

Development of a Computer Vision System for Stone Segmentation During CVAC 2.0 In-Scope Aspiration Ureteroscopy

Kavita Gupta

kag4024@med.cornell.edu

New York Presbyterian Weill Cornell Medicine

Bahram Jafrasteh

New York Presbyterian Weill Cornell Medicine

Michael Tzeng

New York Presbyterian Weill Cornell Medicine

Alexandra Farhangui

New York Presbyterian Weill Cornell Medicine

Patrick Samson

New York Presbyterian Brooklyn Methodist Hospital

Joseph Del Pizzo

New York Presbyterian Weill Cornell Medicine

Michael Amirian

New York Presbyterian Queens Hospital

Qingyu Zhao

New York Presbyterian Weill Cornell Medicine

Research Article

Keywords:

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11254629/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. I am Editorial Board member

Abstract

Objective

To develop and validate StoneClear, an artificial intelligence computer vision system that automatically detects and delineates residual stone fragments (RSFs) during Steerable Ureteroscopic Renal Evacuation (SURE) with the CVAC® 2.0 Aspiration System.

Methods

Intraoperative videos from 10 patients undergoing flexible ureteroscopy with the CVAC® 2.0 device were prospectively collected. Board-certified urologists densely annotated 1,439 stone-containing frames across diverse appearances and selected 977 stone-free frames containing background anatomy, instrumentation, or artifacts (such as urothelium, blood, bubbles, or Randall's plaques). StoneClear was trained to first determine whether a stone was present in each frame (frame-level detection) and then automatically segment the stone region at the pixel level (pixel-level boundaries). Model performance was evaluated using patient-level 5-fold cross-validation.

Results

Mean preoperative stone volume on CT was 10,925.36 mm³ (median: 1,641.49 mm³) and the mean postoperative residual stone volume was 170.20 mm³ (median: 0.0 mm³). We then evaluated StoneClear's ability to identify and localize residual stones in endoscopic video. The model correctly identified the presence or absence of stones in 92.8% of video frames. For frames containing visible stones, the model demonstrated precise pixel-level localization with a mean Dice Similarity Coefficient (DSC) of 0.86, a standard metric measuring spatial overlap between AI predictions and expert surgeon outlines. On frames without visible stones, StoneClear incorrectly identified stones in 13.6% of those frames (false-positive rate), substantially outperforming several existing AI models (such as standard U-Net architectures, which yielded a 78.2% false-positive rate).

Conclusions

StoneClear provides a robust proof-of-concept for real-time AI-assisted intraoperative visual stone assessment within the highly dynamic environment of suction-assisted ureteroscopy. By providing objective visual tracking and accurate boundary localization during the case, this framework establishes a critical foundation for intraoperative residual stone volume estimation to assist surgeons in verifying intraoperative endoscopic clearance.

Introduction

Kidney stone disease is a major and growing public health burden, affecting nearly one in eleven people in the United States.¹ Beyond the acute pain of an initial episode, the disease is defined by its propensity for recurrence: up to half of patients experience another symptomatic stone event within five years.¹ The primary goal of surgical management, most commonly flexible ureteroscopy (fURS) with laser lithotripsy, is to achieve a stone-free status. However, residual stone fragments (RSFs) remain a critical failure point in contemporary endourology. Even small, retained fragments can serve as a nidus for rapid stone regrowth, cause recurrent obstruction, predispose patients to urinary tract infections, and often necessitate repeat surgical intervention.²

Advances in laser technology and endoscope design have improved intraoperative efficiency, yet confirming absolute stone-free status at case conclusion remains elusive. Suction-assisted ureteroscopic systems were developed specifically to address this gap. The CVAC® 2.0 Aspiration System (Calyxo, Inc., Pleasanton, CA) is one such commercial platform designed for in-scope suction lithotripsy. This single-use, 11.9-French digital ureteroscope integrates laser lithotripsy, direct visualization, and simultaneous stone evacuation through independent irrigation and aspiration channels. Within this framework, four angled microjet nozzles fluidize stone dust and debris, while a large-bore outflow lumen evacuates fragments up to approximately 2 mm in diameter. A single trigger controls this entire mechanism without interrupting visualization. Inflow and outflow mechanics are designed to regulate fluid movement and manage intrarenal pressure, though clinical impact on complication rates remains actively under investigation. Despite these advances, reported absolute stone-free rates on postoperative imaging remain low, around 20% for some in-scope aspiration systems and 14% for Flexible and Navigable Suction (FANS), and a large multicenter trial has reported a high-grade stone-free rate of 57.4%.³

