EasyChair Preprint

№ XXX

Intraoperative Absolute Depth Estimation in MVD Surgery

Jinhee Lee, Hwanhee Lee, Jay Park, Ethan Htun, Bruce Xu, Sang Hoon Cho, Sanghoon Lee and Vivek Buch

June 13, 2025

Intraoperative Absolute Depth Estimation in MVD Surgery

[*]Jinhee Lee[2,3] jinny6876@gmail.com

[*]Hwanhee Lee[1,2,3] hwanhee627@gmail.com

[*]Jay J. Park[3] jayjpark@stanford.edu

Ethan Htun[3] ehtun@stanford.edu

Bruce Changlong Xu[3] bcxu1506@alumni.stanford.edu

Sang Hoon Cho[3] yoopoong95@gmail.com

Sanghoon Lee[1,2] vascular@nate.com

[†]Vivek P. Buch[3] vpbuch@stanford.edu

[1]Department of Neurosurgery, Bongseng Memorial Hospital, Busan, South Korea
[2]AI Research Center, Bongseng Memorial Hospital, Busan, South Korea
[3]Surgical Innovation & Machine Interfacing (SIMI) Laboratory, Department of Neurosurgery, Stanford University, CA, USA

***Abstract*—Microvascular decompression (MVD) is a neurosurgical procedure that relieves nerve compression by repositioning or separating offending blood vessels, effectively reducing pain or spasms. Accurate localization of the compression site is crucial for optimal surgical outcomes, as it enables precise identification and decompression of the offending vessel. While horizontal anatomical relationships are easily identified in the surgical view, compressions occurring along the depth axis are more challenging to discern. In this study, we propose a method to measure accurate intraoperative distances during MVD surgery using Depth-Anything-V2. By leveraging the optical properties of standard imaging equipment in conjunction with the depth estimation model, our method computes precise, absolute distances rather than relying solely on relative measurements, achieving distance estimation errors of less than 2 mm compared to intraoperative and preoperative reference measurements.**

***Index Terms*—Deep Learning, Depth Estimation, Surgical Video Analysis, Microvascular Decompression (MVD)**

I. Introduction

Microvascular decompression (MVD) is a neurosurgical procedure designed to relieve nerve compression by repositioning or separating offending blood vessels. The success of MVD critically depends on the precise localization of the compression site, as accurate identification enables targeted decompression and improves patient symptoms [1]. The surgery is typically performed based on preoperative planning using MRI or CT imaging and executed under direct visualization through an operating microscope or endoscope [2]. However, the complexity of anatomical structures and inherent limitations of intraoperative visual information pose significant challenges to precise assessment and decision-making [3]. Consequently, there has been an increasing demand for computer-assisted techniques to provide complementary insights and enhance surgical precision.

Several studies have explored computer vision and AI-based approaches to assist various aspects of MVD procedures. Preoperative efforts have focused on vessel and nerve segmentation using MRI or CT data to improve surgical planning. Tu et al. [4] proposed a multimodal MRI segmentation network for extracting key structures in MVD using knowledge-driven mutual distillation and topological constraints. Intraoperatively, scene analysis techniques have been applied to microscope or endoscope views to segment anatomical structures, detect landmarks, and assist navigation in real time [5], [6]. However, preoperative segmentation lacks adaptability to intraoperative anatomical changes, and most intraoperative scene analysis methods are limited to two-dimensional representations [7], providing limited positional information without depth, which restricts accurate anatomical interpretation and intraoperative decision-making. While approaches such as stereo endoscopy and depth cameras can provide three-dimensional information, their reliance on specialized hardware and complex setups limits practical adoption in typical surgical settings.

In recent years, monocular depth estimation models based on deep learning [8], [9], [10] have achieved remarkable performance in general computer vision tasks. These models predict dense depth maps from single images, providing relative depth information without requiring specialized hardware. Several studies have since explored their application in surgical scenes, demonstrating potential to enhance intraoperative navigation and visualization. Lou et al. [11] showed that fine-tuning a general-purpose depth model on surgical images improves estimation accuracy. However, monocular models still lack metric scale, generating depth maps that are accurate only up to an unknown scaling factor. This limitation becomes critical in MVD, where precise absolute distance measurements are required to evaluate the spatial relationships between anatomical structures.

In this work, we propose a method to estimate objective, real-world distances during MVD surgery by combining a monocular depth estimation model with the optical properties of surgical microscopes. By leveraging the microscope's focal adjustment, we extract relative depth differences and calibrate them with the predicted depth map to obtain absolute

[*]These authors contributed equally. [†] Corresponding authors.

measurements. The approach is validated using preoperative MRI-based 3D reconstructions, demonstrating its accuracy and potential applicability in intraoperative environments.

