A Multi-task Network for Anatomy Identification in Endoscopic Pituitary Surgery

Adrito Das[1(✉)], Danyal Z. Khan[1,2], Simon C. Williams[1,2], John G. Hanrahan[1,2], Anouk Borg[2], Neil L. Dorward[2], Sophia Bano[1,3], Hani J. Marcus[1,2], and Danail Stoyanov[1,3]

[1] Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London, UK
adrito.das.20@ucl.ac.uk
[2] Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London, UK
[3] Department of Computer Science, University College London, London, UK

Abstract. Pituitary tumours are in an anatomically dense region of the body, and often distort or encase the surrounding critical structures. This, in combination with anatomical variations and limitations imposed by endoscope technology, makes intra-operative identification and protection of these structures challenging. Advances in machine learning have allowed for the opportunity to automatically identifying these anatomical structures within operative videos. However, to the best of the authors' knowledge, this remains an unaddressed problem in the sellar phase of endoscopic pituitary surgery. In this paper, PAINet (Pituitary Anatomy Identification Network), a multi-task network capable of identifying the ten critical anatomical structures, is proposed. PAINet jointly learns: (1) the semantic segmentation of the two most prominent, largest, and frequently occurring structures (sella and clival recess); and (2) the centroid detection of the remaining eight less prominent, smaller, and less frequently occurring structures. PAINet utilises an EfficientNetB3 encoder and a U-Net++ decoder with a convolution layer for segmentation and pooling layer for detection. A dataset of 64-videos (635 images) were recorded, and annotated for anatomical structures through multi-round expert consensus. Implementing 5-fold cross-validation, PAINet achieved 66.1% and 54.1% IoU for sella and clival recess semantic segmentation respectively, and 53.2% MPCK-20% for centroid detection of the remaining eight structures, improving on single-task performances. This therefore demonstrates automated identification of anatomical critical structures in the sellar phase of endoscopic pituitary surgery is possible.

Keywords: minimally invasive surgery · semantic segmentation · surgical AI · surgical vision

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43996-4_45.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14228, pp. 472–482, 2023.
https://doi.org/10.1007/978-3-031-43996-4_45

1 Introduction

A difficulty faced by surgeons performing endoscopic pituitary surgery is identifying the areas of the bone which are safe to open. This is of particular importance during the sellar phase as there are several critical anatomical structures within close proximity of each other [9]. The sella, behind which the pituitary tumour is located, is safe to open. However, the smaller structures surrounding the sella, behind which the optic nerves and internal carotid arteries are located, carry greater risk. Failure to appreciate these critical parasellar neurovascular structures can lead to their injury, and adverse outcomes for the patient [9,11].

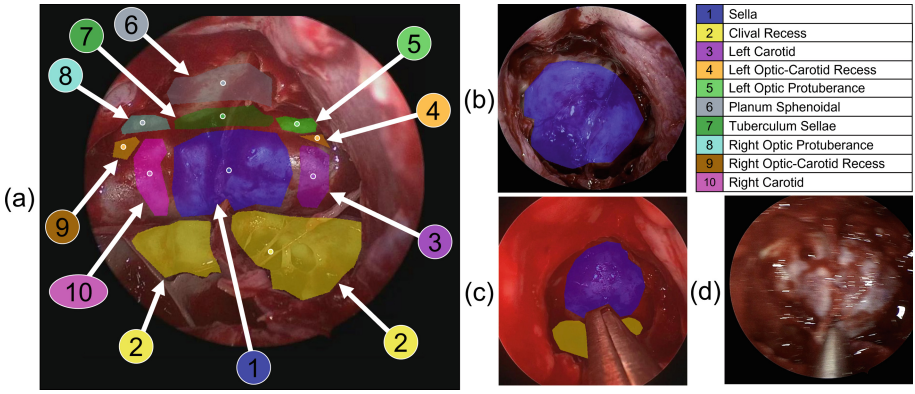

Fig. 1. 10-anatomical-structures semantic segmentation of the sellar phase in endoscopic pituitary surgery. Names and mask colour are given in the legend, with centroids displayed as dots. Example images where each visible critical anatomical structures' mask is annotated by neurosurgeons, are displayed: (a) an image where all structures are clearly seen; (b) an image where only the sella is visible; (c) an image where some structures are occluded by an instrument and biological factors; and (d) an image where none of the structures are identifiable due to blurriness caused by camera movement. (Color figure online)

