Device-Constrained Real-Time EVD Catheter Segmentation

Mustafa Seddiqi✉, Joshua Castillo, Tiberiu Popa, and Marta Kersten-Oertel

Concordia University, Montreal, Canada
mustafa.seddiqi@concordia.ca

Abstract. Ventriculostomy, or external ventricular drain (EVD) placement, is frequently performed freehand in urgent bedside settings, resulting in high catheter misplacement rates. Portable augmented reality (AR) guidance offers a low-cost alternative to conventional navigation systems, but mobile and standalone XR hardware impose strict latency, memory, and power constraints. A key enabling component is reliable real-time segmentation of the thin, low-contrast EVD catheter in unconstrained open-field environments. To address this challenge, we propose a real-time EVD catheter segmentation framework for mobile and XR deployment that includes a structure-aware loss promoting consistency along the catheter trajectory. We construct a scene-structured bedside dataset and complement it with a 3D-driven synthetic data generation pipeline to expand geometric and viewpoint variability. Lightweight backbones are evaluated under real-only, synthetic-only, and combined supervision, with systematic cross-domain and on-device analysis of robustness and accuracy–latency trade-offs under mobile/XR hardware constraints. Our results establish the feasibility of real-time catheter segmentation as a key component of future low-cost AR-guided ventriculostomy workflows. Code is available at: https://github.com/m-seddiqi/evd-catheter-segmentation.

Keywords: Ventriculostomy · Synthetic Data · Thin-Structure Segmentation · Real-Time Inference

1 Introduction

Ventriculostomy, or external ventricular drain (EVD) placement, is one of the most common neurosurgical procedures and is typically performed freehand in emergency and acute care settings [3]. Catheter misplacement remains prevalent and may require repeated insertion attempts with associated complications [17,11]. Guidance aids, including mechanical guides, ultrasound-assisted placement, and navigation-based systems, have been investigated for EVD placement [16,15,19]. However, conventional image-guided techniques can be time-consuming and expensive, limiting their use in urgent bedside settings [19].

Augmented reality (AR) guidance offers a low-cost alternative by overlaying the planned catheter trajectory and ventricular anatomy directly onto the surgical field [20,12,5,4,2]. However, mobile and standalone XR hardware impose

strict latency, memory, and power constraints. Unlike marker-based tracking approaches, our goal is to localize the catheter directly from RGB imagery without fiducial markers, making segmentation a key enabling component.

Achieving this is challenging. Compared with laparoscopic tool segmentation and ultrasound needle tracking [1,6,22,8], catheter segmentation targets a thin, low-contrast, elongated structure occupying a small fraction of the image which is frequently affected by motion blur, specular reflections, and occlusion. Small local segmentation errors can lead to substantial errors in the estimated catheter trajectory, making geometric continuity essential for AR guidance. While foundation models such as SAM provide strong general-purpose segmentation capabilities [18,7], their computational footprint limits practical use on embedded mobile and XR platforms. Furthermore, no publicly available dataset exists for EVD catheter segmentation, and limited real data increases the risk of overfitting under distribution shift.

To our knowledge, prior work has not examined real-time EVD catheter segmentation for mobile and XR deployment, where thin-structure segmentation must be developed and evaluated under the resource constraints of embedded platforms. In this work, we make three contributions. First, we formulate real-time EVD catheter segmentation for mobile and XR deployment and introduce a structure-aware loss that encourages catheter-axis consistency in elongated predictions. Second, we construct a scene-structured bedside dataset with strict session-level partitioning and complement it with a 3D-driven synthetic data generation pipeline to expand geometric and viewpoint variability. Third, we perform systematic cross-domain and on-device evaluation across real-only, synthetic-only, and combined supervision regimes to characterize robustness and accuracy–latency trade-offs under mobile/XR hardware constraints.