II. Related work

A. MVD surgery

Microvascular decompression (MVD) is a neurosurgical procedure used to treat neurovascular compression syndromes such as trigeminal neuralgia (TN) and hemifacial spasm (HFS). TN is primarily caused by compression of the trigeminal nerve (CN5) by the superior cerebellar artery (SCA) or the anterior inferior cerebellar artery (AICA). HFS results from compression of the facial nerve (CN7), typically by the AICA or posterior inferior cerebellar artery (PICA) [1], [12]. The retrosigmoid approach in MVD provides access to the cerebellopontine angle and involves navigating around various neurovascular structures within the posterior fossa, including cranial nerves, arteries such as the SCA and AICA, and veins such as the superior petrosal vein (SPV) [13], [14]. Fig. 1 illustrates the spatial relationships among these key anatomical features relevant to MVD procedures.

Unlike traditional open neurosurgical procedures that require extensive exposure, MVD is a highly precise and minimally invasive surgery that involves opening only a small, well-defined area of the skull to access the affected cranial nerve [15]. Given the compact and confined surgical field, MVD relies on advanced visualization tools such as high-resolution operating microscopes and endoscopes. These tools provide magnified and panoramic views, allowing surgeons to accurately identify neurovascular structures and delicately relieve the compression by repositioning the vessel or inserting Teflon. The limited operative space demands exceptional precision, making MVD one of the most refined and technically challenging procedures in neurosurgery [16], [17].

Although modern operating microscopes and endoscopic systems offer high-resolution visualization, they typically provide two-dimensional images that limit depth perception during MVD procedures [18]. To mitigate the limitations of 2D visualization, technologies such as 3D exoscopes and VR-based surgical planning have been introduced. However, their clinical adoption is limited by cost, complexity, and lack of standardization [2]. In MVD, where the operative field is compact and anatomically dense, accurate real-time depth information is essential for safe vessel manipulation and precise Teflon placement.

B. Depth Estimation in Surgical Imaging

Depth information in medical imaging is crucial for enhancing surgical safety and accuracy, facilitating tasks such as precise lesion localization and improved navigation [19]. Preoperative imaging modalities such as CT and MRI are widely used to obtain volumetric anatomical data, which can be transformed into high-resolution 3D reconstructions using tools like 3D Slicer [20]. However, these scans, acquired preoperatively, cannot reflect real-time anatomical changes, limiting their intraoperative utility [21].

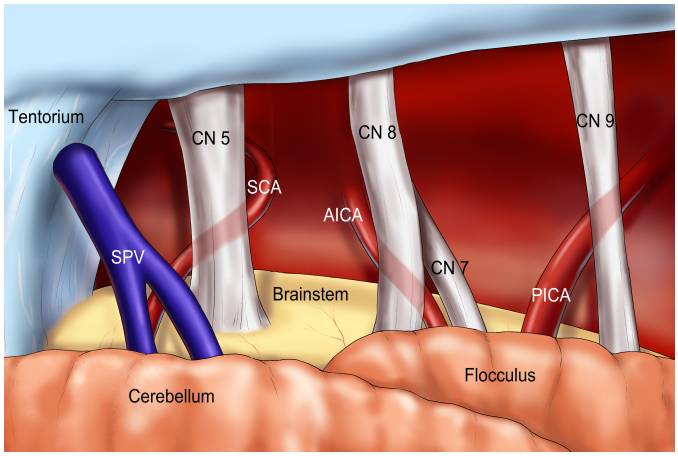

Fig. 1. Simplified schematic of neurovascular anatomy in the cerebellopontine angle from a retrosigmoid approach, showing the relationships among the SPV, cranial nerves (CN5, CN7, CN8, and CN9), and surrounding arteries(SCA, AICA and PICA), as typically visualized during microvascular decompression.

Several imaging techniques have been proposed for real-time depth estimation and explored in surgical applications. Stereo vision reconstructs depth by analyzing images from multiple viewpoints and has been applied in robotic surgery and select stereo-endoscopic systems. RGB-D imaging has shown potential for enhancing spatial awareness in experimental surgical settings. While these techniques offer the potential to improve intraoperative depth perception, their adoption remains limited due to spatial constraints, high costs, and challenges in integrating them into existing surgical workflows.