A key contributing factor to retained fragments is the inherent limitation of subjective human visual inspection under endoscopic lighting. During laser lithotripsy, the operative field is frequently obscured by stone dust, blood, and irrigation fluid—the familiar “snowstorm” effect—making it difficult or impossible to visually confirm complete fragment clearance. Suction-assisted systems like the CVAC 2.0 alter the optical environment by evacuating particulate matter. However, this setting introduces unique visual conditions: fragments mobilized by the microjets traverse the field at high velocity, transiently appearing and disappearing as flow conditions shift. Furthermore, the in-field presence of the suction catheter partially occludes portions of the surgical view. Consequently, surgeons lack an objective, real-time tool to quantify residual stone burden or confirm visual clearance at the conclusion of the procedure.

Computer vision (CV) and artificial intelligence (AI) have emerged as promising approaches to make intraoperative stone assessment more objective. Early efforts focused on preoperative imaging, where deep learning architectures demonstrated accurate stone segmentation on CT and KUB images.^{4–6} More recently, deep learning has been applied directly to fURS video.^{7–10} Despite this progress, virtually all existing intraoperative CV work has been conducted in non-aspirating fURS. To close this gap, we

developed StoneClear, a deep learning framework for real-time detection and pixel-level segmentation of RSFs during Steerable Ureteroscopic Renal Evacuation (SURE) CVAC ureteroscopy. StoneClear combines a transformer-based deep learning model with a novel gating mechanism designed to suppress false-positive stone predictions in visually challenging frames.

Methods

Following institutional review board approval, intraoperative decision-support evaluation was prospectively performed on videos of SURE procedures using the CVAC® 2.0 Aspiration System. All procedures were performed by fellowship-trained endourologists. Only procedures with adequate visualization and uninterrupted recording were included. Preoperative non-contrast computed tomography (CT) imaging was evaluated to establish baseline volumetric stone burden. Postoperative CT scans were reviewed where available to measure post-procedure residual stone volume. Cases demonstrating complete stone absence on postoperative scan were assigned a residual stone volume of 0.0 mm³.

This study used videos from 10 patients' CVAC procedures. Although 10 patient cases represent a modest cohort size typical of initial exploratory medical AI development, the unit of analysis in computer vision is the individual annotated frame. Across these 10 cases, over 1.6 million total video frames were generated, from which a representative sample of 2,416 high-utility frames was selected for dense pixel-level annotation. Frames were extracted and resized to 512 × 512 pixels to balance computational efficiency with preservation of stone morphology.

For each video, board-certified urologists identified frames across a range of clinically relevant appearances. Frames were sampled strategically across various operative stages (early lithotripsy, active evacuation, and final mucosal inspection) to capture maximum visual heterogeneity rather than uniform temporal intervals. In each selected frame, a binary stone mask was manually delineated using MELAGE, an internally adapted open-source annotation platform.^{11,12} Across all patients, 1,439 stone-containing frames were manually segmented. To evaluate false detections, 977 stone-free frames were selected, capturing background renal anatomy, instrumentation, as well as benign anatomical features such as mucosal erythema, blood clots, and Randall's plaques. Fixed mucosal calcifications like Randall's plaques were explicitly annotated as non-stone background to ensure the algorithm learned to distinguish free-floating or treatable residual stone fragments from fixed subepithelial plaques.

Machine Learning Architecture and Clinical Metrics

StoneClear determines, for each frame, whether it contains a stone and localizes it at the pixel level. The AI system operates in two sequential stages: first acting as a 'gatekeeper' (asking: 'Is a stone present in this picture?'), and second acting as an 'outliner' (asking: 'Where exactly are the boundaries of the stone?'). The model is built on a transformer-based deep learning architecture named PraNet-V2,¹³ which extracts visual features from each endoscopic frame (Fig. 1). Unlike traditional computer vision models

that examine small image patches locally, transformer architectures analyze relationships across the entire image simultaneously, helping the model distinguish true stones from background reflections or scope shadows.

To evaluate the model, a patient-level 5-fold cross-validation procedure was used where 8 procedures were used for training and 2 for testing in each fold. We recorded standard evaluation metrics: Frame Classification Accuracy, False-Positive Rate (FPR), False-Negative Rate (FNR), and the **Dice Similarity Coefficient (DSC)**, a standard statistical measure of spatial overlap used in computer vision, ranging from 0.0 (no overlap) to 1.0 (perfect identical overlap). In clinical terms, a DSC above 0.80 represents excellent agreement between the AI's predicted stone outline and the expert surgeon's manual tracing.