2 Methods

2.1 Problem Setting and Deployment Constraints

We formulate catheter segmentation for AR-guided ventriculostomy as a fully automatic, real-time binary segmentation task under device-level constraints. Given an RGB frame $I \in \mathbb{R}^{H\times W\times 3}$ captured from a mobile or head-mounted device (HMD), the objective is to predict a dense catheter mask $M \in \{0,1\}^{H\times W}$. Beyond pixel-wise accuracy, predictions must preserve geometric continuity to ensure stable AR overlays and reliable trajectory visualization. This setting is deployment-driven rather than benchmark-driven. Inference must operate continuously at video rate on embedded mobile GPUs with limited memory, compute, and power budgets. Consequently, model evaluation jointly considers segmentation fidelity and runtime efficiency.

Although large foundation models such as SAM demonstrate strong cross-domain generalization [18,7], their computational footprint limits practical real-time XR deployment. In addition, the thin, low-contrast geometry of EVD catheters leads to severe class imbalance and fragmentation, motivating task-specific optimization under efficiency constraints. We therefore adopt lightweight

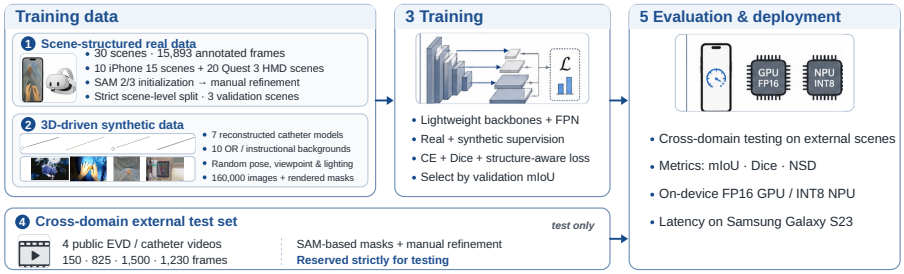

Fig. 1. Overview of real and synthetic data generation, combined-supervision training, cross-domain testing, and FP16/INT8 on-device evaluation.

backbones with a feature pyramid decoder to balance capacity and cost within mobile inference budgets. Fig. 1 summarizes the pipeline.

2.2 Dataset Creation

To evaluate catheter segmentation under simulated bedside conditions within a controlled laboratory environment, we constructed a scene-structured dataset utilizing a handheld mobile device and a head-mounted display. Each scene captures a continuous catheter manipulation session, preserving temporal motion and viewpoint variation while preventing frame-level redundancy across dataset splits. All recordings were conducted without patient involvement and contain no protected health information.

Data collection spanned two platforms to maximize diversity. Ten scenes were recorded using an iPhone 15 under unconstrained room conditions to introduce variability in illumination and background complexity. An additional 20 scenes were captured via a Meta Quest 3 headset, providing synchronized stereo image pairs from a surgeon-centric perspective consistent with the intended AR deployment. In total, the dataset comprises 30 scenes and 15,893 annotated frames.

To prevent temporal leakage, we reserved three scenes (two stereo and one monocular) for validation and used the remaining scenes for training; external videos were strictly test-only. Ground-truth masks were initialized using SAM 2/3 and subsequently manually corrected and quality-checked, particularly in regions affected by motion blur and low contrast.

2.3 Augmentation Strategy for Domain Generalization

To improve robustness under cross-domain variation, we employ a task-aware photometric augmentation pipeline during training. All photometric transformations are applied online to the input image, while ground-truth masks remain unchanged. Standard global appearance perturbations, including Gaussian blur, color jitter, and exposure adjustments, simulate illumination variability and camera differences across recording environments. However, generic photometric

augmentation is insufficient for thin-structure segmentation in open-field neurosurgical settings. The EVD catheter is particularly susceptible to motion-induced blur, specular saturation from surgical lighting, and local contrast degradation, all of which can cause spatially fragmented prediction masks.