Recent advances in deep learning have significantly improved monocular depth estimation, predicting depth maps from a single camera feed using convolutional neural networks or transformer-based architectures [22], [23]. Leveraging large datasets, these models achieve state-of-the-art depth estimation without stereo setups or specialized sensors, demonstrating strong performance across various applications. In medical imaging, particularly endoscopic and microscopic images, monocular depth estimation has been actively explored to enhance spatial perception and surgical precision. Recent studies have proposed supervised and self-supervised approaches for medical data, utilizing anatomical priors, geometry constraints, and uncertainty modeling to improve robustness in complex surgical environments [24], [25].

III. Method

The overall pipeline of our proposed method is illustrated in Fig. 2. Intraoperative depth estimation is first performed using a monocular model to obtain relative depth information. Absolute depth is then calculated by calibrating these relative predictions using optical measurements of known landmark distances acquired intraoperatively. Finally, the computed absolute distances are validated against preoperative measurements derived from MRI-based 3D reconstructions using 3D Slicer.

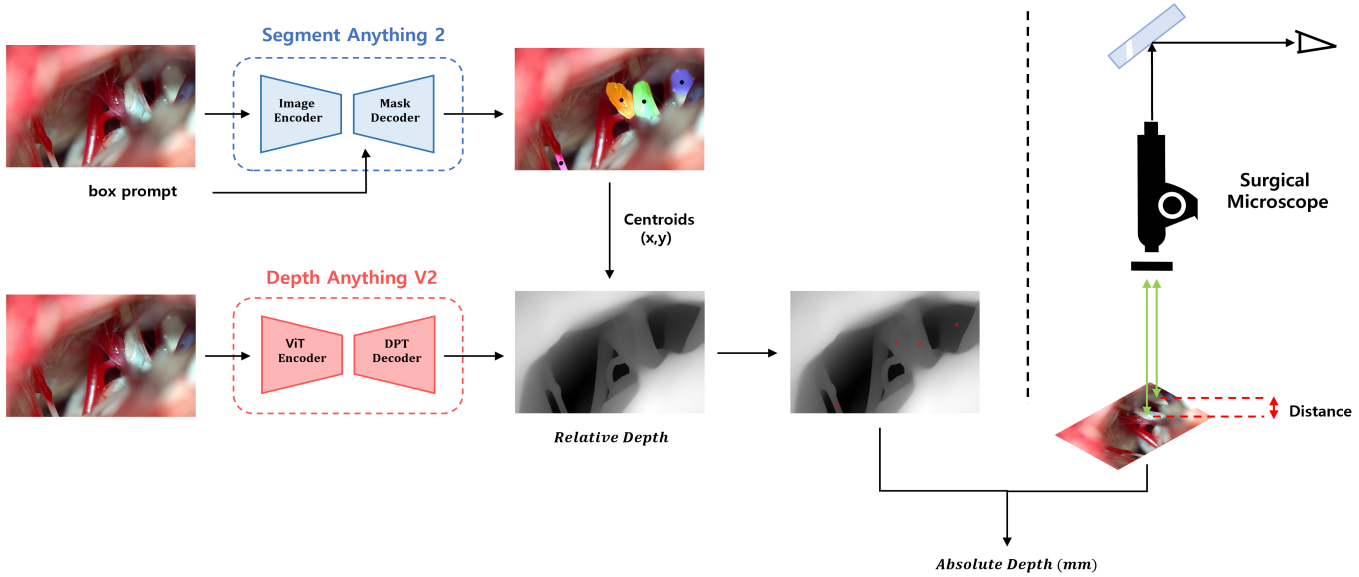

Fig. 2. Overview of the proposed pipeline. Relative depth maps are generated using Depth-Anything-V2 [10], and anatomical structures are segmented with Segment-Anything-Model 2 [26] for centroid-based depth measurement (*left*). Absolute distances are computed by calibrating depth differences using intraoperative microscope focus-based distance measurements between known landmarks (*right*).

A. Relative Depth Estimation

We used a pre-trained Depth-Anything-V2 [10], a state-of-the-art monocular depth estimation model based on a Vision Transformer (ViT) architecture [22], to estimate relative depth within the surgical field. The model, trained on both real and virtual datasets using knowledge distillation from a teacher model, is designed to generalize effectively across a wide range of complex environments, including surgical scenes. To enhance prediction accuracy around critical anatomical structures, we crop the region of interest (ROI) around targeted neurovascular areas. This step reduces interference from irrelevant background elements and focuses the model's prediction capability on clinically relevant region.

B. Absolute Depth Calibration via Optical Scaling

To obtain absolute depth measurements without introducing additional equipment or disrupting the surgical workflow, we utilize a surgical microscope, which is routinely integrated into the neurosurgical setup. This microscope provides intraoperative focal length readouts and allows precise manual focus adjustments at increments of 1 mm. By recording the difference in working distances displayed by the microscope when sequentially focusing on known anatomical landmarks, we derive a reliable physical reference distance for calibration.

