AENEAS Project: Live Image-Based Navigation and Roadmap Generation in Endoscopic Neurosurgery Using Machine Vision

Victor E. Staartjes, MD, PhD*, Gary Sarwin, MSc‡, Alessandro Carretta, MD§, Matteo Zoli, MD, PhD§, Diego Mazzatenta, MD, PhD§, Luca Regli, MD*, Ender Konukoglu, PhD‡, Carlo Serra, MD*

*Machine Intelligence in Clinical Neuroscience (MICN) Laboratory, Department of Neurosurgery, University Hospital Zurich, Clinical Neuroscience Centre, University of Zurich, Zurich, Switzerland; ‡Computer Vision Lab (CVL), ETH Zurich, Zurich, Switzerland; §Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna, Bologna, Italy

Part of this publication have been published in the preprint server ArXiv under the name "Live image-based neurosurgical guidance and roadmap generation using unsupervised embedding" 31 March 2023.

Correspondence: Carlo Serra, MD, Machine Intelligence in Clinical Neuroscience (MICN) Laboratory, Department of Neurosurgery, University Hospital Zurich, Clinical Neuroscience Centre, University of Zurich, Frauenklinikstrasse 10, Zurich 8091, Switzerland. Email: carlo.serra@usz.ch

Received, October 08, 2024; **Accepted,** December 26, 2024; **Published Online,** April 28, 2025.

Operative Neurosurgery 30:1–7, 2026 https://doi.org/10.1227/ons.0000000000001583

BACKGROUND AND OBJECTIVES: Artificial intelligence algorithms have proven capable of replicating cognitive processes. Our aim was to replicate human roadmap generation for endoscopic neurosurgery with a live image-based machine vision method.

METHODS: Surgical videos of a highly standardized surgical approach are labeled and used for algorithm training. After object detection (YOLOv7) to generate bounding boxes for landmark anatomical structures, an autoencoder first encodes the currently detected structures into an estimated position within this anatomical roadmap and then enables extrapolation of structures that are expected to be encountered in forward or backward directions. Average precision of the model applied to the test videos at an intersection-over-union threshold of 0.5 is reported.

RESULTS: In total, 166 anonymized endoscopic recording (3×10^6 labeled video frames) were included. We performed model development using 146 videos and held out 20 videos for evaluation (test set). The performance regarding bounding box detection among the 20 test set videos on average was 53.4. Evaluation of the performance of the autoencoder model in detecting the current position within the roadmap of the surgical approach is evaluated semiquantitatively, showing that the first detection of anatomical structures by the model corresponds well to their label distribution along the latent variable encoding the anatomical roadmap. We also provide videos demonstrating the mixed reality head's up display for anatomical navigation.

CONCLUSION: Our method enables reliable identification of key anatomical structures during endoscopic endonasal trans-sphenoidal surgery in mixed reality. Through encoding detected landmark anatomical structures, a surgical roadmap is encoded. This approach allows for detection of visible anatomical structures and enables extrapolation toward the location of those yet to be dissected in deeper anatomical layers. Further development of such algorithms may pave the way toward adding a mixed reality, real-time anatomical navigation software to the neurosurgeon's armamentarium.

KEY WORDS: Machine vision, Anatomical recognition, Machine learning, Artificial intelligence, Anatomical guidance

The ability of any surgeon to successfully operate on a case is given by 2 elements: a scientific element, the surgeon must know what should be done, and a technical element, manual dexterity. Each surgeon must have some sort of mental "roadmap" to refer during surgery to know what needs to be conducted and must then have the manual dexterity to perform what needs to be performed. Both elements are highly individual. Although dexterity can be improved through laboratory training,

ABBREVIATIONS: AI, artificial intelligence; **AP,** average precision; **IoU,** intersection over union.

robotics may, in the future, unify and standardize surgical techniques, enhancing and expanding their technical capabilities. Building a mental roadmap, however, is a purely cognitive process and is therefore more complex. It is the result of a long learning curve, personal experience, and mentoring one has received and as such is difficult to be standardized, especially if the goal is for the excellence of a few to be replicable by the many.