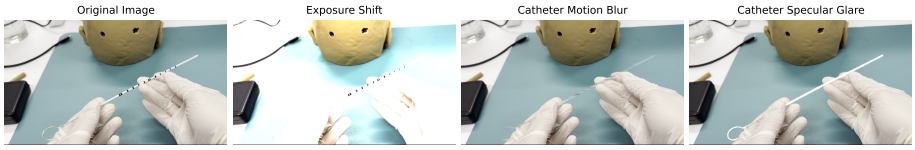

Fig. 2. Examples of the proposed photometric augmentations, including exposure variation, catheter-specific directional motion blur, and synthetic specular highlights.

To explicitly model these failure modes, we introduce catheter-aware perturbations applied selectively to the annotated catheter region. These include directional motion blur aligned with the catheter axis and synthetic specular highlights that emulate strong reflections from overhead lights. By targeting the structure of interest rather than applying image-wide perturbations, this strategy encourages robustness to artifacts that disproportionately affect slender geometries. The pipeline combines standard global photometric perturbations with catheter-aware motion blur and specular highlights to model thin-structure degradation under surgical lighting and motion while preserving mask geometry. Augmentation parameters were fixed across experiments and are released with the code. Fig. 2 illustrates representative effects.

2.4 Synthetic Dataset Generation via 3D Modeling

Photometric augmentation alone does not sufficiently expand geometric diversity in viewpoint, catheter pose, and spatial context. Because thin, elongated structures are highly sensitive to pose and orientation, limited real-world data often leads to overfitting specific viewing configurations. To address this, we introduce a 3D-driven synthetic data generation pipeline inspired by recent 2D-to-3D reconstruction approaches [9]. Specifically, we constructed a library of seven 3D catheter models, reconstructed from publicly available imagery using a single-image reconstruction pipeline followed by manual refinement. These models were aligned and scale-corrected to ensure consistent physical proportions and canonical orientation. Although limited in number and containing minor reconstruction artifacts, they provide meaningful, clinically relevant variations in curvature, diameter, and tip morphology, enabling controlled sampling of pose and viewpoint.

To complement this geometric variation with environmental diversity, we curated ten background images depicting operating-room and instructional settings. Sourced from permissively licensed repositories, these backgrounds contain

no identifiable patient information. Representative reconstructed catheter models and background assets are shown in Fig. 1.

Synthetic samples are generated by pairing a reconstructed catheter model with a randomly selected background, rendering it under randomized rigid transformations (rotation and translation) before compositing. To mimic continuous manipulation while preserving temporal coherence, short video sequences are produced via incremental pose perturbations across consecutive frames. Optional illumination perturbations are also applied to enhance visual realism.

In total, 160,000 synthetic images were generated across diverse catheter and background combinations. Because ground-truth masks are derived directly from rendered 3D geometry, they avoid manual annotation noise. Synthetic rendering used fixed configuration ranges for catheter pose, scale/depth, camera clipping, and tri-light illumination; complete generation settings are released with the code. This synthetic dataset supports controlled evaluation under real-only, synthetic-only, and mixed supervision paradigms, expanding both geometric and contextual diversity beyond our laboratory collection and supporting improved robustness to cross-domain shifts.

3 Experiments

3.1 Experimental Setup

Training Protocol. Main models were trained for 20 epochs using 512×512 and 384×640 crop buckets (batch size 4) and AdamW (LR 10^{-4}, 0.1× for the backbone; weight decay 10^{-4}). In the real+synthetic regime, each epoch used a synthetic shard together with a 20% subsample of the real training frames; exact shard and split manifests are released with the code. Training images were resized with aspect ratio preserved and randomly cropped to the selected bucket; validation and test frames used the closest-aspect bucket with letterbox padding. Image and mask resizing used bilinear and nearest-neighbor interpolation, respectively.

