Computer Vision Approach for Instrument and Anatomy Detection during Transcanal Endoscopic Ear Surgery

Obi Nwosu, MD[1,2]; Krish Suresh, MD[1,2]; Daniel Lee, MD[1,2]; Matthew Crowson, MD[1,2]

[1]Mass Eye and Ear Dept of Otolaryngology, [2]Harvard Medical School

Abstract

Background/Objective: In computer vision (CV), detection tasks involve classifying and bounding a structure of interest. We aimed to develop a proof-of-concept model for automated anatomy and instrument detection during transcanal endoscopic ear surgery (TEES).

Methods: We developed a detection model to identify instruments and middle ear anatomy during TEES for cases of otosclerosis and ossiculoplasty. An internal model was trained and tested on institutional TEES images. This internal model was then externally validated on an images extracted from TEES videos obtained from YouTube. Images in which middle ear anatomy or instruments were not visualized were excluded. Detections were evaluated using recall and intersection over union (IoU).

Results: On the internal dataset, the model achieved an overall recall/IoU of 73%/65% compared to 65%/67% on the external validation dataset. On both the datasets, instrument detection was most precise and accurate. Detection of anatomic structures was generally more accurate for structures with clear, well-defined borders.

Conclusions: A CV approach for instrument and anatomy tracking during TEES is feasible and may contribute to development of augmented reality-based endoscopic ear video systems.

Methods

- 1,045 images from 5 institutional TEES recordings used to train and test an internal model using five-fold cross validation.
- Internal model was then validated on 100 external images from 5 TEES YouTube videos.
- All images manually annotated to create ground-truth annotations. Structures of interest included the chorda tympani, incus, instruments, promontory, and stapedial tendon.
- Open-source CV toolkit, Detectron2, utilized to train and test models[1].
- Model performance determined by accuracy of classification (recall) and precision of detection (IoU) (Figure 1).