The human identification of these structures relies on visual clues, inferred from the impressions made on the bone, rather than direct visualisations of the structures [11]. This is especially challenging as the pituitary tumour often compresses; distorts; or encases the surrounding structures [11]. Neurosurgeons utilise identification instruments, such as a stealth pointer or micro-doppler, to aid in this task [9]. However, once an identification instrument is removed, identification is lost upon re-entry with a different instrument, and so the identification can only be used in referenced to the more visible anatomical landmarks. Automatic identification from endoscopic vision may therefore aid surgeons in this effort while minimising disruption to the surgical workflow [11].

This is a challenging computer vision task due to the narrow camera angles enforced by minimally invasive surgery, which lead to: (i) structure occlusions by

instruments and biological factors (e.g., blood); and (ii) image blurring caused by rapid camera movements. Additionally, in this specific task there are: (iii) numerous small structures; (iv) visually similar structures; and (v) unclear structure boundaries. Hence, the task can be split into two sub-tasks to account for these difficulties in identification: (1) the semantic segmentation of the two larger, visually distinct, and frequently occurring structures (sella and clival recess); and (2) the centroid detection of the eight smaller structures (Fig. 1).

To solve both tasks simultaneously, PAINet (Pituitary Anatomy Identification Network) is proposed. This paper's contribution is therefore:

1. The automated identification of the ten critical anatomical structures in the sellar phase of endoscopic pituitary surgery. To the best of the authors' knowledge, this is the first work addressing the problem at this granularity.
2. The creation of PAINet, a multi-task neural network capable of simultaneously semantic segmentation and centroid detection of numerous anatomical structures within minimally invasive surgery. PAINet uniquely utilises two loss functions for improved performance over single-task neural networks due to the increased information gain from the complementary task.

2 Related Work

Encoder-decoder architectures are the leading models in semantic segmentation and landmark detection [4], with common architectures for anatomy identification including the U-Net and DeepLab families [6]. Improvements to these models include: adversarial training to limit biologically implausible predictions [14]; spatial-temporal transformers for scene understanding across consecutive frames [5]; transfer learning from similar anatomical structures [3]; and graph neural networks for global image understanding [2]. Multi-task networks improve on the baseline models by leveraging common characteristics between sub-tasks, increasing the total information provided to the network [15], and are effective at instrument segmentation in minimally invasive surgery [10].

The most clinically similar works to this paper are: (1) The semantic segmentation of 3-anatomical-structures in the nasal phase of endoscopic pituitary surgery [12]. Here, U-Net was weakly-supervised on centroids, outputting segmentation masks for each structure. Training on 18-videos (367-images), the model achieved statistically significant results ($P < 0.001$) on the hold-out testing dataset of 5-videos (182-images) when compared to a location prior baseline model [12]. (2) The semantic segmentation of: 2-zones (safe or dangerous); and 3-anatomical-structures in laparoscopic cholecystectomy [7]. Here, two PSPNets were fully-supervised on 290-videos (2627 images) using 10-fold cross-validation, achieving 62% mean intersection over union (MIoU) for the 2-zones; and 74% MIoU for the 3-structures [7]. These works are extended in this paper by increasing the number of anatomical structures and the identification granularity.

3 Methods

PAINet: A multi-task encoder-decoder network is proposed to improve performance by exchanging information between the semantic segmentation and centroid detection tasks. EfficientNetB3, pre-trained on ImageNet, is used as the encoder because of its accuracy, computational efficiency and proven generalisation capabilities [13]. The decoder is based on U-Net++, a state-of-the-art segmentation network widely used in medical applications [16]. The encoder-decoder architecture is modified to output both segmentation and centroid predictions by sending the decoder output into two separate layers: (1) a convolution for segmentation prediction; and (2) an average pooling layer for centroid prediction. Different loss functions were minimised for each sub-task (Fig. 2).