One could argue that universal access to care in neurosurgery is limited, especially compared with other medical disciplines with more reproducible standards of care.[1,2]

The introduction of algorithms and highly complex statistical methods of analyzing huge amounts of data, which collectively go by the name of artificial intelligence (AI), offers the possibility of replicating even human cognitive processes,[3] provided that the terms of the process to be replicated are decomposable in quantitative terms. In the medical field, we have already seen the introduction of algorithms to support clinical, diagnostic, prognostic decision-making processes and the analysis of radiological and histological images.[4-10]

The goal of the AENEAS project was to develop AI algorithms capable of replicating the cognitive processes that allow each surgeon to create for each type of surgery a mental roadmap to refer during the operation. The ultimate goal was to be able to achieve a kind of user-friendly "heads-up" that can tell the surgeon in real time based solely on live intraoperative images what to do as if by the surgeon's side was the most experienced in the field.

In a first pilot project, we explored the possibility of developing machine vision algorithms capable of recognizing simple anatomical structures (the nasal septum and turbinates)[11] in the nasal phase of pituitary surgery. This demonstrated the initial premise required for our project: that a computer is able to learn detection of anatomical structures along the trajectory of a surgical approach. In the second step that we present now with this article, we illustrate (1) the performance of a deep learning–based object detection method (YOLO) in detecting an enlarged number of anatomical structures, again in pituitary surgery and (2) a method for generating neurosurgical roadmaps capable of mapping the surgeon's current location using live image feedback, being able to predict which structure should appear going forward or backward.

METHODS

Overview of the Roadmap Concept

Intraoperative orientation is based on thorough analysis of preoperative images and exact knowledge of relevant surgical anatomy, which a surgeon primarily acquires as a result of repeated visual exposure to the anatomy being road mapped. The idea behind our methodological approach was to try to replicate this human learning process through repeated visual exposure and have an algorithm learn to recognize anatomical structures by watching a multitude of surgical videos of the same approach and extracting the spatial relationships between the detected structures.

Data Acquisition and Labeling

To ensure the generalizability of our algorithm, we extracted videos of the same type of transnasal trans-sphenoidal endoscopic pituitary surgery (TSS) intervention from surgical video database. Surgeries were performed by 4 different surgeons. Surgical video frames were labeled with dedicated software (VoTT) by an experienced neurosurgical resident (A. C.) and checked according to an internal quality check protocol by a senior neurosurgeon (C. S.). Both right- and left-nostril cases were included. Labeling was performed into 16 different classes as presented in Table. The use of patient data from the pituitary registry was approved by the local ethical review boards (KEK 2023-02265). All patients signed research consent forms.

Machine Vision Approach

The underlying technical machine vision concepts underlying our algorithm are reported in Sarwin et al 2023.[12] In brief, while our prototype algorithm[11] had worked with single-pixel annotations and a weakly supervised U-Net[13] neural network to perform heatmap regression based on the approach described by Payer et al,[14,15] the roadmap

TABLE. Quantitative Performance of the Object Detection Model Among the 20 Held-Out Videos for Testing

Anatomical structure	AP_{50}
General	
Average	53.4
Instrument	94.4
Nasal phase	
Nasal Septum	78.6
Superior nasal meatus	40.6
Middle nasal meatus	63.8
Inferior nasal meatus	62.9
Nasal floor	65.3
Choana	54.3
Sphenoethmoidal recess	41.1
Sphenoid ostium	43.1
Sphenoid phase	
Sphenoidal sinus	74.0
Rostrum	17.9
Left osseous carotid	32.4
Right osseous carotid	21.6
Planum	34.9
Sellar floor	70.6
Clival recess	58.9

AP, average precision; IoU, intersection over union.
We provide values for AP at an IoU value of 0.5 (AP_{50}), corresponding to the performance of the model in placing bounding boxes on detected anatomical structures.

concept requires a different model architecture.[12] Conceptually, we first applied an object detection algorithm (YOLOv7)[16] to put bounding boxed around anatomical objects. The model was trained on sparsely annotated video data for 16 classes (see Table) and then allowed us to generate bounding boxes for every single frame.

