Anatomically-Enhanced URO dot AI: Multi-Stage Fine-Tuning of Foundation Models for Precise Urinary Stone Segmentation

Byeongin Moon*1 BLUEIN@AIDOT.AI
Jaehoon Oh*2 OJJAI@HANYANG.AC.KR
Young Eun Yoon*3 UROLOGISTYOON@HANYANG.AC.KR
Dong Keon Lee4 STOLENEGG@GMAIL.COM
Gihun Lee1 1007HOON@AIDOT.AI
Mingyeong Park1 CYZLALSRUD96@AIDOT.AI
Heesun Choi1 HEESUN@AIDOT.AI
Hansol Choi†1 SOLCHOI@AIDOT.AI
Jaehoon Jeong†1 JMAN@AIDOT.AI
1 *AIDOT Inc., Seoul, Republic of Korea*
2 *Department of Emergency Medicine, College of Medicine, Hanyang University, Seoul, Republic of Korea*
3 *Department of Urology, College of Medicine, Hanyang University, Seoul, Republic of Korea*
4 *Department of Emergency Medicine, College of Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea*

Editors: Under Review for MIDL 2026

Abstract

Automated segmentation of urinary stones in non-contrast CT (NCCT) is challenging due to small lesions, class imbalance, and voxel sparsity. We propose a hierarchical fine-tuning framework based on VISTA3D with (1) anatomical mapping to define regions of interest and (2) organ-aware stone segmentation. On 119 test cases, the method achieved 95.69% stone-level and 96.64% patient-level sensitivity, demonstrating improved detection performance through anatomical context.

Keywords: Urinary Stone, Foundation Model, Hierarchical Fine-Tuning, VISTA3D.

1. Introduction

Accurate localization of urinary stones is essential for clinical intervention. While research primarily focuses on kidney stones (Li et al., 2022; Cui et al., 2021; Elton et al., 2022), automated detection of ureteral and bladder stones remains challenging due to the ureter's tortuous anatomy and a scarcity of annotated datasets. Furthermore, conventional statistical probability maps struggle to generalize across diverse patient anatomies (Längkvist et al., 2018).

To address this, we propose URO dot AI, leveraging the VISTA3D (Liu et al., 2024) foundation model pre-trained on large-scale medical data. Building on our previous work

---

* Contributed equally
† Corresponding author

on organ segmentation (Jang et al., 2024), we first establish explicit anatomical priors to define the boundaries of the urinary tract. This hierarchical approach effectively restricts the search space, isolating micro-lesions within the Kidney-Ureter-Bladder (KUB) region. Consequently, our framework enhances KUB-wide detection precision, boosting intra-renal sensitivity while mitigating the severe false positives and morphological challenges associated with the ureter.

2. Methodology

2.1. Hierarchical Learning Framework Overview

The proposed pipeline (Figure 1) integrates ROI localization via UNETR (Hatamizadeh et al., 2022) and a two-stage hierarchical fine-tuning of the VISTA3D foundation model. This progression from coarse localization to refined anatomical mapping enables the detection of microscopic stones within the complex morphology of the urinary tract.

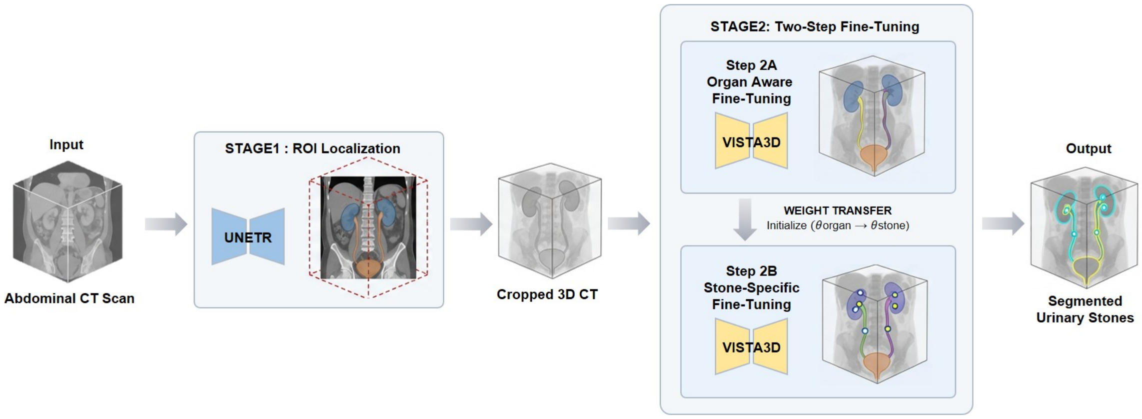

Figure 1: Overview of the proposed multi-stage hierarchical fine-tuning framework.

2.2. Stage 1: ROI Localization and Cropping (UNETR)

To optimize memory efficiency and narrow the search space, UNETR generates coarse kidney and bladder masks. These define a 3D ROI crop that encompasses the entire urinary trajectory while discarding irrelevant anatomical regions.