Results

In total, 10 patients (age 65.8 ± 13.1 years; 7 females, Table 1) undergoing fURS were included. Preoperatively, patients presented with a mean baseline stone volume of 10,925.36 mm³ (median: 1,641.49 mm³; range: 487.75 mm³ to 101,481.53 mm³). Among patients with completed postoperative CT abdomen and pelvis imaging, the mean post-operative residual stone volume of 170.20 mm³.

Table 1

Summary of the study dataset, including total video frames, the subset of manually annotated stone-containing frames, and the subset of selected stone-free frames used for model development.

Patient	# of Manually Annotated Stone-Containing Frames	# of Selected Stone-free Frames	Total Video Frames
P1	61	66	53,959
P2	84	105	107,970
P3	375	112	301,300
P4	93	98	230,346
P5	150	106	133,536
P6	98	110	179,525
P7	172	106	134,970
P8	53	89	54,000
P9	98	139	157,247
P10	255	46	269,794
Total	1,439	977	1,622,647

Across 5-fold patient-level cross-validation of the endoscopic frames, StoneClear distinguished stone-containing from stone-free frames with 92.8% accuracy (Table 2). On stone-free frames, StoneClear generated false-positive stone segmentations in 13.6% (133 of 977) of frames. While an ideal false-

positive rate would be 0%, in medical computer vision a baseline false-positive rate is evaluated relative to standard models. Without the classification-gating mechanism, standard segmentation architectures erroneously predicted stones in up to 78.2% of stone-free frames due to optical reflections and mucosal artifacts (Table 2). Thus, the proposed gating mechanism reduced false-positive segmentations by more than fivefold.

Table 2

Performance comparison of kidney stone detection and segmentation with two model variants: replacing the transformer-based visual feature extractor with a U-Net based extractor (first row) and removing the gating mechanism (second row), each compared with the proposed full model (third row).

Model	Stone Detection Accuracy (%)	False Positive Rate (%)	Dice Similarity Coefficient (DSC)
U-Net+Gating	88.21 ± 2.60	24.75 ± 7.76	0.77 ± 0.05
Transformer w.o. Gating	66.98 ± 10.31	78.18 ± 17.30	0.89 ± 0.02
Proposed Model (Transformer + Gating)	92.77 ± 3.79	13.60 ± 7.53	0.86 ± 0.16

For frames with visible stones, StoneClear achieved a low false-negative rate (1.8%) and a mean Stone DSC of 0.86 (Fig. 2). Representative frames, expert ground-truth annotation, and model predictions are shown in Fig. 3. Representative failure cases where StoneClear produced inaccurate segmentations were specifically examined (Fig. 4), occurring primarily in frames heavily obscured by dense dust or large reflective stones.

Discussion

Achieving complete stone clearance remains a central goal in endourology. While suction-assisted ureteroscopy platforms like CVAC 2.0 enhance fragment evacuation, clinical studies demonstrate that residual stone fragments (RSFs) persist postoperatively in a notable subset of cases. The ASPIRE randomized controlled trial, published in December 2024, demonstrated that the SURE procedure with the CVAC system achieved superior stone clearance and significantly reduced residual stone volume (RSV) compared with standard ureteroscopy, with a 73% reduction in two-year healthcare utilization. However, the reported stone-free rate in the ASPIRE trial was only 48%, and real-world multicenter studies, including a large Northwell series, achieved 57% at best.¹⁴ The CLEARANCE study, evaluating the second-generation CVAC 2.0, demonstrated 96% mean stone clearance with a mean RSV of 14.1 mm³ at 30 days, confirming that a meaningful proportion of patients retain clinically relevant residual fragments.¹⁵ A critical distinction must be drawn between intraoperative 'endoscopic visual clearance' (what the surgeon sees through the scope at case conclusion) and formal 'postoperative stone-free status' (typically assessed via non-contrast CT at 30 to 90 days). High reported visual clearance rates often exceed true radiographic stone-free rates because small fragments hidden in unexamined calyces, obscured by minor mucosal bleeding, or buried behind infundibular neck folds remain undetected during endoscopic inspection.

StoneClear was specifically designed as an intraoperative visual decision-support tool—not a replacement for postoperative radiologic imaging. Its primary function is to enhance the surgeon's real-time visual acuity during endoscopic inspection, helping identify small fragments floating in the field or resting on the mucosa before the scope is withdrawn.

The visual environment during suction-assisted ureteroscopy presents distinct challenges: high-velocity microjet fluid flow, rapidly moving stone particles, and partial occlusion by the suction catheter tip. In this setting, subjective surgeon estimation of remaining stone burden is prone to error. In our dataset, informal intraoperative surgeon perception often overestimated true visual clearance prior to systematic re-inspection of all calyces. StoneClear provides an automated, objective overlay that highlights unevacuated fragments during active suction and final endoscopic inspection.

