SegPrompt: Using Segmentation Map as a Better Prompt to Finetune Deep Models for Kidney Stone Classification

Wei Zhu[1] Runtao Zhou[1] Yuan Yao[1] Timothy Campbell[2] Rajat Jain[2] Jiebo Luo[1]

[1]University of Rochester [2]University of Rochester Medical Center

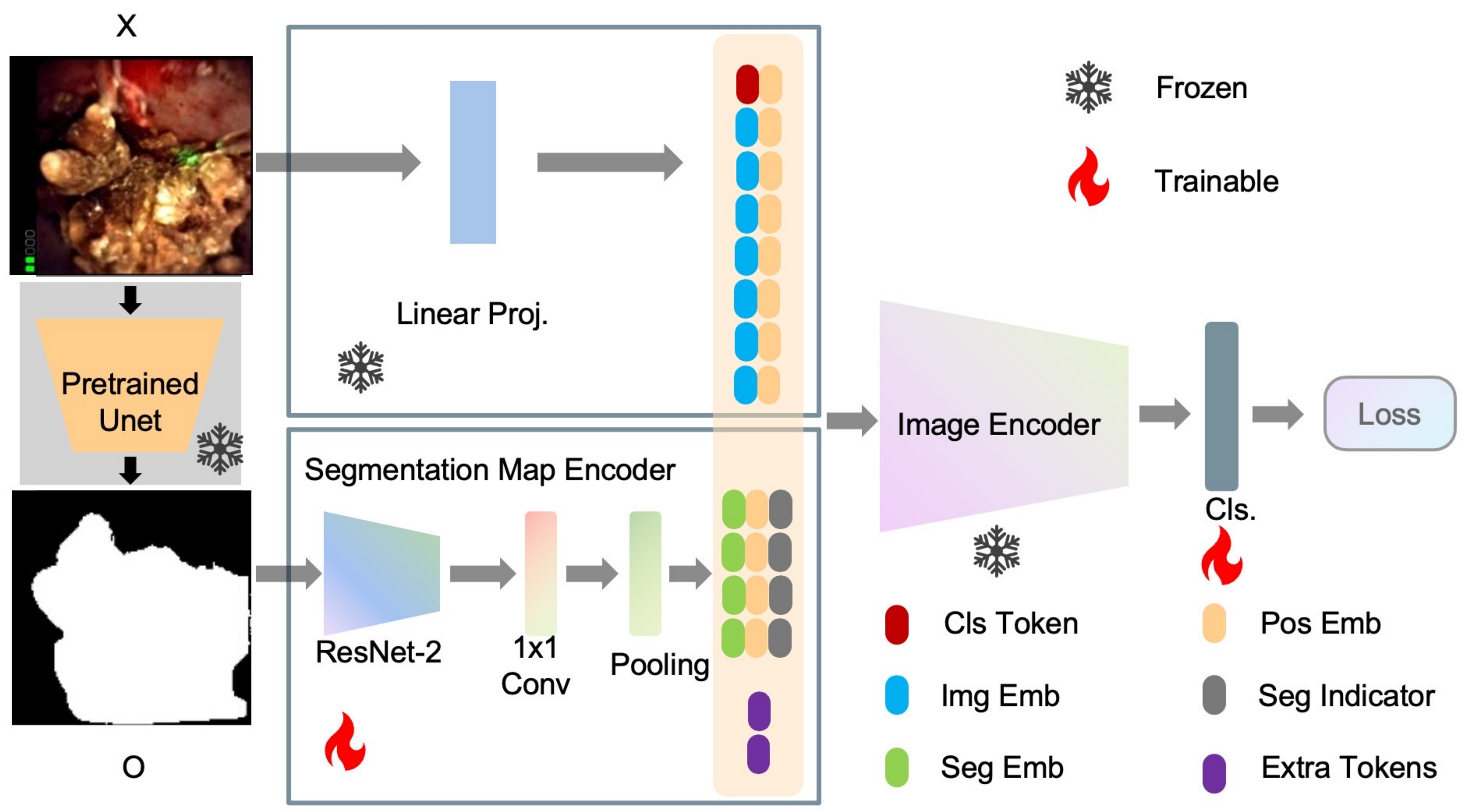
Background

- **Kidney stone disease** (KSD) affects 10% of the US population during their lifetime and results in billions of dollars of the annual cost to society. [1]
- The past standard of fragmenting the stone into small pieces and classify with chemical analysis in laboratory typically takes 1-2 months to get the result, even though the patients can be in a critical condition and suffer from great pain. [2]
- For automatic kidney stone classification, **the limited kidney stone image data** makes it hard to obtain a robust deep model that could generalize to unseen cases with vanilla finetuning. [3]