To implement the roadmap concept, we encoded the current location of each bounding box into a 1-dimensional (linear) latent space representing the surgical path extending from the nostril (beginning) to the sellar floor (end). This allows localization of the relative position of specific anatomical structures along the TSS approach (Figure 1). We excluded the surgical instrument from the YOLO output because its appearance is not related to a specific location in the surgical path. We then use 2 decoders to generate class probabilities and bounding box coordinates for each structure, to create the real-time anatomical "head's up display" (HUD) based on the position in the latent space. This correlation process is in a way not unlike an experienced neurosurgeon who relies on several anatomical landmarks to determine the current stage of the surgical approach, which in turn then allows the exact localization of each single anatomical structure.

Evaluation

To assess the performance of the object detection model, we report the average precision (AP) of the model applied to the test videos at an intersection-over-union (IoU) threshold of 0.5. This metric, also known as the area under the precision-recall curve, provides a measure of how accurately the model can detect and localize objects—in this case, the anatomy and the surgical instruments. Higher AP values indicate better performance and fewer false-positives. IoU is crucial in this metric because it evaluates how closely the predicted bounding boxes match the actual object locations. Specifically, a predicted bounding box is classified as a true-positive if it overlaps by at least 50% with the actual bounding box.

To assess the performance of the roadmap autoencoder, we qualitatively compare the probability of encountering specific anatomical structures within certain parts of the single-dimensional encoded latent space with their first detection in every test set video.

RESULTS

In total, 166 anonymized endoscopic recording (in 166 different patients) were included and labeled. This led to a total of approximately 19 000 labeled video frames. We performed model development using 146 videos and held out 20 videos for evaluation (test set).

Anatomical Structure Detection

The performance (AP for IoU values of 0.5) regarding bounding box detection is reported in Table. Overall, among the 20 test set videos, the average AP_{50} among all anatomical structures was 53.4, ranging from 17.9 for rostrum and 21.6 for right osseous carotid to 78.6 for the nasal septum and a maximum of 94.4 for the instrument class. When comparing these numbers with the reported AP of 69.7% of the YOLOv7 model on the COCO data set, which is a widely used benchmark for object detection task, it becomes evident that some classes are detected with state-of-the-art accuracy, whereas for others, the model's accuracy could be improved (for example, for the rostrum, see Table). The variation in performance across various classes can largely be attributed to differences in the amount of labeled data available for each class, with more labels often correlating with higher performance.

Roadmap Generation

Evaluation of the performance of the autoencoder model in detecting the current position within the roadmap of the surgical approach and of the finally generated bounding boxes (as a sort of mixed-reality HUD) is demonstrated in Figures 2 and 3, the latter showing that the first detection of anatomical structures by the model corresponds well to their label distribution along the latent variable encoding the anatomical roadmap.

Deployment

We demonstrate the deployment of AENEAS as a mixed-reality tool for endoscopic pituitary surgery in Video 1 (nasal phase) and Video 2 (sellar phase).

DISCUSSION

In this study, we have demonstrated how it is possible to use machine vision algorithms to correctly recognize intraoperative