Computer vision applications in surgical specialties outside of urology offer valuable parallels for evaluating AI accuracy. In gastroenterology, real-time computer vision systems for colorectal polyp detection (CADE) and segmentation (CADx) routinely operate at DSC thresholds between 0.78 and 0.85 and false-positive rates of 10–15% while achieving substantial reductions in missed adenomas during screening endoscopy.^{16–20} Similarly, automated surgical phase recognition and instrument tracking in laparoscopic cholecystectomy utilize comparable deep learning architectures. These cross-disciplinary successes suggest that AI overlay systems do not require 100% precision to provide meaningful intraoperative assistance, provided they meaningfully reduce surgeon oversight.^{21,22}

Limitations

This study has limitations. The present cohort of 10 procedures is modest and reflects the early, single-center nature of this proof-of-concept study; nonetheless, dense frame-level annotation yielded 2,416 labeled frames that served as the effective unit of model training and evaluation, and patient-level cross-validation ensured that every procedure contributed to testing. In addition, annotated frames were purposively selected for clear stone visibility, and quantitative evaluation across the full length of each case was precluded by the need for dense, frame-by-frame ground-truth annotation of more than 1.6 million video frames. The annotated set was nonetheless constructed to include challenging, lower-quality frames, such as fragments undergoing high-velocity evacuation and fragments partially occluded by the suction catheter; validation on full-length, unselected case recordings remains an important direction for future work.

Conclusion

StoneClear provides proof-of-concept that transformer-based deep learning with classification-driven gating delivers accurate, robust stone detection and segmentation in the visually demanding CVAC ureteroscopy environment. By providing objective real-time spatial boundary delineation, this technology offers a valuable tool for intraoperative visual verification during suction-assisted lithotripsy.

Declarations

1. Funding: This work was supported by the Mastercard Weill Cornell AI/Big Data/Enabling Technologies Seed Grant Program, awarded to Kavita Gupta.

2. Clinical trial number: Not applicable.

3. Human Ethics and Consent to Participate: This study was approved by the Weill Cornell Medicine Institutional Review Board (IRB #25-08029166) and was conducted in accordance with the Declaration of Helsinki. The IRB waived the requirement for informed consent because the study used de-identified surgical videos from retrograde intrarenal surgery (RIRS).

4. Consent to Publish: Not applicable.

Social media caption

StoneClear, a transformer-based AI with a classification-gated false-positive filter, achieves 92.8% stone detection accuracy and DSC 0.86 during suction-assisted ureteroscopy, the first CV system built for this environment. #Endourology #AI

References

1. Shastri S, Patel J, Sambandam KK, et al. Kidney Stone Pathophysiology, Evaluation and Management: Core Curriculum 2023. *Am J Kidney Dis.* 2023;82:617–34.
2. Wong VKF, Que J, Kong EK, et al. The Fate of Residual Fragments After Percutaneous Nephrolithotomy: Results from the Endourologic Disease Group for Excellence Research Consortium. *J Endourol.* 2023;37:617–22.
3. Gauhar V, Traxer O, Castellani D, et al. Could Use of a Flexible and Navigable Suction Ureteral Access Sheath Be a Potential Game-changer in Retrograde Intrarenal Surgery? Outcomes at 30 Days from a Large, Prospective, Multicenter, Real-world Study by the European Association of Urology Urolithiasis Section. *Eur Urol Focus.* 2024;10:975–82.
4. Elton DC, Turkbey EB, Pickhardt PJ, et al. A deep learning system for automated kidney stone detection and volumetric segmentation on noncontrast CT scans. *Med Phys.* 2022;49:2545–54.
5. Huang Z-H, Liu Y-Y, Wu W-J, et al. Design and Validation of a Deep Learning Model for Renal Stone Detection and Segmentation on Kidney–Ureter–Bladder Images. *Bioengineering.* 2023;10:970.
6. Ye Y, Chipusu K, He L, et al. Urological diagnostics based on kidney stone detection in CT imaging using YOLOv8 deep learning framework. *Front Med.* 2026;13:1702159.
7. Setia SA, Stoeber ZA, Floyd C, et al. Computer Vision Enabled Segmentation of Kidney Stones During Ureteroscopy and Laser Lithotripsy. *J Endourol.* 2023;37:495–501.