Architectures. We adopt `FastViT-T8` [21] with an FPN decoder [13] as the default backbone due to its accuracy–efficiency trade-off for mobile deployment. We also evaluate `FastViT-SA12`, `FastViT-SA36` [21], `MobileNetV3-Large` [10], and `ConvNeXt-Tiny` [14], all with the same decoder.

Loss Function. We optimize the composite objective $\mathcal{L} = \alpha\mathcal{L}_{\text{CE}} + \beta\mathcal{L}_{\text{Dice}} + \lambda\mathcal{L}_{\text{struct}}$, where the weights are set to $\alpha = 0.35$, $\beta = 0.65$, and $\lambda = 0.02$. The cross-entropy term uses a background weight of 0.5, while the Dice term provides foreground-overlap supervision. The structure-aware term complements these pixel-wise objectives by constraining the predicted response relative to the dominant catheter geometry. Given the ground-truth mask G, we estimate its principal axis through covariance analysis of the foreground coordinates and compute the absolute perpendicular distance of each pixel from this axis. Let μ_G, σ_G denote the mean and standard deviation of these distances over the ground-truth foreground, and let μ_P, σ_P denote the corresponding probability-weighted statistics for the predicted foreground probabilities P. We define $\mathcal{L}_{\text{struct}} = \rho((\mu_P -$

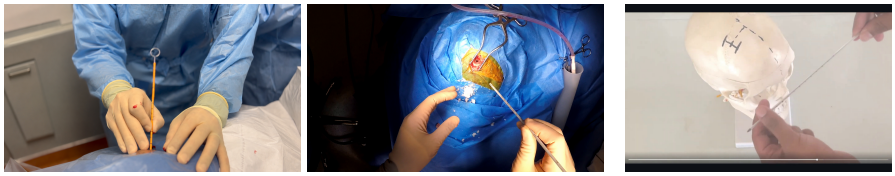

Fig. 3. Representative frames from three of the four external evaluation scenes. Scene 1: real, non-surgeon view (150 frames); Scene 2: real, surgeon-aligned view (825 frames); Scene 3: demonstration, non-surgeon view (1500 frames); Scene 4 (not shown): demonstration, non-surgeon view (1230 frames). Sequences vary substantially in viewpoint, catheter appearance, motion blur, and background relative to our data.

$\mu_G)/\max(\mu_G, 1)) + \gamma\,\rho((\sigma_P - \sigma_G)/\max(\sigma_G, 1))$, where ρ denotes the Smooth L1 penalty and $\gamma = 0.5$. Matching these statistics encourages the predicted foreground to remain concentrated around the ground-truth catheter axis, penalizing off-axis spread and fragmented responses that deviate from the dominant catheter geometry. This provides a lightweight, catheter-specific geometric regularizer that complements the cross-entropy and Dice objectives.

Metrics and Model Selection. We report mean Intersection over Union (mIoU), Dice, and Normalized Surface Dice (NSD) on the foreground class, with NSD computed using a 2-pixel surface tolerance. Models are selected by validation mIoU; Dice and NSD provide complementary overlap- and boundary-sensitive evaluation.

3.2 Cross-Domain Evaluation on External Videos

To evaluate robustness under distribution shift, we test on four publicly available videos depicting EVD placement or catheter manipulation, hereafter referred to as Scenes 1–4. These sequences are strictly reserved for testing and differ substantially from our collected data in viewpoint (surgeon-aligned vs. non-surgeon), catheter morphology and scale, motion blur, and background context (illustrated in Fig. 3). These external recordings are utilized solely for evaluation and are not redistributed. Furthermore, no identifiable patient information is retained in our analysis, and frames are presented strictly for illustrative purposes.

Ground-truth masks for all evaluated frames are generated using the same semi-automated annotation pipeline described in Section 2.2, consisting of SAM-based initialization followed by manual refinement. We report per-scene mIoU, Dice, and NSD, computing a macro-average across scenes (with equal weight per sequence) to avoid bias toward longer recordings. This protocol enables a systematic comparison of supervision strategies and backbone architectures under realistic cross-domain shift.