2.3. Stage 2: Two-Step Fine-Tuning Strategy (VISTA3D)

We utilize VISTA3D, a medical foundation model pre-trained on 11,454 CT scans, as our backbone. To address the challenges of data scarcity and the narrow, variable structure of the ureter, we implement a hierarchical knowledge transfer strategy:

- **Step 2A (Organ-Aware)**: The model is specialized to segment nine structures—the bladder, kidneys, and three ureteral segments (proximal, mid, distal—to establish robust spatial priors.

- **Step 2B (Stone-Specific)**: Using the organ context as a spatial constraint, the model undergoes targeted fine-tuning to differentiate stones from other high-density extra-urinary voxels.

2.4. Datasets

All datasets utilized in this study consist of non-contrast CT (NCCT) scans. The Organ Dataset (448 train / 72 val / 56 test) and Stone Dataset (809 train / 219 val / 119 test) were employed for organ and stone segmentation, respectively.

3. Results

The performance of our hierarchical approach (Table 1) significantly surpassed the baseline, boosting Stone-PPV from 44.37% to 60.79% while maintaining 95.69% sensitivity. This 16.42% absolute improvement confirms that organ-aware priors effectively suppress false positives in anatomically complex regions. Regional analysis (Table 2) further highlights 100% sensitivity in proximal ureters and the bladder, validating the framework's ability to resolve diagnostic ambiguities in 3D volumes.

Table 1: Performance summary of the proposed method.

Method	**Patient-level Metrics**			**Stone-level Metrics**	
	Sensitivity (Strict)	**Sensitivity** (Non-Strict)	**Mean PPV**	**Sensitivity**	**PPV**
Stone FT only	87.39%	95.80%	56.16%	92.34%	44.37%
Organ & Stone FT (proposed)	93.28%	96.64%	72.37%	95.69%	60.79%

Note. FT: Fine-tuning; Strict: all stones detected; Non-strict: at least one stone detected per patient.

Table 2: Stone-level sensitivity across different anatomical regions.

Method	**R-Kid**	**L-Kid**	**R-Prox**	**L-Prox**	**R-Mid**	**L-Mid**	**R-Dist**	**L-Dist**	**Bladder**
Stone FT only	90.38%	90.24%	97.37%	100.00%	100.00%	50.00%	80.00%	100.00%	100.00%
Organ & Stone FT (proposed)	96.15%	92.68%	100.00%	100.00%	83.33%	75.00%	95.00%	88.89%	100.00%

Note. FT: Fine-tuning; R/L: Right/Left; Kid: Kidney; Prox/Mid/Dist: Proximal/Mid/Distal Ureter.

4. Conclusion

We demonstrate that hierarchical anatomical priors in a 3D foundation model significantly enhance microscopic stone detection by constraining the search space and suppressing false positives, even under data sparsity. This framework provides a robust, scalable solution for automated urological diagnostics, successfully bridging large-scale foundation models with specialized clinical tasks.

Acknowledgments

This work was conducted by AIDOT Inc. The authors are grateful to Hanyang University Hospital and Seoul National University Bundang Hospital for their essential support in providing the clinical datasets, which were accessed and used under the approval of their respective Institutional Review Boards (IRBs).

References

Yingpu Cui, Zhaonan Sun, Shuai Ma, Weipeng Liu, Xiangpeng Wang, Xiaodong Zhang, and Xiaoying Wang. Automatic detection and scoring of kidney stones on noncontrast ct images using s.t.o.n.e. nephrolithometry: Combined deep learning and thresholding methods. *Molecular Imaging and Biology*, 23(3):436–445, 2021. doi: 10.1007/s11307-020-01554-0.

Daniel C Elton, Evrim B Turkbey, Perry J Pickhardt, and Ronald M Summers. A deep learning system for automated kidney stone detection and volumetric segmentation on noncontrast ct scans. *Medical Physics*, 49(4):2545–2554, 2022. doi: 10.1002/mp.15518.

Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. UNETR: Transformers for 3D medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 574–584, 2022. doi: 10.1109/WACV51458.2022.00181.

Dong-Hyun Jang, Juncheol Lee, Young-Jin Jeon, et al. Kidney, ureter, and urinary bladder segmentation based on non-contrast enhanced computed tomography images using modified u-net. *Scientific Reports*, 14(1):15325, 2024. doi: 10.1038/s41598-024-66045-6.

Dan Li, Chuda Xiao, Yang Liu, Zhuo Chen, Haseeb Hassan, Liyilei Su, Jun Liu, Haoyu Li, Weiguo Xie, Wen Zhong, and Bingding Huang. Deep segmentation networks for segmenting kidneys and detecting kidney stones in unenhanced abdominal ct images. *Diagnostics*, 12(8):1788, 2022. doi: 10.3390/diagnostics12081788.

Pengbo Liu, Yabo Yuan, Yucheng Tang, Dong Yang, Daguang Xu, et al. VISTA3D: A unified segmentation foundation model for 3D medical imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11654–11664, 2024.

Martin Längkvist, Johan Jendeberg, Per Thunberg, Amy Loutfi, and Mats Lidén. Computer aided detection of ureteral stones in thin slice computed tomography volumes using convolutional neural networks. *Computers in Biology and Medicine*, 97:153–160, 2018. doi: 10.1016/j.compbiomed.2018.04.021.