8. Lu D, Wu Y, Acar A, et al. ASSIST-U: A system for segmentation and image style transfer for ureteroscopy. *Healthc Tech Lett*. 2024;11:40–7.
9. Cabo J, Lu D, Floyd C et al. The Evaluation of Computer Vision-Based Automated Performance Metrics for Endoscopic Kidney Stone Surgery. *J Endourol* 2025; end.2024.0895.
10. Lu D, Deol ES, Koyama T, et al. A computer vision model for automated kidney stone segmentation and evaluation of its performance vs surgeons. *BJU Int*. 2026;137:87–94.
11. Jafrasteh B, Lubián-López SP, Trimarco E, et al. MGA-Net: A novel mask-guided attention neural network for precision neonatal brain imaging. *NeuroImage*. 2024;300:120872.
12. Jafrasteh B, Lubián-López SP, Benavente-Fernández I. MELAGE: A purely python based Neuroimaging software (Neonatal). *arXiv preprint arXiv:230907175*.
13. Hu B-C, Ji G-P, Shao D et al. PraNet-V2: Dual-Supervised Reverse Attention for Medical Image Segmentation. *Epub ahead of print 15 April 2025*. 10.48550/ARXIV.2504.10986
14. Matlaga BR, Mueller TJ, Johnson B, et al. A Prospective, Randomized, Noninferiority Study to Evaluate the Safety and Effectiveness of Steerable Ureteroscopic Renal Evacuation Compared with Standard Ureteroscopy: 30-Day Results of the ASPIRE Study. *J Endourol*. 2025;39:10–8.
15. Eisner BH, Pathak N, Sabnis R, et al. A Prospective Study of Steerable Ureteroscopic Renal Evacuation Using the Second-Generation CVAC® Aspiration System: Results from the CLEARANCE Study. *J Endourol*. 2026;40:455–61.
16. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, Liu P, Li L, Song Y, Zhang D, Li Y, Xu G, Tu M, Liu X. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 2019;68(10):1813–9.
17. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, Ferrara E, Spadaccini M, Alkandari A, Fugazza A, Anderloni A, Galtieri PA, Pellegatta G, Carrara S, Di Leo M, Craviotto V, Lamona L, Lorenzetti R, Andrealli A, Antonelli G, Wallace M, Sharma P, Rosch T, Hassan C. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology*. 2020;159(2):512–e5207.
18. Hassan C, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, Antonelli G, Yu H, Areia M, Dinis-Ribeiro M, Bhandari P, Sharma P, Rex DK, Rösch T, Wallace M, Repici A. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc*. 2021;93(1):77–e856. 10.1016/j.gie.2020.06.059. . Epub 2020 Jun 26. PMID: 32598963.
19. Hsieh YH, Tang CP, Tseng CW, Lin TL, Leung FW. Computer-Aided Detection False Positives in Colonoscopy. *Diagnostics (Basel)*. 2021;11(6):1113.
20. Tudela Y, Majó M, de la Fuente N, Galdran A, Krenzer A, Puppe F, Yamlahi A, Tran TN, Matuszewski BJ, Fitzgerald K, Bian C, Pan J, Liu S, Fernández-Esparrach G, Histace A, Bernal J. A complete benchmark for polyp detection, segmentation and classification in colonoscopy images. *Front Oncol*. 2024;14:1417862.

21. Shinozuka K, Turuda S, Fujinaga A, Nakanuma H, Kawamura M, Matsunobu Y, Tanaka Y, Kamiyama T, Ebe K, Endo Y, Etoh T, Inomata M, Tokuyasu T. Artificial intelligence software available for medical devices: surgical phase recognition in laparoscopic cholecystectomy. *Surg Endosc*. 2022;36(10):7444–52.
22. Fernicola A, Palomba G, Capuano M, De Palma GD, Aprea G. Artificial intelligence applied to laparoscopic cholecystectomy: what is the next step? A narrative review. *Updates Surg*. 2024;76(5):1655–67.