Specifically, we first measure the actual physical distance between two clearly identifiable anatomical landmarks via microscope focus adjustment. This measured distance serves as a reference scale factor, enabling us to establish a proportional relationship between pixel-based relative depth values from the Depth-Anything-V2 model and their corresponding real-world distances. We then apply a segmentation model to automatically delineate the anatomical structures of interest within the surgical scene. For each segmented structure, we select the centroid of the segmented mask as the representative point for absolute depth calculation. By applying the computed scale factor, relative depth predictions are converted into absolute distances across the entire surgical field. This procedure is illustrated in detail in Fig. 2.

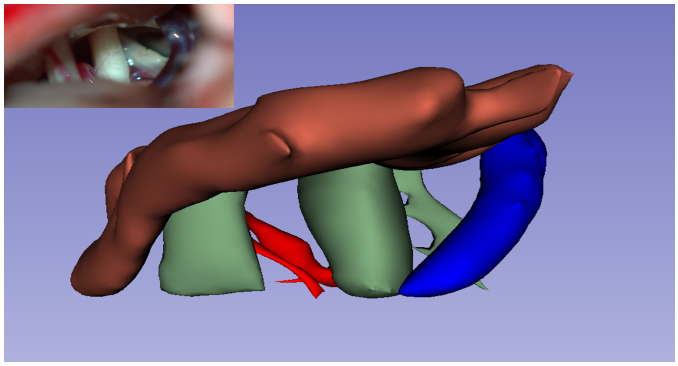

Fig. 3. Synchronized views for depth validation. The top-left inset shows the intraoperative microscopic view, while the main panel displays the corresponding 3D reconstruction using 3D Slicer. Distance between target structures was measured in both views after manual alignment.

C. Validation of Absolute Depth Using 3D Slicer

To validate the absolute depth measurements obtained during surgery, we utilized 3D Slicer, a widely adopted open-source platform for medical image analysis and 3D reconstruction [20]. 3D Slicer generates volumetric anatomical models based on preoperative MRI scans, providing accurate reference

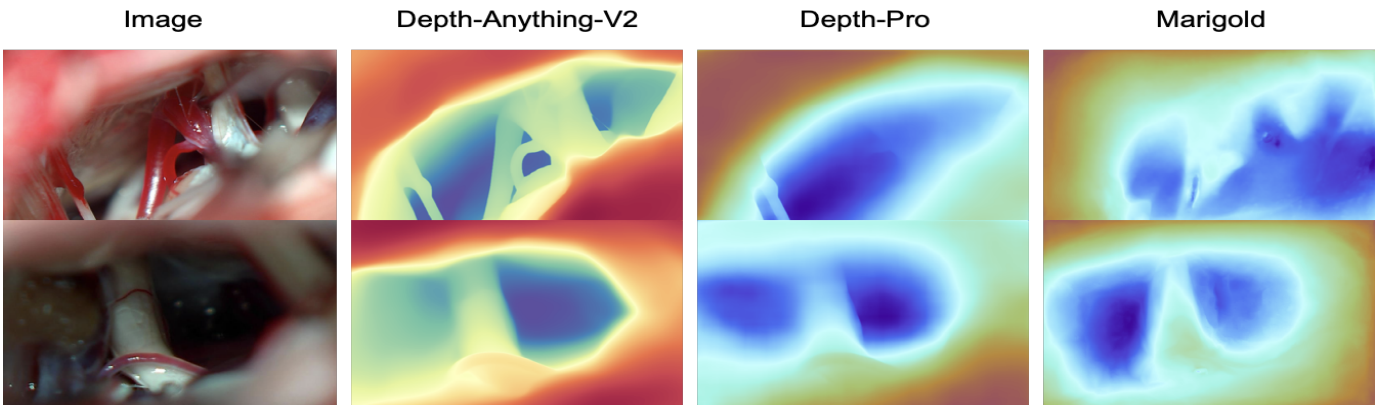

Fig. 4. Qualitative comparison of depth estimation models in representative MVD surgical scenes. Depth-Anything-V2 [10] outperforms DepthPro [9] and Marigold [8] in producing stable and anatomically consistent depth maps.