Introduction

- We propose SegPrompt, a real-time stone-type prediction method based on the deep neural network which can be trained with limited training data.
- SegPrompt integrates the segmentation map into the training process to make the model aware of the regions of interest, which intuitively benefits the classification training process [4].
- The segmentation map is obtained with a pretrained UNet.
- SegPrompt only prompt-tunes a small part of the model, thus alleviating the overfitting problem and improving the classification performance.

Method



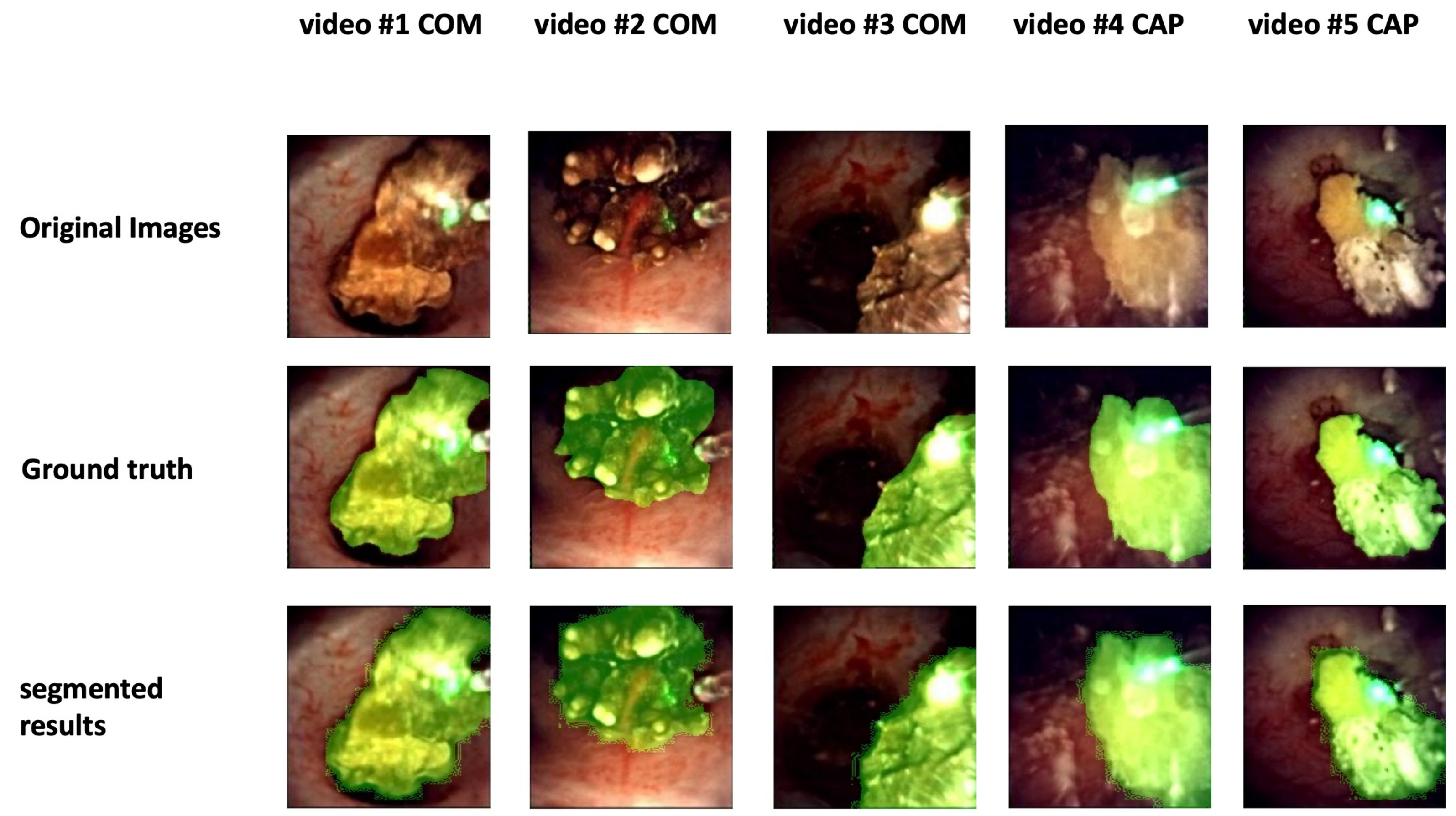
Step 1: Extract the segmentation map with a pretrained UNet.
Step 2: Encode segmentation map into embeddings by the first two blocks of a pretrained ResNet18.
Step 3: Add the position embedding and segmentation indicator to the segmentation embedding to obtain *segmentation tokens*.
Step 4: Concatenate the image tokens, segmentation tokens, and extra learnable tokens.
Step 5: Feed all tokens to the transformer backbone.
Notes: Only update the segmentation map encoder and the last classifier during training.