3.3 Training Regimes and Backbone Trade-offs

Supervision regime. Table 1 summarizes the cross-domain comparison across supervision regimes and backbone architectures. Combined supervision yields

Table 1. Cross-domain performance (%) of backbone architectures across four external videos (mean ± std). Unless otherwise indicated, models are trained with combined supervision (real + synthetic). * Real-only supervision. † Synthetic-only supervision. Best macro-average per metric is highlighted in bold.

Backbone	Scene 1			Scene 2			Scene 3			Scene 4			Avg		
	mIoU ↑	Dice ↑	NSD ↑	mIoU ↑	Dice ↑	NSD ↑	mIoU ↑	Dice ↑	NSD ↑	mIoU ↑	Dice ↑	NSD ↑	mIoU ↑	Dice ↑	NSD ↑
FastViT-T8*	51.1±1.3	67.6±1.1	56.9±1.0	46.0±0.4	63.0±0.4	62.0±1.3	29.4±8.2	45.1±9.9	33.3±6.4	31.6±4.1	47.9±4.7	53.9±4.4	39.5±3.3	55.9±3.8	51.5±2.7
FastViT-T8†	70.1±4.5	82.3±3.1	71.5±2.3	55.6±0.8	71.5±0.7	73.0±0.6	22.0±6.0	35.9±8.1	27.9±5.8	29.7±6.9	45.6±8.2	44.1±11.5	44.3±0.7	58.8±0.6	54.1±1.0
FastViT-T8	68.0±3.8	80.9±2.7	70.8±3.1	67.8±1.5	80.8±1.0	82.1±1.2	31.7±4.7	48.0±5.4	36.5±3.0	44.0±1.7	61.1±1.6	68.4±0.8	52.9±1.3	67.7±1.4	64.4±1.0
FastViT-SA12	74.2±4.9	85.2±3.2	71.6±3.0	70.6±0.5	82.8±0.3	83.4±0.2	36.9±0.6	53.9±0.7	42.6±1.2	32.2±2.0	48.7±2.3	52.5±3.9	53.5±0.4	67.6±0.0	62.5±0.6
MobileNetV3-L	57.0±1.3	72.6±1.1	60.0±0.9	56.9±2.3	72.5±1.8	73.8±1.6	19.7±3.1	32.9±4.3	25.8±2.7	34.9±1.1	51.7±1.2	53.5±0.9	42.1±0.8	57.4±1.2	53.3±0.7
FastViT-SA36	73.4±5.2	84.6±3.4	72.6±4.1	63.1±5.2	77.3±3.9	79.4±3.1	51.4±5.1	67.8±4.4	60.7±4.3	40.6±0.3	57.7±0.3	68.6±1.3	57.1±1.2	71.9±0.9	70.3±1.7
ConvNeXt-T	80.0±0.2	88.9±0.2	77.4±0.2	73.6±2.3	84.8±1.5	87.0±1.1	44.0±4.0	61.1±3.8	48.2±2.3	50.1±2.3	66.7±2.0	78.3±2.4	**61.9±2.2**	**75.4±1.9**	**72.7±1.5**

the strongest and most consistent macro-average across metrics, indicating improved robustness to viewpoint and appearance shift. Synthetic-only training performs competitively when geometric variation is well covered, but degrades on sequences with appearance artifacts not captured by simulation. Real-only training underperforms overall, suggesting limited diversity in real data alone. These results support a complementary effect: synthetic data expands geometric coverage, while real data anchors appearance realism.

Backbone capacity. Under combined supervision, higher-capacity backbones improve cross-domain robustness, particularly in NSD, reflecting better structural continuity of the thin catheter. `ConvNeXt-Tiny` achieves the highest macro-average, with `FastViT-SA36` close behind and strong under pronounced viewpoint shifts. Lighter models (`FastViT-T8`, `MobileNetV3-Large`) show larger drops on the most shifted videos, highlighting limits in representational capacity.