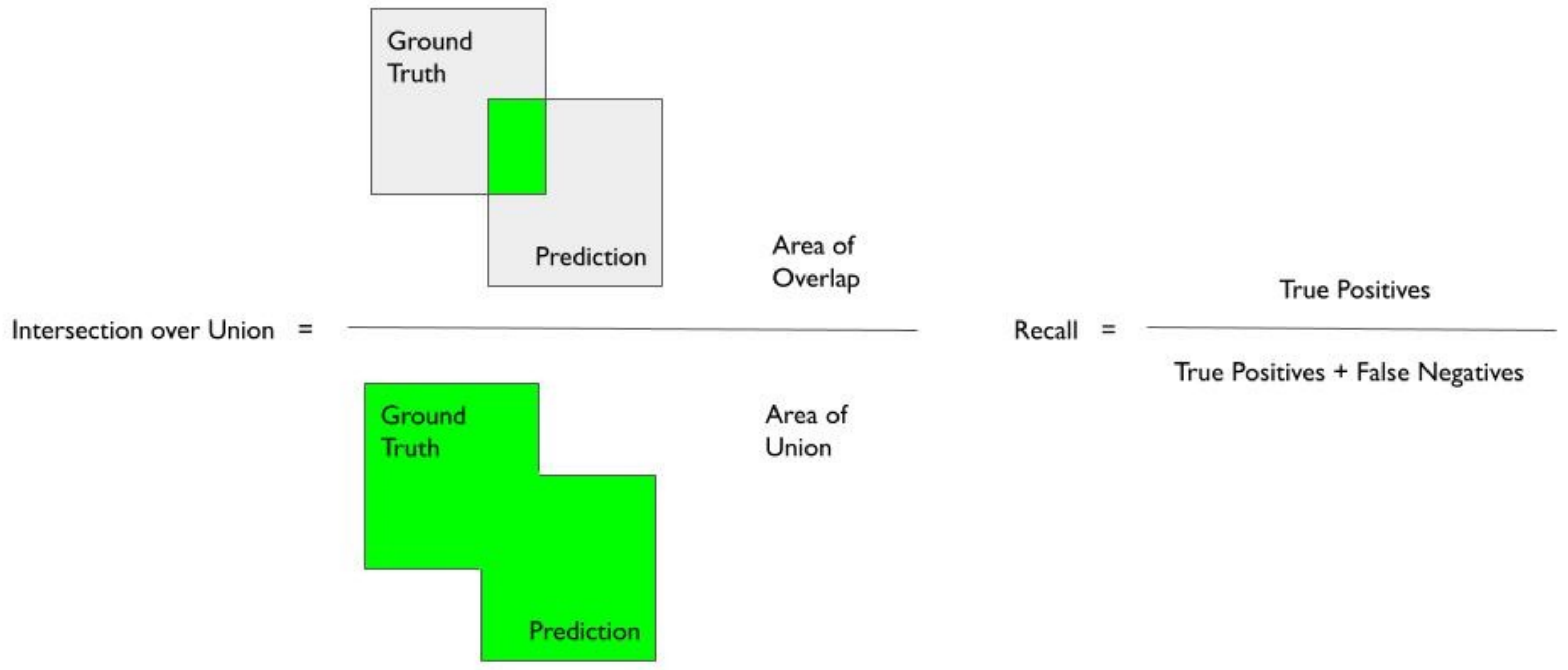

Figure 1. IoU, a precision metric, represents the ratio of the area of overlap to the area of union between the ground truth and predicted annotations. Recall represents the true positivity rate.

Results:

instrument/anatomy detection during TEES is feasible

- Structure-specific detection performance on both both internal and external datasets is reported in **Table 1.**

	Five-Fold Validation		External Validation	
Structure	Recall ± std	IoU ± std	Recall	IoU
inst	94 ± 4%	84 ± 3%	91%	82%
in	89 ± 12%	67 ± 10%	52%	74%
ct	81 ± 11%	69 ± 4%	75%	68%
pr	69 ± 36%	54 ± 11%	22%	41%
st	30 ± 26 %	51 ± 35%	10 %	70%

Table 1. Structure specific performance on both internal and external testing. ct – chorda tympani, in – incus, inst – instrument, pr – promontory, st – stapedial tendon

- Examples of model predictions on both external and internal images are displayed in **Figure 2-3.** Percentages represent model confidence of prediction.
- A demonstration of live anatomic and instrument tracking can be accessed at the QR code below.