Figures

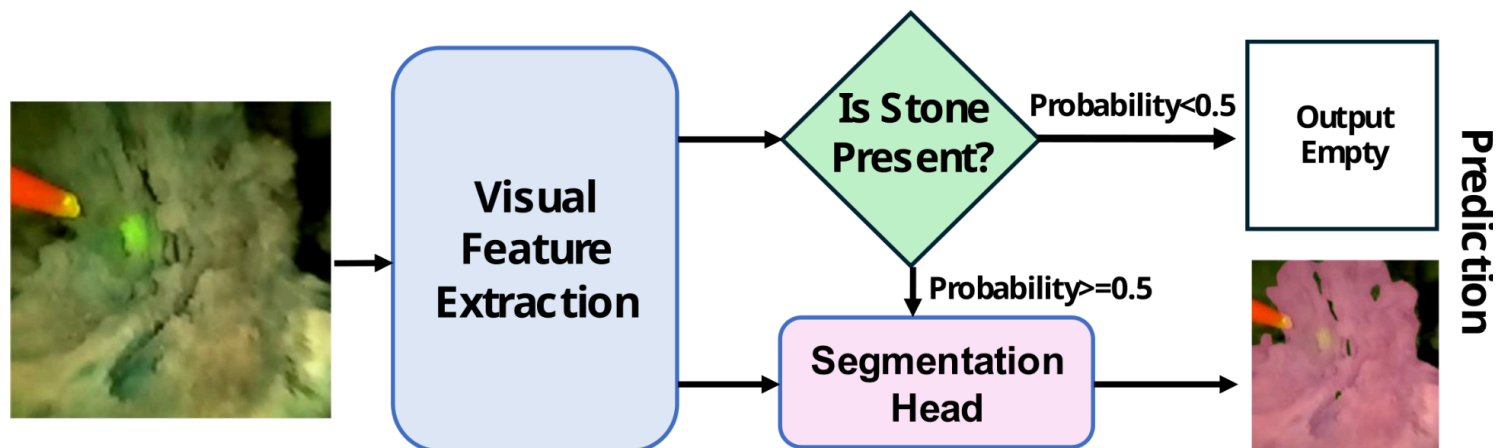


Figure 1

Model architecture: The system utilizes a transformer to extract visual features from an endoscopic frame. Then a gating classifier identifies if the frame contains stones. If a stone is present, a specialized segmentation model generates pixel-wise stone localization.