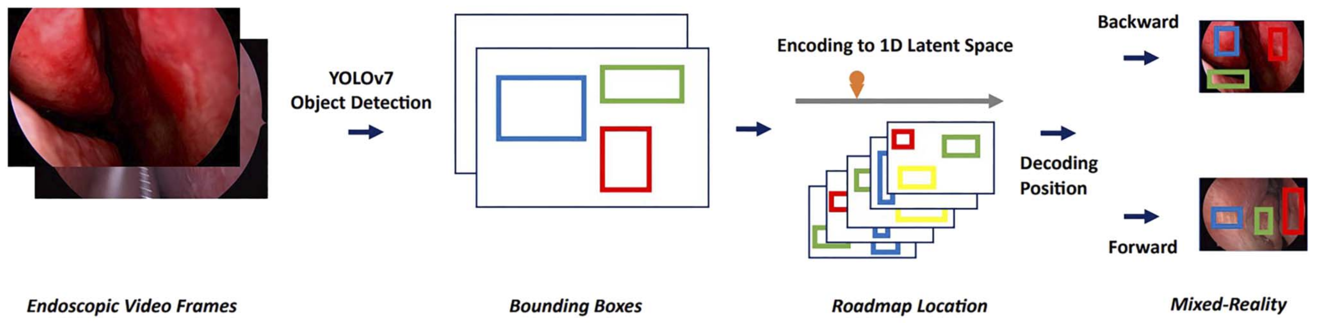

FIGURE 1. *Representation of methodological approach. A sequence of video frames processed to detect bounding boxes relative to various anatomical structures. A neural network then encodes the sequence of detections into a 1D latent variable correlating with the position along the surgical path. Given the current position along the surgical path, an estimation of anatomical structures in the forward or backward directions can be obtained, by extrapolating the current value of the latent variable.*

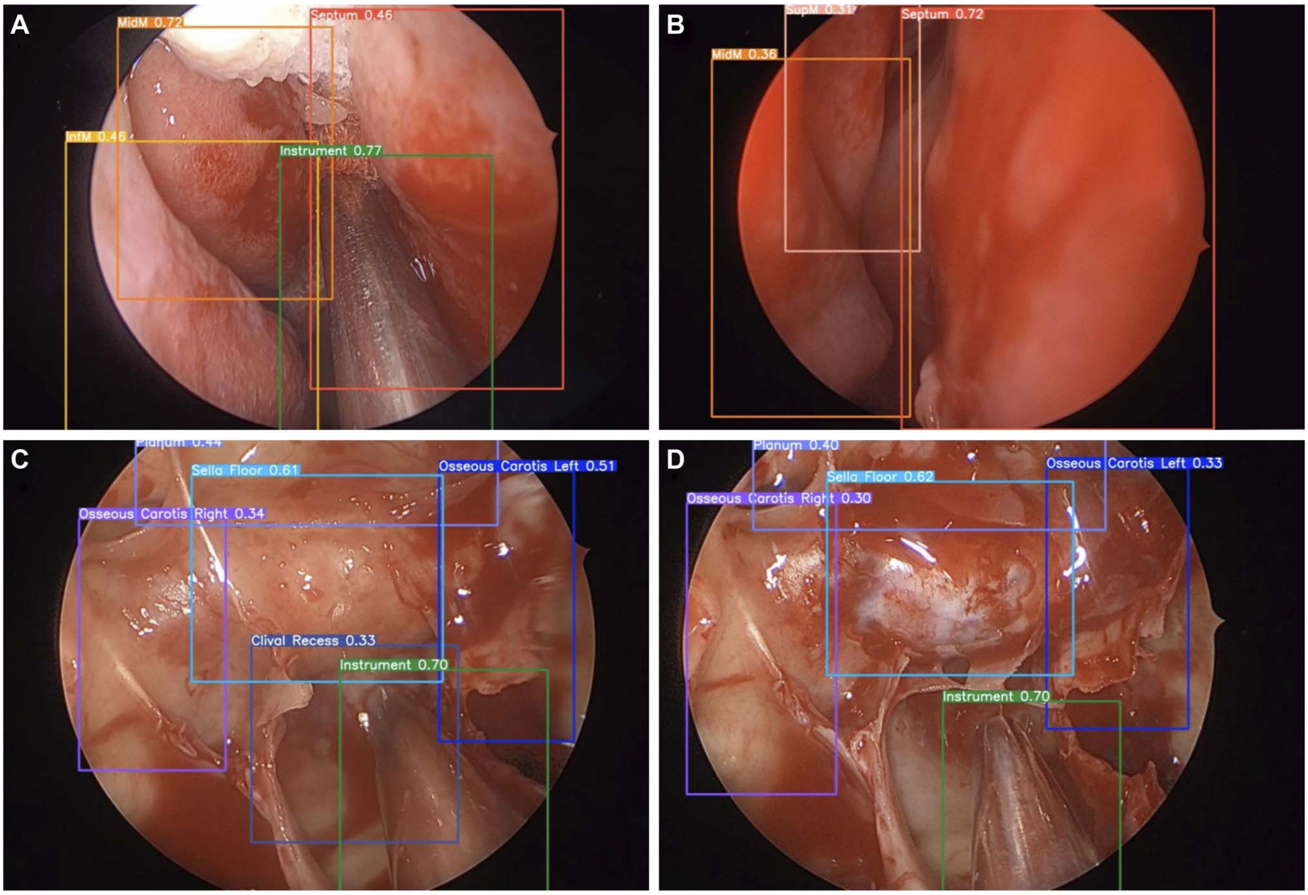

FIGURE 2. *Sequence of snapshot depicting different stages of the surgical approach. The frames were extracted from the hold-out videos on which the algorithm was tested. Note how in both the nasal phase (frame* **A** *and* **B***) and the sphenoidal phase (frame* **C** *and* **D***) the different anatomical structures relevant to the surgical approach are recognized by the algorithm.*