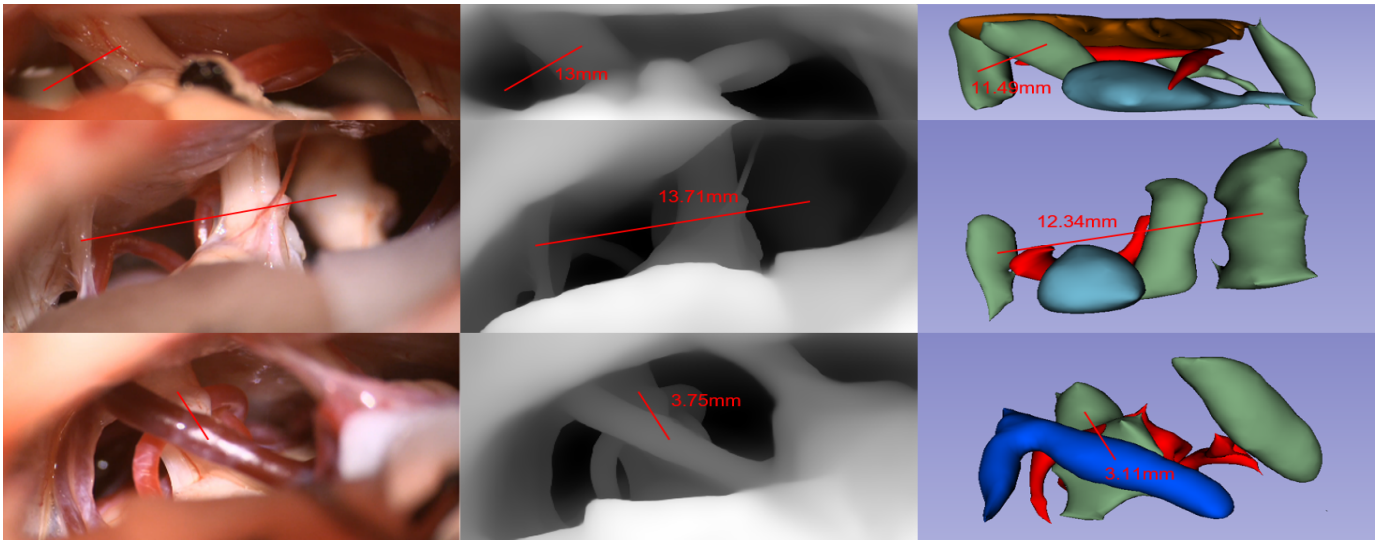

Fig. 5. Validation of absolute distances using 3D Slicer. Each row shows the surgical view, the Depth-Anything-V2 depth map with annotations, and the corresponding 3D Slicer measurement. The annotated values represent distances in the depth direction.

distances for evaluating intraoperative depth measurements. From these volumetric models, we extracted inter-structural distances that served as reference values to assess the accuracy of our absolute depth estimates [27].

To facilitate a direct comparison, we manually aligned the intraoperative surgical scene captured by the microscope with the corresponding 3D reconstructed view from 3D Slicer. This manual alignment ensured consistency in perspective between the intraoperative images and preoperative models. After alignment, distances between corresponding target anatomical structures were measured and directly compared between the two modalities. Fig. 3 illustrates examples of synchronized views and the procedure for measuring distances.

IV. Results

A. Experimental Setup

Our experiments were conducted using intraoperative video recordings from 11 patients who underwent MVD surgery for TN or HFS. The operation was performed by a single neurosurgeon using the same surgical method for all cases. All frames were recorded using a ZEISS TIVATO 700 surgical microscope [28] at a resolution of 1920 × 1080 pixels, and manually cropped to focus on regions of anatomical interest. For anatomical segmentation, we utilized the pre-trained Segment-Anything-Model 2 (SAM2) [26] to identify target structures accurately.

Since no monocular depth estimation models specifically designed for MVD currently exist, we qualitatively evaluated several general-purpose models, including Depth-Anything-V2 [10], Depth-Pro [9], and Marigold [8]. As shown in Fig. 4, Depth-Anything-V2 qualitatively outperformed other candidates, producing clearer and more stable depth maps in representative MVD surgical scenes, particularly in regions involving CN5, CN8, and adjacent vessels. Therefore, we selected Depth-Anything-V2 as the primary model for subsequent experiments. For these experiments, we utilized the

TABLE I

REPRESENTATIVE ABSOLUTE DISTANCE MEASUREMENTS BETWEEN ANATOMICAL STRUCTURES. ESTIMATED DISTANCES FROM THE DEPTH MODEL ARE COMPARED WITH CORRESPONDING INTRAOPERATIVE MICROSCOPE MEASUREMENTS AND PREOPERATIVE 3D SLICER REFERENCES.