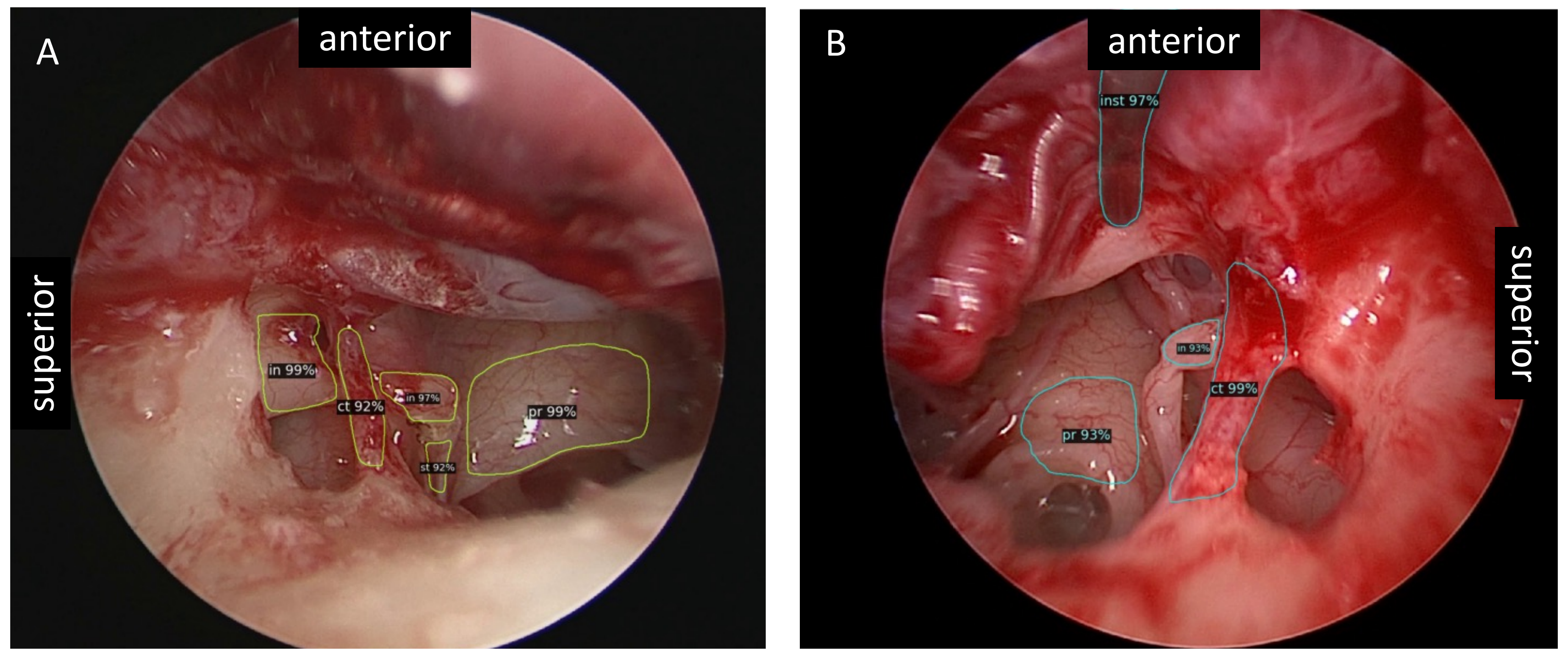

Figure 2. Model predictions on internal test images. A) Right ear, the model correctly identifies each structure. B) Left ear, the model correctly identifies all structures except the stapedial tendon which was underrepresented in the distribution of ground-truth annotations. ct – chorda tympani, in – incus, inst – instrument, pr – promontory, st – stapedial tendon