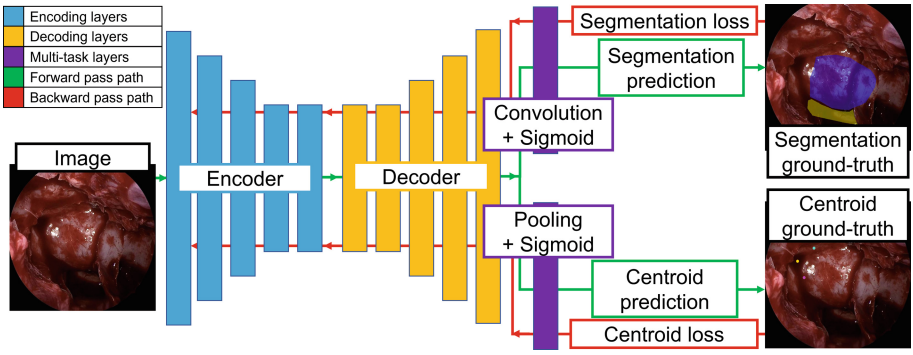

Fig. 2. Multi-task (semantic segmentation and centroid detection) architecture diagram. Notice the two output layers (convolution and pooling) and two loss functions.

Ablation studies and granular details are provided below. The priority was to find the optimal sella segmentation model, as it is required to be opened to access the pituitary tumour behind it, indicating the surgical "safe-zone".

Semantic Segmentation: First, single-class sella segmentation models were trialed. 8-encoders (pre-trained convolution neural networks) and 15-decoders were used, with their selection based off architecture variety. Two loss functions were also used: (1) distribution-based logits cross-entropy; and (2) region-based Jaccard loss. Boundary-based loss functions were not trialed as: (1) the boundary of the segmentation masks are not well-defined; and (2) in the cases of split structures (Fig. 1c), boundary-based loss functions are not appropriate [8]. The decoder output is passed through a convolution layer and sigmoid activation.

For multi-class sella and clival recess segmentation, the optimal single-class model was extended by: (1) sending through each class to the loss function separately (multi-class separate); and (2) sending both classes through together (multi-class together). An extension of logits cross-entropy, logits focal loss, was used instead as it accounts for data imbalance between classes.

Centroid Detection: 5-models were trialed: 3-models consisted of encoders with a convolution layer and linear activation; and 2-models consisted of encoder-decoders with an average pooling layer and sigmoid activation with 0.3 dropout. Two distance-based loss functions were trialed: (1) mean squared error (MSE); and (2) mean absolute error (MAE). Loss was calculated for all structures simultaneously as a 16 dimensional output (8 centroids $\times$ 2 coordinates) and set to 0 for a structure if ground-truth centroids of that structure was not present.

4 Experimental Setup

Evaluation Metrics: For sella segmentation the evaluation metric is intersection over union (IoU), as commonly used in the field [4,8]. For multi-class segmentation, the model that optimises clival recess IoU without reducing the previously established sella IoU is chosen. Precision and recall are also given.

For centroid detection, the evaluation metric is mean percentage of correct keypoints (MPCK) with the threshold set to 20%, indicating the mean number of predicted centroids falling within 144 pixels of the ground-truth centroid. This is commonly used in anatomical detection tasks as it ensures the predictions are close to the ground-truth while limiting overfitting [1]. MPCK-40% and MPCK-10%, along with the mean percentage of centroids that fall within their corresponding segmentation mask (mean percentage of centroid masks (MPCM)) are given as secondary metrics. For multi-task detection, MPCK-20% is optimised such that sella IoU does not drop from the previously established optimal IoU.

Network Parameters: 5-fold cross-validation was implemented with no hold-out testing. To account for structure data imbalance, images were randomly split such that the number of structures in each fold is approximately even. Images from a singular video were present in either the training or validation dataset.

Each model was run for with a batch size of 5 for 20 epochs, where the epoch with the best primary evaluation metric on the validation dataset was kept. The optimising method was Adam with varying initial learning rates, with a separate optimiser for each loss function during multi-task training.

All images were scaled to 736 $\times$ 1280 pixels for model compatibility, and training images were randomly augmented within the following parameters: shift in any direction by up to 10%; zooming in or out about the image center by up to 10%; rotation about the image center clockwise or anticlockwise by up to $\pi/6$; increasing or decreasing brightness, contrast, saturation, and hue by up to 10%.