Subject	Reference [CN5-CN8] (mm)		Target Structures	**Measured (mm)**		**Estimated (mm)**	**Errors (mm)**	
	Microscope	3D-Slicer		Microscope	3D-Slicer		Microscope	3D-Slicer
P1	13	11.49	CN5–CN9	14.0	10.74	10.4	3.6	0.34
			CN8–CN9	1.0	0.78	2.6	1.6	1.82
			CN5–CN6	3.0	5.972	2.97	0.03	3.002
			CN8–CN6	16.0	17.37	15.97	0.03	1.4
P2	12	10.4	CN5–CN9	14.0	12.34	13.71	0.29	1.37
			CN8–CN9	1.0	1.954	1.71	0.71	0.244
P3	12	11.27	CN5–CN9	12.0	14.06	10.8	1.2	3.26
			CN8–CN9	0.0	2.786	1.2	1.2	1.586
P4	11	10.46	CN5–CN9	12.0	10.72	19.09	7.09	8.37
			CN8–CN9	1.0	0.231	8.08	7.08	7.849
P5	10	9.411	CN5–CN9	10.0	7.829	8.39	1.61	0.561
			CN8–CN9	0.0	1.595	1.61	1.61	0.015
P6	9	9.827	CN5–CN9	9.0	9.492	7.4	1.6	2.92
			CN8–CN9	0.0	0.17	1.6	1.6	1.43
P7	11	9.223	CN5–CN9	8.0	5.505	11.25	3.25	5.745
			CN8–CN9	3.0	3.699	0.25	2.75	3.449
P8	7	10.91	CN8–artery	-	5.222	5.25	-	0.028
			CN5–vein	-	7.781	10.75	-	2.969
			CN8–vein	-	3.11	3.75	-	0.64
P9	11	10.58	CN5–CN9	12.0	13.03	3.92	8.08	9.11
			CN8–CN9	1.0	2.401	7.08	6.08	4.679
P10	16	14.00	CN5–CN9	13.0	13.23	5.73	7.27	7.5
			CN8–CN9	3.0	0.663	10.26	7.26	9.597
P11	8	6.154	CN5–artery	-	6.154	4.19	-	1.964
			CN8–artery	-	0.507	3.81	-	3.303
Average Error	-	-	-	-	-	-	3.197	3.326
Average Error w/o extreme cases	-	-	-	-	-	-	**1.504**	**1.897**

pre-trained Depth-Anything-V2 model configured with a ViT-L encoder and an input resolution of 518×518. Both the Depth-Anything-V2 and SAM2 were used as off-the-shelf pre-trained models without additional fine-tuning on surgical datasets.

B. Absolute Depth Measurement

To demonstrate the effectiveness of our method, we compared the estimated absolute distances between various anatomical structures derived from the depth maps with validation measurements independently obtained from 3D Slicer reconstructions. Figure 5 shows representative examples, including the original surgical images, corresponding predicted depth maps, and aligned 3D Slicer views annotated with identical inter-point measurements. The intraoperative images and 3D Slicer reconstructions were visually aligned to provide consistent anatomical perspectives. Estimated absolute distances from the depth maps and reference distances from 3D Slicer are overlaid in millimeters for direct comparison.

Table I summarizes the quantitative comparisons for multiple anatomical structure pairs, including CN5, CN6, CN8, CN9, and adjacent vessels. The estimated absolute depths closely match the reference measurements derived from microscope focal adjustments and 3D Slicer reconstructions, achieving an overall accuracy within approximately 3 mm for most evaluated anatomical structure pairs. When excluding extreme cases, the error is further reduced to approximately 1.5–1.8 mm, indicating a notably lower margin of error.

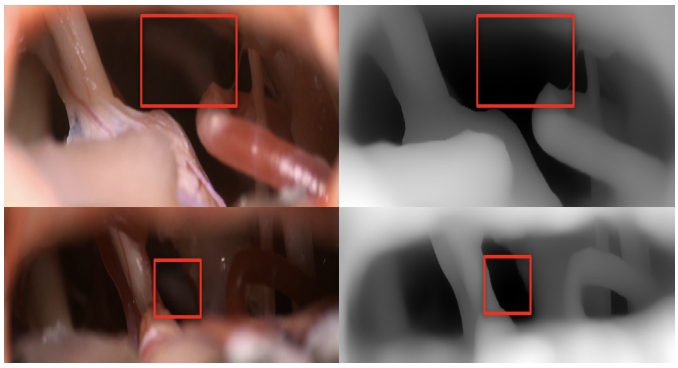

Fig. 6. Depth information is noticeably reduced for deep structures, indicating the model's limited ability to preserve distant depth variations.