Quantization and On-Device Deployment. INT8 models were calibration-quantized from ONNX and compiled through Qualcomm AI Hub at 384×640 (batch size 1) for TFLite/NPU inference. Latency was measured for model inference only, excluding preprocessing and postprocessing, after warm-up, and the minimum per-image latency is reported. Table 2 reports Scene 1 accuracy and latency under FP32, FP16 (GPU), and INT8 (NPU). FP16 largely preserves FP32 accuracy across architectures, making it a practical reduced-precision option. INT8 substantially reduces latency, but accuracy degradation is architecture-dependent: `ConvNeXt-Tiny` remains stable, whereas `FastViT-T8` exhibits larger drops, particularly in NSD. These results support the feasibility of using the segmentation output in a future markerless mobile/XR guidance pipeline, where catheter masks could support downstream stereo trajectory reconstruction.

Table 2. Accuracy (%) and on-device latency on Scene 1 under different numerical precisions. FP16 (GPU) and INT8 (NPU) models are benchmarked on Samsung Galaxy S23. Latency denotes minimum per-image inference time.

Backbone	Params (M)	FP32 (Scene 1)			FP16 GPU (Scene 1)				INT8 NPU (Scene 1)			
		mIoU ↑	Dice ↑	NSD ↑	mIoU ↑	Dice ↑	NSD ↑	Lat. (ms)	mIoU ↑	Dice ↑	NSD ↑	Lat. (ms)
FastViT-T8	8.93	65.3	79.0	68.6	61.3	76.0	65.9	21.5	37.1	54.1	50.4	5.8
MobileNetV3-L	7.98	57.9	73.4	60.6	58.0	73.4	60.6	20.5	55.2	71.1	59.2	6.3
FastViT-SA12	16.55	77.7	87.5	73.7	79.1	88.3	74.0	22.4	72.3	83.9	71.0	6.4
FastViT-SA36	36.49	69.7	82.2	69.7	74.3	85.2	72.7	29.5	69.9	82.3	71.0	10.8
ConvNeXt-T	32.91	80.2	89.0	77.5	80.1	88.9	77.1	27.0	80.2	89.0	77.0	13.5

3.4 Ablation and Failure Analysis

Ablation on Scene 2. All configurations were trained for 10 epochs, and the checkpoint with the highest validation mIoU was selected. Using `FastViT-T8`, we progressively introduce each component and report Scene 2 test mIoU: weak augmentation (43.0) → strong augmentation (50.8) → +structure loss (56.1) → +catheter-aligned motion blur (57.7) → +random exposure (53.8). While validation mIoU increases monotonically across additions, the final configuration shows reduced performance on Scene 2. This indicates that aggressive global exposure perturbations, despite improving validation mIoU, may attenuate already low catheter contrast in real surgical footage under challenging illumination.
Structure-aware regularization. Smaller backbones often misclassify bright linear edges as catheter, producing spurious elongated masks when the catheter is absent or barely visible. The structure-aware term reduces these failures by matching predicted and ground-truth normal-distance statistics around the catheter axis, encouraging catheter-axis consistency and penalizing off-axis fragmented responses; larger backbones exhibit fewer such errors.
Low-contrast collapse. All models may fail when the catheter becomes near-transparent, occasionally collapsing to an empty mask. This reflects the inherent sensitivity of thin-structure segmentation to extreme contrast degradation.
Foundation-model instability under blur. Prompt-based foundation models may fail at initialization under strong motion blur, yielding missed or distorted masks. In the same frames, task-specific backbones, particularly larger ones, remain more stable, indicating better robustness to deployment-level blur.