The code is written in Python 3.8 using PyTorch 1.8.1, run on a single NVIDIA Tesla V100 Tensor Core 32-GB GPU using CUDA 11.2, and is available at https://github.com/dreets/pitnet-anat-public. For PAINet, a batch size of 5 utilised 29-GB and the runtime was approximately 5-min per epoch. Valuation runtime is under 0.1-s per image and therefore a real-time overlay on-top of the endoscope video feed is feasible intra-operatively.

5 Dataset Description

Images: Images come from 64-videos of endoscopic pituitary surgery where the sellar phase is present [9], recorded between 30 Aug 2018 and 20 Feb 2021 from The National Hospital of Neurology and Neurosurgery, London, United Kingdom. All patients have provided informed consent, and the study was registered with the local governance committee. A high-definition endoscope (Hopkins Telescope, Karl Storz Endoscopy) was used to record the surgeries at 24 frames per second (fps), with at least 720p resolution, and stored as mp4 files. 10-images corresponding to 10-s of the sellar phase immediately preceding sellotomy were extracted from each video at 1 fps, and stored as 720p png files. Video upload and annotation was performed using Touch Surgery^{TM} Enterprise.

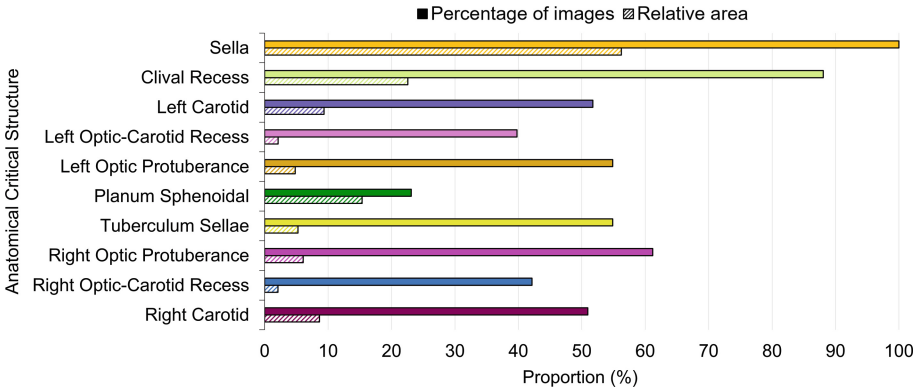

Fig. 3. Distribution of 10 anatomical structures in 635-images: The top bars display structure frequency; and the bottom bars display each structures area relative to the total area covered by all structures in a given image (mean-averaged across all images).

Annotations: Expert neurosurgeons identified 10-anatomical-structures as critical based on the literature (Fig. 1a) [9,11]. 640-images were manually segmented to obtain ground-truth segmentations. A two-stage process was used: (1) two neurosurgeons segmented each image, with any differences settled through discussion; (2) two consultant neurosurgeons independently peer-reviewed the segmentations. Only visible structures were annotated (Fig. 1b); if the structures were occluded, the segmentation boundaries were drawn around these occlusions (Fig. 1c); and if an image is too blurry to see the structures no segmentation boundaries were drawn - this excluded 5 images (Fig. 1d). The center of mass of each segmentation mask was defined as the centroid.

The sella is present in all 635-images (Fig. 3). Other than the clival recess, the remaining 8-structures are found in less than 65% of images, with planum sphenoidal found in less than 25% of images. Moreover, the area covered by these

8-structures are small, with several covering less than 10% of the total area covered by all structures in a given image. Furthermore, most smaller structures boundaries are ambiguous as they are hard to define even by expert neurosurgeons. This emphasizes the challenge of identifying smaller structure in computer vision, and supports the need for detection and multi-task solutions.

6 Results and Discussion

Quantitative evaluation is calculated for: single-class sella segmentation (Table 1); single-class, multi-class, and PAINet 2-structures segmentation (Table 2); multi-class, and PAINet 8-structures centroid detection (Tables 3 and 4).