V. DISCUSSION

This study demonstrates that absolute depth can be estimated in real time by calibrating depth estimation models with actual measurements. This enables depth perception from 2D surgical views, potentially improving the identification of offending vessels in MVD surgery by clarifying complex anatomical structures. However, we find that our method does not perform effectively in scenes with large depth disparities. As shown in Figure 6, while Depth-Anything-V2 is effective at capturing relative depth, it tends to lose resolution in regions that are significantly farther from the nearest structures. In such cases, depth values for distant anatomical features are compressed

toward the lower end of the value range or inaccurately predicted, despite their physical presence in the scene. When these underestimated values are used for absolute scaling, the resulting distances can become unrealistically large. In cases with extreme depth disparities, the error can reach up to 8mm.

Additionally, we observed a baseline discrepancy of approximately 2mm between intraoperative microscope measurements and preoperative 3D Slicer reconstructions, which may contribute to residual error. In our future efforts, using phantom models with knowledge of the objective distance may help improve the accuracy of calibration, and we plan to utilize advanced surgical microscopes equipped with integrated depth sensing modules. As more MVD procedures are conducted with such equipment, it will become feasible to collect accurate intraoperative depth measurements for robust validation. Based on this data, we aim to construct a dedicated dataset for MVD and develop a depth estimation model capable of predicting absolute depth in real time, ultimately enabling 3D reconstruction to enhance spatial awareness during surgery.

VI. Conclusion

In this work, we present a method to estimate absolute depth during MVD surgery by combining monocular depth prediction with intraoperative optical calibration. The approach enables real-time distance measurement between anatomical structures and shows accuracy within 2mm compared to 3D Slicer references. These results demonstrate that integrating these methodologies effectively bridges the gap between preoperative static imaging and real-time intraoperative analysis, enhancing spatial awareness during surgery.

References

[1] P. J. Jannetta, "Observations on the etiology of trigeminal neuralgia, hemifacial spasm, acoustic nerve dysfunction and glossopharyngeal neuralgia. definitive microsurgical treatment and results in 117 patients," *Neurochirurgia*, vol. 20, no. 05, pp. 145–154, 1977.

[2] I. I. Okon, S. S. Menon, M. Osama, M. Aiman, L. F. F. Paleare, D. L.-P. Eliseo III, M. D. Shafqat, C. O. Ezeaku, M. Y. Ferreira, Y. Razouqi *et al.*, "Microvascular decompression: a contemporary update," *BMC surgery*, vol. 25, no. 1, p. 20, 2025.

[3] S. Lee, K. M. Joo, and K. Park, "Challenging microvascular decompression surgery for hemifacial spasm," *World Neurosurgery*, vol. 151, pp. e94–e99, 2021.

[4] R. Tu, D. Zhang, C. Li, L. Xiao, Y. Zhang, X. Cai, and W. Si, "Multimodal mri segmentation of key structures for microvascular decompression via knowledge-driven mutual distillation and topological constraints," *International Journal of Computer Assisted Radiology and Surgery*, vol. 19, no. 7, pp. 1329–1338, 2024.

[5] R. Bai, X. Liu, S. Jiang, and H. Sun, "Deep learning based real-time semantic segmentation of cerebral vessels and cranial nerves in microvascular decompression scenes," *Cells*, vol. 11, no. 11, p. 1830, 2022.

[6] J. J. Park, N. Doiphode, X. Zhang, L. Pan, R. Blue, J. Shi, and V. P. Buch, "Developing the surgeon-machine interface: using a novel instance-segmentation framework for intraoperative landmark labelling," *Frontiers in Surgery*, vol. 10, p. 1259756, 2023.

[7] J. Watrinet, L. Wenzel, J. Fürmetz, P. Augat, P. Blum, C. Neidlein, M. Bormann, F. Stuby, and C. von Rueden, "Possibilities and limits of intraoperative 2d imaging in trauma surgery," *Unfallchirurgie (Heidelberg, Germany)*, 2023.

[8] B. Ke, A. Obukhov, S. Huang, N. Metzger, R. C. Daudt, and K. Schindler, "Repurposing diffusion-based image generators for monocular depth estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 9492–9502.

[9] A. Bochkovskii, A. Delaunoy, H. Germain, M. Santos, Y. Zhou, S. R. Richter, and V. Koltun, "Depth pro: Sharp monocular metric depth in less than a second," *arXiv preprint arXiv:2410.02073*, 2024.

[10] L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, and H. Zhao, "Depth anything v2," *Advances in Neural Information Processing Systems*, vol. 37, pp. 21 875–21 911, 2024.

