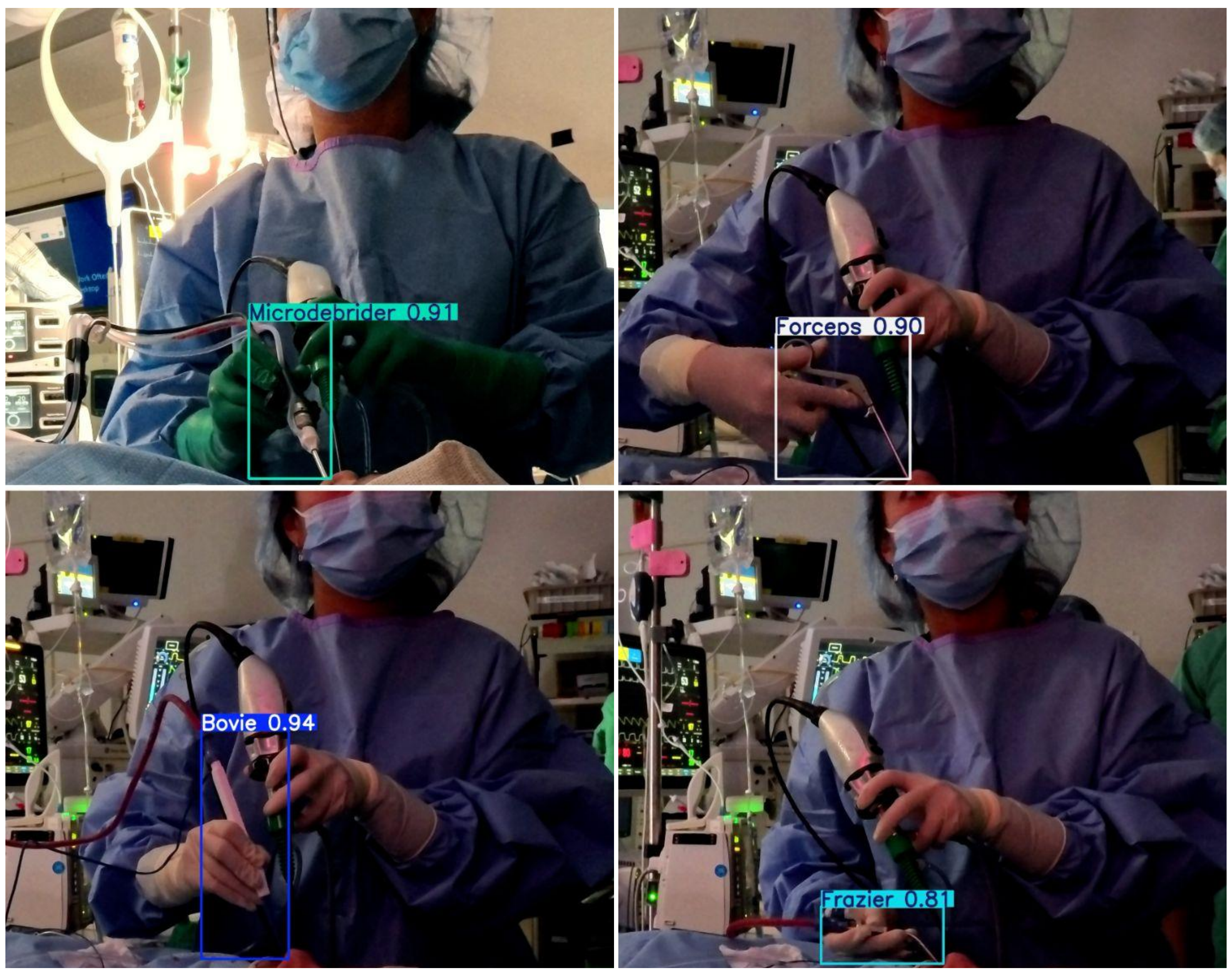

Figure 1: Sample predictions on still-frames. Representative model outputs with corresponding confidence level

CONCLUSIONS

- Computer vision-based instrument detection in endoscopic sinus and skull base surgery is feasible, accurate, and efficient
- High model performance and rapid inference support potential real-time applications without disruption to surgical workflow
- Scalable video capture using compact action cameras enables broad adoption across operating environments
- Represents an objective and quantitative tool for surgical training, performance benchmarking, and workflow optimization

REFERENCES

1. Choksi S et al. Bringing Artificial Intelligence to the operating room: edge computing for real-time surgical phase recognition. *Surg Endosc*. 2023;37(11):8778-8784.
2. Yanik E et al. Video-based formative and summative assessment of surgical tasks using deep learning. *Sci Rep*. 2023;13(1):1038.
3. Ward TM et al. Computer vision in surgery. *Surgery*. 2021;169(5):1253-1256.
4. Ganry L et al. Modified GoPro Hero 6 and 7 for Intraoperative Surgical Recording-Transformation Into a Surgeon-Perspective Professional Quality Recording System. *J Oral Maxillofac Surg*. 2019;77(8):1703.e1-1703.e6.