The optimal model for single-class sella segmentation achieved 65.4% IoU, utilising an EfficientNetB3 encoder; U-Net++ decoder; Jaccard loss; and a 0.001 initial learning rate. Reductions in IoU are seen when alternative parameters are used, highlighting their impact on model performance.

Table 1. Selected models performance on the single-class sella segmentation task (5-fold cross validation). The model with the highest IoU is given in top-most row (in bold), with each row changing one model parameter (in italics), where the highest and lowest IoU for a given model parameter change is shown. Complete results for all models can be found in Supplementary Material Table 1. *Initial learning rate.

Decoder	Encoder	Loss	Rate*	IoU	Recall	Precision
U-Net++	**EfficientNetB3**	**Jaccard**	**0.001**	**65.4±1.6**	**79.3±2.6**	**79.5±3.9**
DeepLabv3+	EfficientNetB3	Jaccard	0.001	63.7 ± 2.9	77.0 ± 4.7	79.5 ± 2.1
PSPNet	EfficientNetB3	Cross-Entropy	0.001	54.5 ± 4.1	74.3 ± 1.9	68.4 ± 3.5
U-Net++	EfficientNetB3	*Cross-Entropy*	0.001	64.8 ± 2.3	79.4 ± 2.5	78.6 ± 2.6
U-Net++	*Xception*	Jaccard	0.001	60.1 ± 4.6	78.2 ± 2.9	77.9 ± 4.3
U-Net++	*ResNet18*	Jaccard	0.001	56.3 ± 3.2	73.3 ± 4.0	74.2 ± 5.2
U-Net++	EfficientNetB3	Jaccard	*0.0001*	63.4 ± 1.3	72.7 ± 4.7	83.6 ± 6.3
U-Net++	EfficientNetB3	Jaccard	*0.01*	34.7 ± 9.5	68.7 ± 8.2	69.6 ± 8.8

Table 2. Selected models performance for sella and clival recess segmentation (5-fold cross-validation). All models use an EfficientNetB3 encoder and U-Net++ decoder. The best performing network, as determined by sella IoU, is displayed in bold.

Training	Loss	Sella IoU	Clival Recess IoU
Single-Class	Jaccard	65.4 ± 1.6	53.4 ± 5.9
Multi-Class Separate	Focal	65.4 ± 2.1	54.2 ± 8.5
Multi-Class Together	Focal	65.6 ± 2.6	49.9 ± 4.8
Multi-task (PAINet)	**Focal**	**66.1 ± 2.3**	**54.1 ± 5.0**

Using the optimal sella model configuration, 53.4% IoU is achieved for single-class clival recess segmentation. Extending this to multi-class and PAINet training improves both sella and clival recess IoU to 66.1% and 54.1% respectively.

The optimal model for centroid detection achieves 51.7% MPCK-20%, with minor deviations during model parameter changes. This model, ResNet18 with MSE loss, outperforms the more sophisticated models, as these models over-learn image features in the training dataset. However, PAINet leverages the additional information from segmentation masks to achieve an improved 53.2%.

The per structure PCK-20% indicate performance is positively correlated with the number of images where the structure is present. This implies the limiting factor is the number of images rather than architectural design.

Table 3. Selected models performance for centroid detection (5-fold cross-validation). The model with the highest MPCK-20% is given in the last row (in bold). Complete results for all models can be found in Supplementary Material Table 2.

Training	Decoder	Encoder	Loss	MPCK-20	MPCK-40	MPCK-10	MPCM
Multi-Class	–	ResNet18	MSE	51.7 ± 9.2	58.0 ± 7.0	28.4 ± 5.9	09.7 ± 2.8
Multi-Class	–	ResNet18	*MAE*	51.4 ± 9.4	57.9 ± 7.8	34.2 ± 5.3	10.9 ± 2.5
Multi-Class	–	EfficientNetB3	MSE	46.8 ± 8.3	55.5 ± 5.7	19.5 ± 3.2	06.3 ± 3.0
Multi-Class	DeepLabv3+	ResNet18	MSE	50.4 ± 6.2	57.8 ± 5.8	20.1 ± 5.4	06.5 ± 1.8
Multi-Class	U-Net++	ResNet18	MSE	50.1 ± 4.2	58.0 ± 6.8	26.1 ± 2.7	09.7 ± 1.2
Multi-Class	U-Net++	EfficientNetB3	MSE	48.1 ± 3.7	56.2 ± 4.7	26.4 ± 4.8	06.8 ± 1.6
PAINet	**U-Net++**	**EfficientNetB3**	**MSE**	$\mathbf{53.2 \pm 5.9}$	$\mathbf{58.0 \pm 6.9}$	$\mathbf{39.6 \pm 3.2}$	$\mathbf{13.4 \pm 2.9}$

Table 4. PAINet 8-structures centroid detection performance (5-fold cross validation).