[11] A. Lou, Y. Li, Y. Zhang, and J. Noble, "Surgical depth anything: Depth estimation for surgical scenes using foundation models," *arXiv preprint arXiv:2410.07434*, 2024.

[12] M. R. McLaughlin, P. J. Jannetta, B. L. Clyde, B. R. Subach, C. H. Comey, and D. K. Resnick, "Microvascular decompression of cranial nerves: lessons learned after 4400 operations," *Journal of neurosurgery*, vol. 90, no. 1, pp. 1–8, 1999.

[13] P. J. Jannetta, "Microsurgery of cranial nerve cross-compression," *Neurosurgery*, vol. 26, pp. 607–615, 1979.

[14] A. L. Rhoton Jr, "The cerebellopontine angle and posterior fossa cranial nerves by the retrosigmoid approach," *Neurosurgery*, vol. 47, no. 3, pp. S93–S129, 2000.

[15] J. Huang, Y. Zhan, Y. Li, L. Jiang, K. Wang, Z. Wu, Y. Xie, and Q. Shi, "The efficacy and safety of¡ 2 cm micro-keyhole microvascular decompression for hemifacial spasm," *Frontiers in Surgery*, vol. 8, p. 685155, 2021.

[16] P. Jannetta, "Outcome after microvascular decompression for typical trigeminal neuralgia, hemifacial spasm, tinnitus, disabling positional vertigo, and glossopharyngeal neuralgia (honored guest lecture)," *Clinical neurosurgery*, vol. 44, pp. 331–383, 1997.

[17] A. S. S. Andersen, T. B. Heinskou, P. Rochat, J. B. Springborg, N. Noory, E. A. Smilkov, L. Bendtsen, and S. Maarbjerg, "Microvascular decompression in trigeminal neuralgia-a prospective study of 115 patients," *The journal of headache and pain*, vol. 23, no. 1, p. 145, 2022.

[18] B. Feng, X. Chai, Y. Yu, H. Xia, J. Wu, Y. Ren, P. Yu, and Y. Zhu, "Endoscopic microvascular decompression for primary trigeminal neuralgia: surgical experience and early outcomes," *Scientific Reports*, vol. 15, no. 1, p. 10289, 2025.

[19] F. Y. L. Sop, M. D'Ercole, A. Izzo, A. Rapisarda, E. Ioannoni, A. Caricato, A. Olivi, and N. Montano, "The impact of neuronavigation on the surgical outcome of microvascular decompression for trigeminal neuralgia," *World neurosurgery*, vol. 149, pp. 80–85, 2021.

[20] L. Zhou, P. Lei, P. Song, Z. Li, H. Zhang, H. Wei, L. Gao, Q. Hua, H. Ye, Q. Chen *et al.*, "Clinical application of 3d slicer reconstruction and 3d printing localization combined with neuroendoscopy technology in vps surgery," *Scientific Reports*, vol. 15, no. 1, p. 2609, 2025.

[21] I. J. Gerard, M. Kersten-Oertel, K. Petrecca, D. Sirhan, J. A. Hall, and D. L. Collins, "Brain shift in neuronavigation of brain tumors: A review," *Medical image analysis*, vol. 35, pp. 403–420, 2017.

[22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[23] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby *et al.*, "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[24] Y. Liu and S. Zuo, "Self-supervised monocular depth estimation for gastrointestinal endoscopy," *Computer Methods and Programs in Biomedicine*, vol. 238, p. 107619, 2023.

[25] M. S. Zeinoddin, C. Lena, J. Qu, L. Carlini, M. Magro, S. Kim, E. De Momi, S. Bano, M. Grech-Sollars, E. Mazomenos *et al.*, "Dares: Depth anything in robotic endoscopic surgery with self-supervised vector-lora of the foundation model," *arXiv preprint arXiv:2408.17433*, 2024.

[26] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson *et al.*, "Sam 2: Segment anything in images and videos," *arXiv preprint arXiv:2408.00714*, 2024.

[27] J. Egger, T. Kapur, A. Fedorov, S. Pieper, J. V. Miller, H. Veeraraghavan, B. Freisleben, A. J. Golby, C. Nimsky, and R. Kikinis, "Gbm volumetry using the 3d slicer medical image computing platform," *Scientific reports*, vol. 3, no. 1, p. 1364, 2013.

[28] Carl Zeiss Meditec AG, *ZEISS TIVATO 700 Instructions for Use, Version 1.0*, Carl Zeiss Meditec AG, Jena, Germany, 2020. [Online]. Available: https://www.bioclinicalservices.com.au/carl-zeiss/clinical/tivato-700-instructions-for-use-ver-1-0-sept-2020