4 Discussion and Conclusion

We formulated EVD catheter segmentation as a thin-structure segmentation problem under limited data and mobile deployment constraints, motivated by AR-guided ventriculostomy. Unlike conventional segmentation benchmarks, this setting requires not only high overlap accuracy but also longitudinal geometric consistency while operating within strict latency and memory budgets. We therefore evaluated segmentation performance jointly with cross-domain robustness and on-device efficiency.

Cross-domain evaluation demonstrates that thin-structure segmentation is particularly sensitive to distribution shift. While real-only training exhibited limited generalization and synthetic-only supervision lacked sufficient appearance realism, combining both sources consistently produced the strongest performance, suggesting that geometric diversity from 3D-generated data complements real-world appearance cues. Larger-capacity backbones further improved overlap and boundary stability, albeit at increased computational cost, emphasizing the importance of balancing segmentation accuracy with deployment efficiency. Finally, the proposed structure-aware loss improved segmentation performance by encouraging catheter-axis consistency, reducing off-axis fragmentation errors that could otherwise destabilize downstream AR trajectory estimation.

Our study has several limitations. The dataset is relatively small and was collected in a controlled, non-clinical environment. Moreover, we evaluated the segmentation component in isolation rather than within a complete AR guidance system. Integration with an AR platform, together with clinical and usability evaluation, remains future work.

In summary, we demonstrate that reliable real-time EVD catheter segmentation can be achieved within realistic mobile inference budgets. These results support the feasibility of deploying catheter segmentation as a core component of future low-cost AR-guided EVD systems and provide a deployment-oriented benchmark for thin-structure segmentation under resource constraints.

Acknowledgments. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPIN-2021-03477 and RGPIN-2017-06722).

Disclosure of Interests. The authors declare no competing interests.

References

1. Ahmed, F.A., Yousef, M., Ahmed, M.A., Ali, H.O., Mahboob, A., Ali, H., Shah, Z., Aboumarzouk, O., Al Ansari, A., Balakrishnan, S.: Deep learning for surgical instrument recognition and segmentation in robotic-assisted surgeries: a systematic review. Artificial Intelligence Review **58**, 1 (2025). https://doi.org/10.1007/s10462-024-10979-w
2. de Almeida, A.G.C., Fernandes de Oliveira Santos, B., Oliveira, J.L.M.: A neuronavigation system using a mobile augmented reality solution. World Neurosurgery **167**, e1261–e1267 (2022). https://doi.org/10.1016/j.wneu.2022.09.014
3. Amoo, M., Henry, J., Javadpour, M.: Common trajectories for freehand frontal ventriculostomy: a systematic review. World Neurosurgery **146**, 292–297 (2021). https://doi.org/10.1016/j.wneu.2020.11.065
4. Asadi, Z., Castillo, J.P., Asadi, M., Sinclair, D.S., Kersten-Oertel, M.: iSurgARy: a mobile augmented reality solution for ventriculostomy in resource-limited settings. Healthcare Technology Letters **12**, e12118 (2025). https://doi.org/10.1049/htl2.12118
5. Bagher Zadeh Ansari, N., Léger, É., Kersten-Oertel, M.: VentroAR: an augmented reality platform for ventriculostomy using the Microsoft HoloLens. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization **11**(4), 1225–1233 (2023). https://doi.org/10.1080/21681163.2022.2156394
6. Bakiya, K., Savarimuthu, N.: Transfer learning for surgical instrument segmentation in open surgery videos: a modified u-net approach with channel amplification. Signal, Image and Video Processing **18**(11), 8061–8076 (2024). https://doi.org/10.1007/s11760-024-03451-3
7. Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Surís Coll-Vinent, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., Afouras, T., Mavroudi, E., Xu, K., Wu, T.H., Zhou, Y., Momeni, L., Hazra, R., Ding, S., Vaze, S., Porcher, F., Li, F., Li, S., Kamath, A., Cheng, H.K., Dollár, P., Ravi, N., Saenko, K., Zhang, P., Feichtenhofer, C.: SAM 3: segment anything with concepts. In: International Conference on Learning Representations (ICLR) (2026)