Structure	Left Carotid	Left Optic-Carotid Recess	Left Optic Protuberance	Planum Sphenoidale	Tuberculum Sellae	Right Optic Protuberance	Right Optic-Carotid Recess	Right Carotid
PCK-20%	68.3 ± 9.2	37.6 ± 4.8	53.9 ± 7.3	27.6 ± 2.4	76.1 ± 2.8	72.0 ± 9.1	34.8 ± 3.0	55.4 ± 8.5

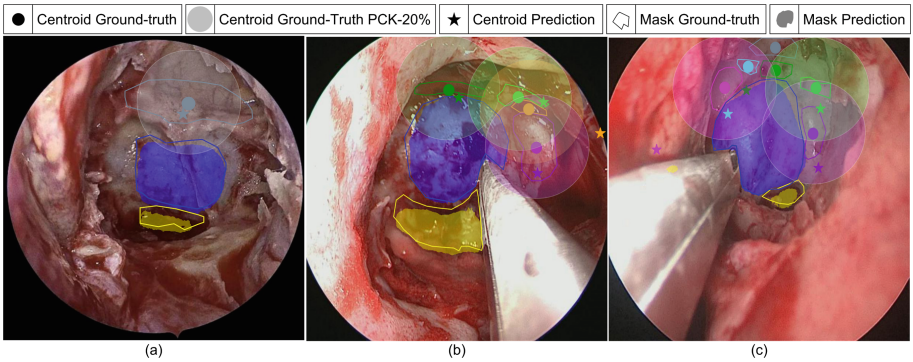

Fig. 4. PAINet predictions for three videos, displaying images with: (a) strong; (b) typical; and (c) poor performances. (The color map is given in Fig. 1.)

Qualitative predictions of the best performing model, PAINet, are displayed in Fig. 4. The segmentation predictions look strong, with small gaps from the ground-truth. However, this is expected as structure boundaries are not well-defined. The centroid predictions are weaker: in (a) the planum sphenoidale (grey) is predicted within the segmentation mask; in (b) three structures are within their segmentation mask, but the left optic-carotid recess (orange) is predicted in a biologically implausible location; and in (c) this is repeated for the right carotid (pink) and no structures are within their segmentation masks.

7 Conclusion

Identification of critical anatomical structures by neurosurgeons during endoscopic pituitary surgery remains a challenging task. In this paper, the potential of automating anatomical structure identification during surgery was shown. The proposed multi-task network, PAINet, designed to incorporate identification of both large prominent structures and numerous smaller less prominent structures, was trained on images of the sellar phase of endoscopic pituitary surgery. Using 635-images from 64-surgeries annotated by expert neurosurgeons and various model configurations, the robustness of the PAINet was shown over single task networks. PAINet achieved 66.1% (+0.7%) and 54.1% IoU (+0.7%) for sella and clival recess segmentation respectively, a higher performance than other minimally invasive surgeries [7]. PAINet also achieved 53.2% MPCK-20% (+1.5%) for detection of the remaining 8-structures. The most important structures to identify and avoid, the carotids and optic protuberances, have high performance, and therefore demonstrate the success of PAINet. This performance is greater than similar studies in endoscopic pituitary surgery for different structures [12] but lower than anatomical detection in other surgeries [1]. Collecting data from more pituitary surgeries will support incorporating anatomy variations and achieving generalisability. Furthermore, introducing modifications to the model architecture, such as the use of temporal networks [5], will further boost performance required for real-time video clinical translation.