SCAN HERE FOR A LIVE DEMO

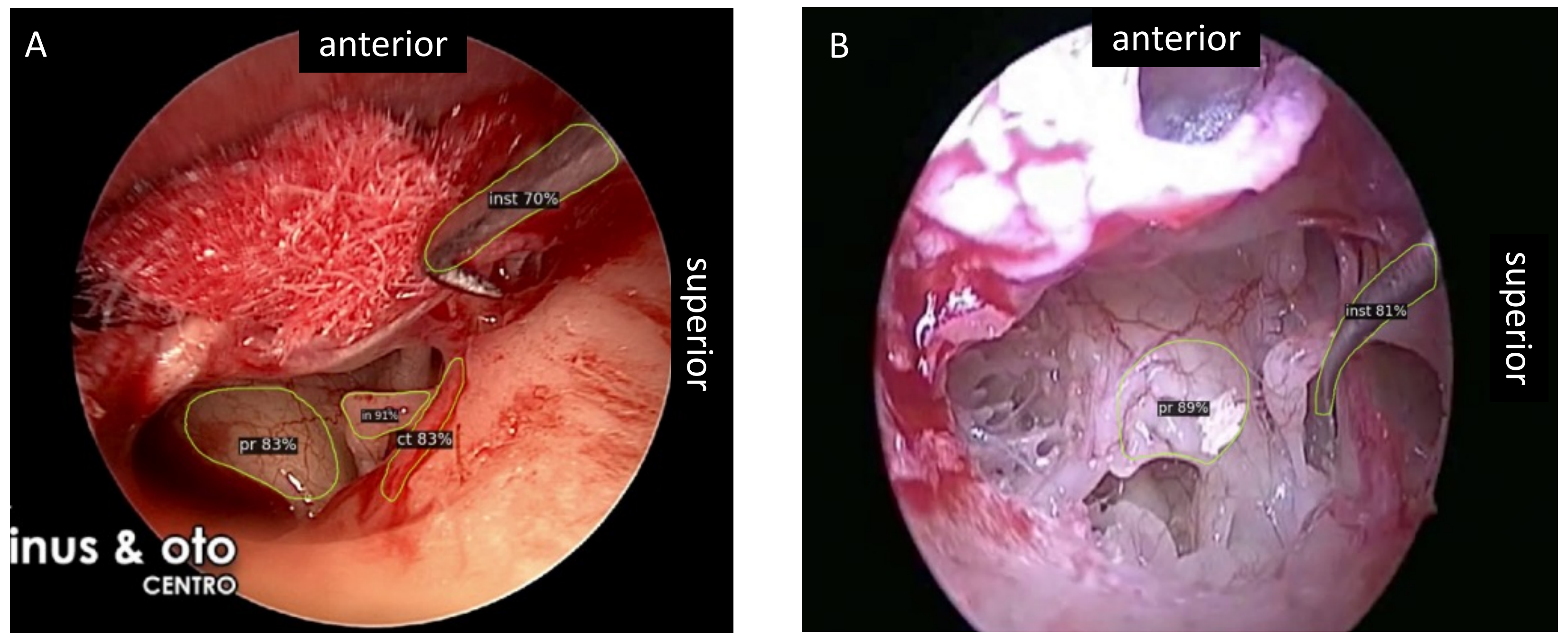

Figure 3. Model predictions on external YouTube TEES images. A) Left ear, high-quality image where model performs well. B) Left ear, poorer quality image where model performance falls; note missed detections of the chorda tympani, incus, and stapedial tendon. ct – chorda tympani, in – incus, inst – instrument, pr – promontory, st – stapedial tendon. *Images accessed publicly under the Creative Commons License. Credit: The image in panel A is obtained from a video uploaded by Dr. João Flávio Nogueira. The image in panel B is obtained from a video uploaded by Dr. Gouda Ramesh.*

Discussion

- Excellent instrument detection given unique appearance in the surgical field
- Anatomic structures with distinct borders (ie incus, chorda) tended to have better detection than those with indistinct boundaries (ie promontory)
- Limitations:
  - Model trained on high-definition endoscopic video data and is thus biased towards high-quality image data.
  - Model trained with video data from only ossiculoplasty or stapes surgery. Thus, performance cannot be generalized to all TEES cases.
- Future directions:
  - Retraining with more heterogenous data (both in types of surgeries included and in video quality)
  - Development of companion model for depth prediction from two-dimensional video data

Conclusions

- We developed a CV detection model for tracking anatomy and instruments during TEES.
- This approach could be used for development of novel training, navigation and robotic-assisted surgical platforms.

Contact

Matthew Crowson, MD

Mass Eye and Ear/Harvard Department of Otolaryngology

mcrowson@meei.harvard.edu

References

1. https://github.com/facebookresearch/detectron2
2. Kozin ED, Lee DJ, Pollak N. Getting Started with Endoscopic Ear Surgery. Otolaryngol Clin North Am. 2021;54(1):45-57. doi:10.1016/j.otc.2020.09.009
3. Mahadevkar S V., Khemani B, Patil S, et al. A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions. IEEE Access. 2022;10:107293-107329. doi:10.1109/ACCESS.2022.3209825
4. You E, Lin V, Mijovic T, Eskander A, Crowson MG. Artificial Intelligence Applications in Otology: A State of the Art Review. Otolaryngol Head Neck Surg. 2020;163(6):1123-1133. doi:10.1177/0194599820931804
5. Ward TM, Mascagni P, Ban Y, et al. Computer vision in surgery. Surgery. 2021;169(5):1253-1256. doi:10.1016/j.surg.2020.10.039