8. Che, H., Qin, J., Chen, Y., Ji, Z., Yan, Y., Yang, J., Wang, Q., Liang, C., Wu, J.: Improving needle tip tracking and detection in ultrasound-based navigation system using deep learning-enabled approach. IEEE Journal of Biomedical and Health Informatics **28**(5), 2930–2942 (2024). https://doi.org/10.1109/JBHI.2024.3353343
9. Chen, X., Chu, F.J., Gleize, P., Liang, K.J., Sax, A., Tang, H., Wang, W., Guo, M., Hardin, T., Li, X., Lin, A., Liu, J.W., Ma, Z., Sagar, A., Song, B., Wang, X., Yang, J., Zhang, B., Dollár, P., Gkioxari, G., Feiszli, M., Malik, J.: SAM 3D: 3Dfy anything in images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7220–7232 (2026)
10. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., Adam, H.: Searching for MobileNetV3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1314–1324 (2019)
11. Hulou, M.M., Maglinger, B., McLouth, C.J., Reusche, C.M., Fraser, J.F.: Freehand frontal external ventricular drain (EVD) placement: accuracy and complications. Journal of Clinical Neuroscience **97**, 7–11 (2022). https://doi.org/10.1016/j.jocn.2021.12.036
12. Lin, M.A., Siu, A.F., Bae, J.H., Cutkosky, M.R., Daniel, B.L.: HoloNeedle: augmented reality guidance system for needle placement investigating the advantages of three-dimensional needle shape reconstruction. IEEE Robotics and Automation Letters **3**(4), 4156–4162 (2018). https://doi.org/10.1109/LRA.2018.2863381
13. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2117–2125 (2017)
14. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A ConvNet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11976–11986 (2022)
15. Manfield, J.H., Yu, K.K.H.: Real-time ultrasound-guided external ventricular drain placement: technical note. Neurosurgical Focus **43**(5), E5 (2017). https://doi.org/10.3171/2017.7.FOCUS17148
16. O'Leary, S.T., Kole, M.K., Hoover, D.A., Hysell, S.E., Thomas, A., Shaffrey, C.I.: Efficacy of the Ghajar Guide revisited: a prospective study. Journal of Neurosurgery **92**(5), 801–803 (2000). https://doi.org/10.3171/jns.2000.92.5.0801
17. Raabe, C., Fichtner, J., Beck, J., Gralla, J., Raabe, A.: Revisiting the rules for freehand ventriculostomy: a virtual reality analysis. Journal of Neurosurgery **128**(4), 1250–1257 (2018). https://doi.org/10.3171/2016.11.JNS161765
18. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollár, P., Feichtenhofer, C.: SAM 2: segment anything in images and videos. In: International Conference on Learning Representations (ICLR) (2025)
19. Sarrafzadeh, A., Smoll, N., Schaller, K.: Guided (VENTRI-GUIDE) versus freehand ventriculostomy: study protocol for a randomized controlled trial. Trials **15**, 478 (2014). https://doi.org/10.1186/1745-6215-15-478
20. Schneider, M., Kunz, C., Pal'a, A., Wirtz, C.R., Mathis-Ullrich, F., Hlaváč, M.: Augmented reality–assisted ventriculostomy. Neurosurgical Focus **50**(1), E16 (2021). https://doi.org/10.3171/2020.10.FOCUS20779
21. Vasu, P.K.A., Gabriel, J., Zhu, J., Tuzel, O., Ranjan, A.: FastViT: a fast hybrid vision transformer using structural reparameterization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5785–5795 (2023)

22. Yan, W., Ding, Q., Chen, J., Yan, K., Tang, R.S.Y., Cheng, S.S.: Learning-based needle tip tracking in 2D ultrasound by fusing visual tracking and motion prediction. Medical Image Analysis **88**, 102847 (2023). https://doi.org/10.1016/j.media.2023.102847