surgical anatomy in TSS surgery. In addition, we have shown how it is also possible to replicate and automate the cognitive process of constructing a neurosurgical roadmap using AI. These advancements pave the way for diverse research directions, all aimed at modeling and democratizing access to even the most complex surgical procedures.

Roadmap Concept

Some AI applications have been reported showing, as we also previously did,[11,12] that computers can learn to recognize surgical anatomy without being given prior information.[17-19]

The addition of the roadmap concept represents, however, a novel approach for detecting anatomical structures. Although no explicit anatomical atlas is provided, the encoded latent variable acts as a sort of "implicit" anatomical atlas that is constructed for each anatomical region or surgical procedure. Performance seems to be markedly increased compared with our prior architecture, and the ability to extrapolate toward other anatomical structures that are likely to come into view during the next surgical step is added. Especially in endoscopic surgery, the ability to also forecast structures that are expected during backward (eg, outwards) movement is also crucial as the endoscope is frequently passed inside and out of the nasal cavity again and again. This fact required some adjustments. First, to validate the roadmap concept and assess its correlation with depth along the surgical path, we considered the first appearance of each structure, with the assumption that deeper structures should appear later than shallower ones. Second, multiple instruments are frequently passed inside and outside, so the bounding boxes for the class "instrument" had to be excluded from construction of the roadmap, as it was of course only poorly correlated with the surgical roadmap. Beyond its current limitations, the most important aspect of the roadmap concept is that it can, in our opinion, represent the foundation from which to build the future of automation in neurosurgery and surgery more generally.

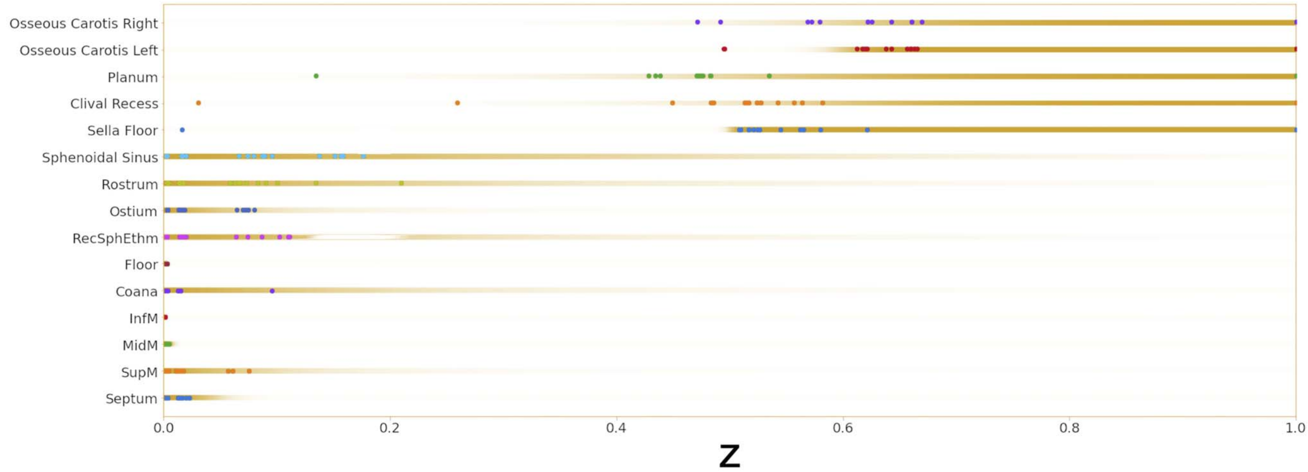

FIGURE 3. *The normalized generated confidences of each class along the latent space. This visualizes the probability of finding a certain anatomical structure at a specific point in the latent space. In addition, video frames of 20 test videos responsible for the first appearances of anatomical structures in every video have been encoded and overlaid onto the confidence intervals to demonstrate that their locations correlate with the beginning of these intervals.*