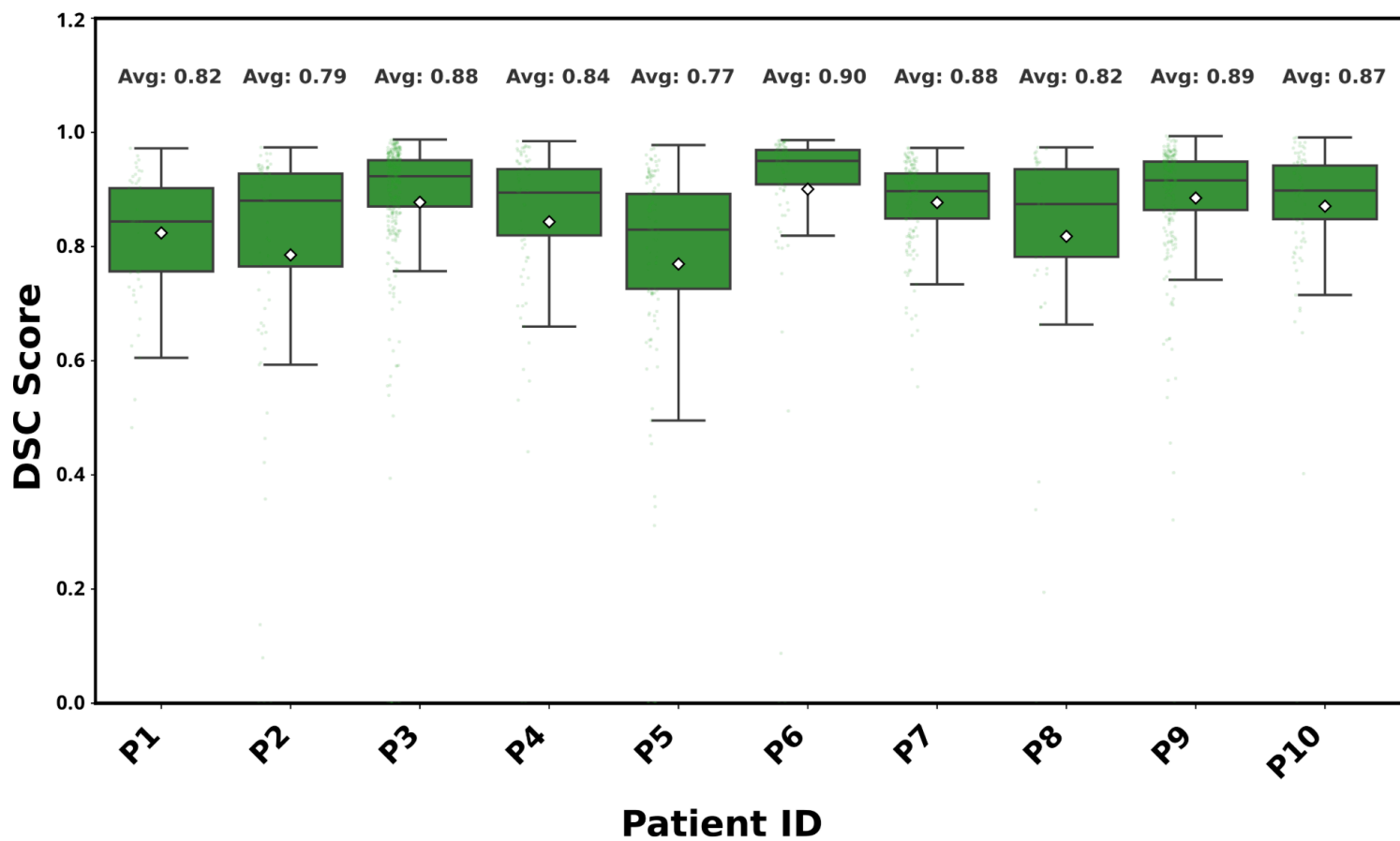


Figure 2

Segmentation accuracy (DSC) for individual patients (P1–P10) based on 5-fold cross-validation.

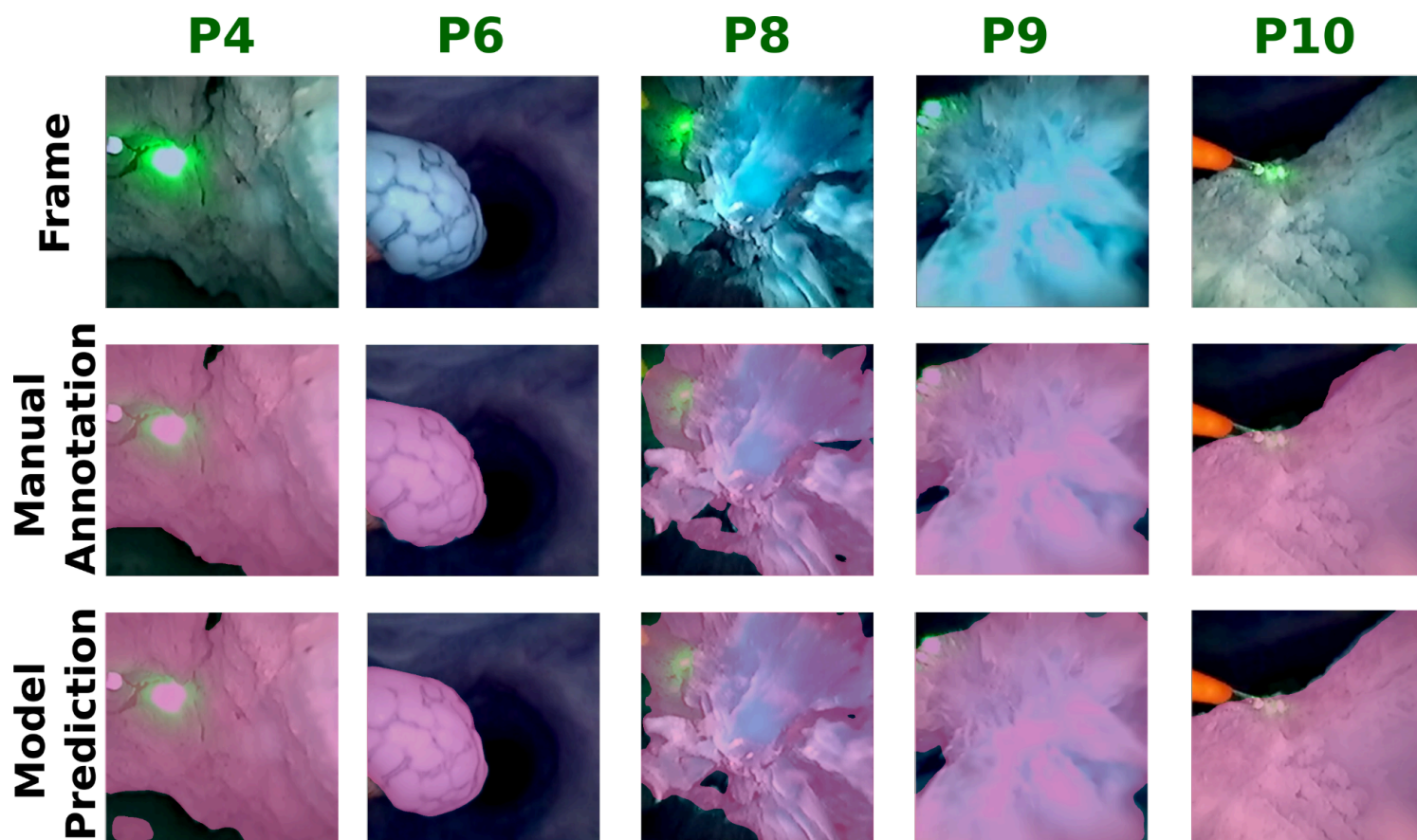


Figure 3

Exemplar segmentation outcomes: The rows display the original endoscopic frame, the manually annotated ground truth, and the segmentation results generated by the model.