Acknowledgements. This research was funded in whole, or in part, by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145/Z/16/Z]; the Engineering and Physical Sciences Research Council (EPSRC) [EP/P027938/1, EP/R004080/1, EP/P012841/1, EP/W00805X/1]; and the Royal Academy of Engineering Chair in Emerging Technologies Scheme. AD is supported by EPSRC [EP/S021612/1]. HJM is supported by WEISS [NS/A000050/1] and by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at University College London (UCL). DZK and JGH are supported by the NIHR Academic Clinical Fellowship. DZK is supported by the Cancer Research UK (CRUK) Predoctoral Fellowship. With thanks to Digital Surgery Ltd, a Medtronic company, for access to Touch Surgery™ Enterprise for both video recording and storage.

References

1. Danks, R.P., et al.: Automating periodontal bone loss measurement via dental landmark localisation. Int. J. Comput. Assist. Radiol. Surg. **16**(7), 1189–1199 (2021). https://doi.org/10.1007/s11548-021-02431-z
2. Gaggion, N., Mansilla, L., Mosquera, C., Milone, D.H., Ferrante, E.: Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: applications to chest x-ray analysis. IEEE Trans. Med. Imaging **42**(2), 546–556 (2023). https://doi.org/10.1109/tmi.2022.3224660
3. Gu, R., et al.: Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures. IEEE Trans. Med. Imaging **42**(1), 245–256 (2023). https://doi.org/10.1109/tmi.2022.3209798
4. Hao, S., Zhou, Y., Guo, Y.: A brief survey on semantic segmentation with deep learning. Neurocomputing **406**, 302–321 (2020). https://doi.org/10.1016/j.neucom.2019.11.118
5. Jin, Y., Yu, Y., Chen, C., Zhao, Z., Heng, P.A., Stoyanov, D.: Exploring intra- and inter-video relation for surgical semantic scene segmentation. IEEE Trans. Med. Imaging **41**(11), 2991–3002 (2022). https://doi.org/10.1109/tmi.2022.3177077
6. Liu, L., Wolterink, J.M., Brune, C., Veldhuis, R.N.J.: Anatomy-aided deep learning for medical image segmentation: a review. Phys. Med. Biol. **66**(11), 11TR01 (2021). https://doi.org/10.1088/1361-6560/abfbf4
7. Madani, A., et al.: Artificial intelligence for intraoperative guidance using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. Ann. Surg. **276**(2), 363–369 (2020). https://doi.org/10.1097/sla.0000000000004594
8. Maier-Hein, L., Reinke, A., Godau, P., et al.: Metrics reloaded: pitfalls and recommendations for image analysis validation (2022). https://doi.org/10.48550/arxiv.2206.01653
9. Marcus, H.J., et al.: Pituitary society expert Delphi consensus: operative workflow in endoscopic transsphenoidal pituitary adenoma resection. Pituitary **24**(6), 839–853 (2021). https://doi.org/10.1007/s11102-021-01162-3
10. Marullo, G., Tanzi, L., Ulrich, L., Porpiglia, F., Vezzetti, E.: A multi-task convolutional neural network for semantic segmentation and event detection in laparoscopic surgery. J. Personal. Med. **13**(3), 413 (2023). https://doi.org/10.3390/jpm13030413
11. Patel, C.R., Fernandez-Miranda, J.C., Wang, W.H., Wang, E.W.: Skull base anatomy. Otolaryngol. Clin. North Am. **49**(1), 9–20 (2016). https://doi.org/10.1016/j.otc.2015.09.001
12. Staartjes, V.E., 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. **21**(4), 242–247 (2021). https://doi.org/10.1093/ons/opab187
13. Tan, M., Le, Q.V.: EfficientNet: rethinking model scaling for convolutional neural networks. arXiv (2019). https://doi.org/10.48550/ARXIV.1905.11946
14. Wang, P., Peng, J., Pedersoli, M., Zhou, Y., Zhang, C., Desrosiers, C.: CAT: constrained adversarial training for anatomically-plausible semi-supervised segmentation. IEEE Trans. Med. Imaging, 1 (2023). https://doi.org/10.1109/tmi.2023.3243069

15. Zhang, Y., Yang, Q.: An overview of multi-task learning. Natl. Sci. Rev. **5**(1), 30–43 (2017). https://doi.org/10.1093/nsr/nwx105
16. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1