Intraoperative Navigation

With a range of options which provide intraoperative guidance,[20-23] only fluorescence (ie, 5-aminolevulinic acid) is picture in picture. However, it only depicts pathology. This highlights the need for a modality that can provide valuable information in the same format or modality as the 1 used during operation and which thereby could make navigation more streamlined without having to take your eyes of the surgical scene. Our algorithm has the possibility to showcase valuable information such as the surgical anatomy and roadmap directly superimposed on the video stream, which could readily be turned off as well with the touch of a button or even voice command.

Moreover, it has the advantage of deciphering a visual input at the same time as the operating surgeon, in the same modality (ie, optic) in which it is presented and in which the operator will have to make a decision. In other words, there is no intermediation of other instrumentation whose information is returned in a different format that will then have to be somehow translated by the operator in order for it to be usefully usable.

It has previously been discussed that a real-time intraoperative anatomical navigation method may get into trouble when encountering pathological anatomy, such as in subarachnoid hemorrhage, neuro-oncology, revision surgeries, or even simply severely atrophied nasal mucosa.[24] Although the prior architecture certainly could not and while the current model also does not handle pathological anatomy, the roadmap concept may be one of the keys toward enabling this technology in distorted anatomy. The algorithm, similarly to a human neurosurgeon, currently recognizes anatomical landmarks and recognizes where it is currently situated. It can then extrapolate the location of other anatomical structures relative to the currently detected landmarks and could thus also be able to know where a tumor is. This would however necessitate some preoperative information or, alternatively, the ability to detect distorted anatomical landmarks or pathological tissue as a separate class. In short, this is a research direction that will have to be pursued in the near future to fully understand the potential of the roadmap concept and of real-time visual image based navigation.

Future Perspectives

Properly implemented, mixed-reality concepts could assist in improving intraoperative anatomical orientation, in the case of AENEAS by providing labeling of anatomical structures currently visible within the endoscopic/microscopic video feed, and the ability to extrapolate toward structures not yet visible within the current images, by providing bounding boxes as a form of HUD. Importantly, HUDs or mixes reality in general must never distract from important, real-world visual inputs by obstruction vision or by reducing the surgeon's alertness.[25,26] For this reason, tools such as AENEAS would have to be integrated into endoscopes and microscopes as on-demand features, able to be turned on and off at will and only when necessary. Another question is whether neurosurgeons would accept this form of assistance or if there might be a significant uncertainty or distrust toward the algorithm. Although there are no specific data regarding an application such as AENEAS, a recent survey demonstrated that neurosurgeons seem to be comfortable with intraoperative assistance by AI aides.[27]

This also links into the next issue with such technical adjuncts in the neurosurgical operating room: resident training and reproducibility. It could be argued that real-time anatomical labeling could improve the learning effects reached during a single procedure—in addition to the potential benefit of reducing surgical complications created by spatial disorientation, such as carotid lacerations. As outlined in Khan et al[28] and Sarker et al,[29] even just detecting the different stages of an endoscopic surgical approach can be beneficial regarding surgical training and may improve operative workflows and efficiency. Having a well-validated, high-performance anatomical annotation algorithm available may be similar in effect regarding anatomical orientation to being assisted by a highly experienced neurosurgeon, which could prove particularly useful in conditions known in real-world practice where 1 is faced with a nondeferrable emergency situation (for example, a pituitary apoplexy with acute visual impairment) without having the human expertise at hand. Although there must never be too much reliance on any single technical adjunct, it is also true that reasoning by analogy with civil aviation autopilot systems could lead to a drastic reduction in the incidence of trivial but potentially grievous errors not only in the most difficult situations but also and especially in routine ones, which are then also the most numerous and high impact ones.