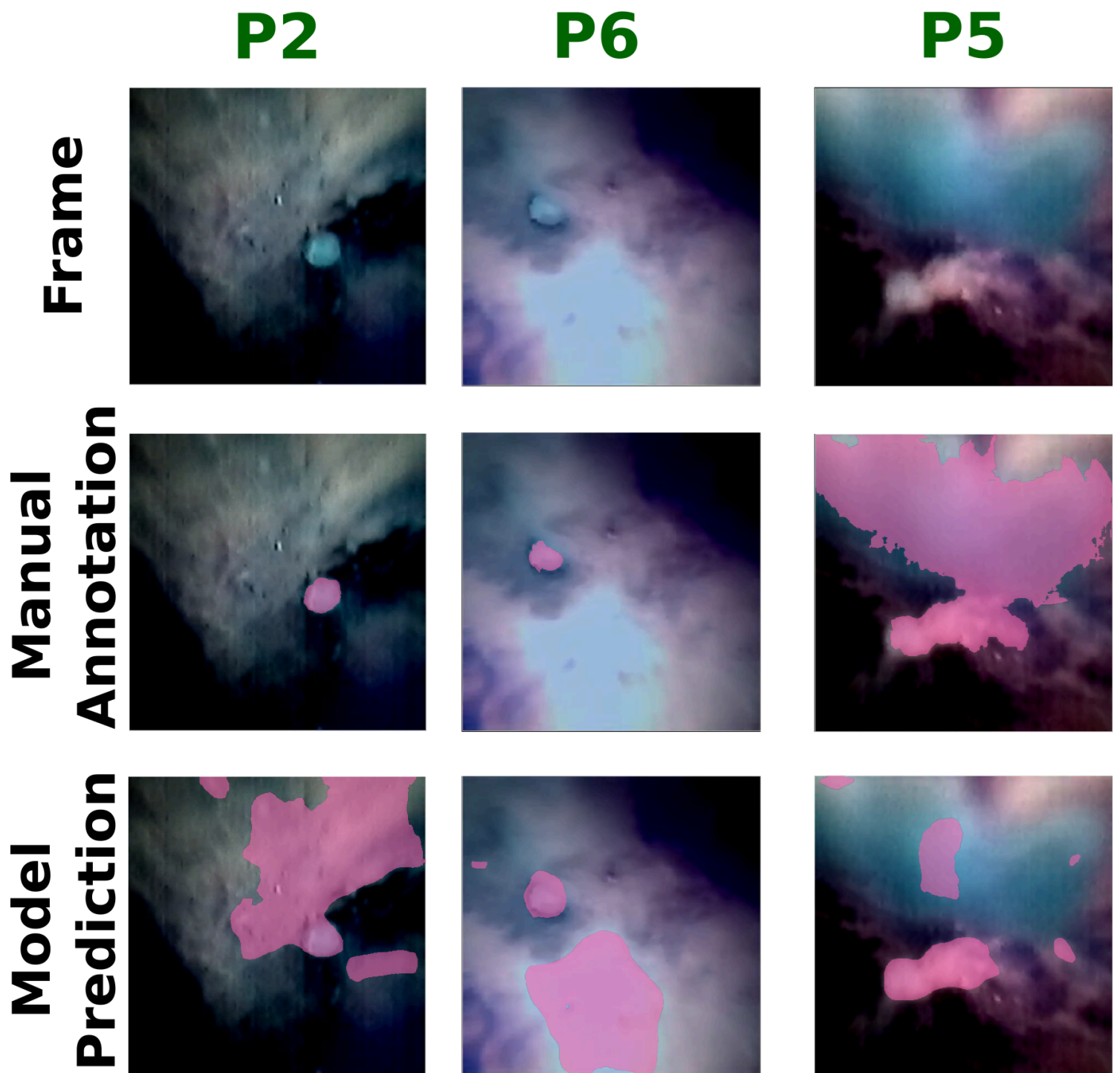


Figure 4

Representative failure cases with poor localization. Each column is one example; rows show the original endoscopic frame, the manually annotated ground truth, and the segmentation results generated by the model. In all three cases the stone was correctly detected but poorly localized (Dice 0.08 to 0.31): in frames obscured by stone dust and irrigation turbulence the model over-segments background regions, whereas for a large stone it captures only part of the true extent.