The quantitatively and qualitatively promising performance of our machine vision methods indicate that they may be a promising component for creating fully automated anatomical recognition software that is integrated into the surgical workflow as mixed reality. Further development must focus on widening the applications of AENEAS, first to other relatively consistent surgical approaches such as the pterional craniotomy and sylvian fissure split or the retrosigmoid craniotomy and then in pathological anatomy as discussed above. In combination with larger training samples and a higher number of anatomical landmarks, implementation studies integrating the interface into the neurosurgeon's operative visualization devices may be the next step toward real-time intraoperative anatomical guidance.

Limitations

As described in more detail above, pathological anatomy cannot yet be handled by our algorithm, and it can only be applied to 1 specific surgical approach. It thus remains unclear, whether this approach does bring any measurable benefit, may that be regarding reducing complications, operative times, or empowering and democratizing microsurgical training.

CONCLUSION

Our method enables reliable identification of key anatomical structures during TSS surgery in a mixed reality context. Through encoding detected landmark anatomical structures, a surgical roadmap is established. This approach allows not only for detection of superficially visible anatomical structures but also enables extrapolation toward the location of those yet to be dissected in deeper anatomical layers. In this way, we could replicate the human cognitive process that allows surgeon to orientate within surgical procedures. Further development of such algorithms may pave the way toward adding a mixed reality, real-time anatomical navigation software to the neurosurgeon's armamentarium with the ultimate goal of being able in the future to model and, thus, make available to the widest possible audience, any kind of surgical procedure, even those nowadays performed by very few highly specialized operators only.

Funding

This research is supported by SNF Project IZKSZ3_218786 research grant.

Disclosures

The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Skjold-Ødegaard B, Søreide K. Standardization in surgery: friend or foe? *Br J Surg.* 2020;107(9):1094-1096.
2. Dobson GP. Trauma of major surgery: a global problem that is not going away. *Int J Surg.* 2020;81:47-54.
3. Siemens G, Marmolejo-Ramos F, Gabriel F, et al. Human and artificial cognition. *Comput Educ Artif Intel.* 2022;3(9):100107.
4. Muscas G, Orlandini S, Becattini E, et al. Radiomic features associated with extent of resection in glioma surgery. In: Staartjes VE, Regli L, Serra C, eds. *Machine Learning in Clinical Neuroscience*. Vol 134. Acta Neurochirurgica Supplement. Springer International Publishing; 2022:341-347.
5. Zoli M, Staartjes VE, Guaraldi F, et al. Machine learning-based prediction of outcomes of the endoscopic endonasal approach in Cushing disease: is the future coming? *Neurosurg Focus.* 2020;48(6):e5.
6. Zanier O, Zoli M, Staartjes VE, et al. Machine learning-based clinical outcome prediction in surgery for acromegaly. *Endocrine.* 2021;75(2):508-515.
7. Voglis S, van Niftrik CHB, Staartjes VE, et al. Feasibility of machine learning based predictive modelling of postoperative hyponatremia after pituitary surgery. *Pituitary.* 2020;23(5):543-551.
8. Staartjes VE, Serra C, Muscas G, et al. Utility of deep neural networks in predicting gross-total resection after transsphenoidal surgery for pituitary adenoma: a pilot study. *Neurosurg Focus.* 2018;45(5):e12.
9. Van Niftrik CH, Van Der Wouden F, Staartjes VE, et al. Machine learning algorithm identifies patients at high risk for early complications after intracranial tumor surgery: registry-based cohort study. *Neurosurgery.* 2019;85(4):e756-e764.
10. Staartjes V, Serra C, Muscas G, et al. Machine learning-based predictive analytics in transsphenoidal surgery for pituitary adenoma. *J Neurol Surg B Skull Base.* 2019; 80(S 01):A143.
11. Staartjes VE, Volokitin A, Regli L, Konukoglu E, Serra C. Machine vision for real-time intraoperative anatomic guidance: a proof-of-concept study in endoscopic pituitary surgery. *Oper Neurosurg.* 2021;21(4):242-247.
12. Sarwin G, Carretta A, Staartjes V, et al. Live image-based neurosurgical guidance and roadmap generation using unsupervised embedding. In: Frangi A, de Bruijne M, Wassermann D, Navab N, eds. *Information Processing in Medical Imaging.* Springer Nature Switzerland; 2023:107-118.
13. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*;2015:234-241. Accessed August 27, 2020. http://arxiv.org/abs/1505.04597
14. Payer C, Štern D, Bischof H, Urschler M. Multi-label whole heart segmentation using CNNs and anatomical label configurations. In: Pop M, Sermesant M, Jodoin PM, et al., eds. *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Lecture Notes in Computer Science. Springer International Publishing; 2018:190-198.
15. Payer C, Stern D, Bischof H, Urschler M. Regressing heatmaps for multiple landmark localization using CNNs. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*. Springer International Publishing AG; 2016:230-238.

16. Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR);2023:7464-7475.
17. Petscharnig S, Schöffmann K. Learning laparoscopic video shot classification for gynecological surgery. *Multimed Tools Appl.* 2018;77(7):8061-8079.
18. Twinanda AP, Marescaux J, de Mathelin M, Padoy N. Classification approach for automatic laparoscopic video database organization. *Int J Comput Assist Radiol Surg.* 2015;10(9):1449-1460.
19. Takiyama H, Ozawa T, Ishihara S, et al. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. *Scientific Rep.* 2018;8(1):7497-7498.
20. Gerard IJ, Kersten-Oertel M, Petrecca K, Sirhan D, Hall JA, Collins DL. Brain shift in neuronavigation of brain tumors: a review. *Med Image Anal.* 2017;35:403-420.
21. Serra C, Stauffer A, Actor B, et al. Intraoperative high frequency ultrasound in intracerebral high-grade tumors. *Ultraschall in der Medizin (Stuttgart, Germany: 1980).* 2012;33(7):e306-e312.
22. Rammeloo E, Schouten JW, Krikour K, et al. Preoperative assessment of eloquence in neurosurgery: a systematic review. *J Neurooncol.* 2023;165(3):413-430.
23. Picart T, Gautheron A, Caredda C, et al. Fluorescence-guided surgical techniques in adult diffuse low-grade gliomas: state-of-the-art and emerging techniques: a systematic review. *Cancers.* 2024;16(15):2698.
24. Ruzevick JJ, Strickland BA, Zada G. Commentary: Machine vision for real-time intraoperative anatomic guidance: a proof-of-concept study in endoscopic pituitary surgery. *Oper Neurosurg.* 2021;21(4):e302-e303.
25. Previc FH, Ercoline WR. *Spatial Disorientation in Aviation*. AIAA; 2004.
26. Ling B, Bo L, Bingzheng S, Lingcun Q, Chengqi X, Yafeng N. A cognitive study of multicolour coding for the head-up display (HUD) of fighter aircraft in multiple flight environments. *J Phys Conf Ser.* 2019;1215(1):012032.
27. Layard Horsfall H, Palmisciano P, Khan DZ, et al. Attitudes of the surgical team toward artificial intelligence in neurosurgery: international 2-stage cross-sectional survey. *World Neurosurg.* 2021;146:e724-e730.
28. Khan DZ, Luengo I, Barbarisi S, et al. Automated operative workflow analysis of endoscopic pituitary surgery using machine learning: development and preclinical evaluation (IDEAL stage 0). *J Neurosurg*. 2022;137(1):51–58.
29. Sarker SK, Chang A, Albrani T, Vincent C. Constructing hierarchical task analysis in surgery. *Surg Endosc.* 2008;22(1):107-111.

Acknowledgments

Conception and design of the work: Serra, Konukoglu, Sarwin, Staartjes. Acquisition data: Serra, Carretta, Mazzatenta, Zoli, Regli. Analysis and interpretation of data: Serra, Konukoglu, Sarwin, Staartjes. Drafting the work: Serra, Sarwin, Staartjes. Critical Revision of the Work for important intellectual content: Serra, Carretta, Mazzatenta, Zoli, Regli. Final approval of the version to be published: Serra. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: All Authors.

VIDEO 1. Extract from video hold out depicting the performance of the algorithm in the nasal phase of the surgical approach.

VIDEO 2. Extract from video hold out depicting the performance of the algorithm in the sphenoidal phase of the surgical approach.