References

[1] GoApi Chewcharat and Gary Curhan. Trends in the prevalence of kidney stones in the united states from 2007 to 2016. Urolithiasis, 49(1):27–39, 2021.
[2] Gilberto Ochoa-Ruiz, Vincent Estrade, Francisco Lopez, Daniel Flores-Araiza, Jonathan ElBeze, Dinh-Hoan Trinh, Miguel Gonzalez-Mendoza, Pascal Eschw`ege, Jacques Hubert, and Christian Daul. On the in vivo recognition of kidney stones using machine learning. arXiv preprint arXiv:2201.08865, 2022.
[3] Yuhang Zang, Wei Li, Kaiyang Zhou, Chen Huang, and Chen Change Loy. Unified vision and language prompt learning. arXiv preprint arXiv:2210.07225, 2022.
[4] Muhammad Attique Khan, Tallha Akram, Muhammad Sharif, Tanzila Saba, Kashif Javed, Ikram Ullah Lali, Urcun John Tanik, and Amjad Rehman. Construction of saliency map and hybrid set of features for efficient segmentation and classification of skin lesion. Microscopy research and technique, 82(6):741–763, 2019.

Result

Data: 1496 kidney stone images from 5 videos (867 COM stones + 629 CAP stones)



Qualitative Results.

- Improves the F1 score from **96.07% -> 99.45%** (compared with VPT)

Methods	Accuracy	Precision	Recall	F1	AUC
FT	95.07 ± 3.2	94.21 ± 4.7	95.46 ± 2.7	93.73 ± 5.9	95.37 ± 3.4
FT-crop	94.34 ± 4.0	94.14 ± 4.2	94.19 ± 3.9	92.96 ± 5.6	94.17 ± 3.4
FT-concat	95.18 ± 2.5	94.91 ± 2.5	95.13 ± 2.3	94.53 ± 3.0	95.10 ± 2.9
ResNet50	94.75 ± 2.7	94.06 ± 4.1	95.49 ± 2.0	93.55 ± 5.1	95.18 ± 1.6
ResNet50-crop	95.28 ± 3.8	95.71 ± 3.3	94.66 ± 3.7	94.32 ± 4.4	94.58 ± 3.7
ResNet50-concat	96.72 ± 2.6	96.44 ± 3.3	96.93 ± 2.5	96.44 ± 2.8	96.84 ± 2.5
VPT	96.87 ± 1.1	96.60 ± 1.7	96.65 ± 1.3	96.07 ± 2.8	96.52 ± 1.8
VPT-Deep	95.85 ± 0.8	95.49 ± 1.2	95.60 ± 1.2	95.13 ± 2.4	95.32 ± 1.7
SegPrompt	$\mathbf{99.56 \pm 0.3}$	$\mathbf{99.45 \pm 0.4}$	$\mathbf{99.60 \pm 0.3}$	$\mathbf{99.45 \pm 0.5}$	$\mathbf{99.57 \pm 0.4}$
SegPrompt-Deep	99.19 ± 0.3	99.06 ± 0.3	99.24 ± 0.3	99.26 ± 0.2	99.23 ± 0.5

Ablation studies.

- Number of segmentation tokens l_s

# l_s	Accuracy	Precision	Recall	F1	AUC
25	98.81 ± 1.0	98.75 ± 0.9	98.73 ± 1.0	98.69 ± 0.9	98.70 ± 0.8
36	99.13 ± 0.6	99.16 ± 0.4	98.87 ± 1.2	98.78 ± 1.4	98.92 ± 0.6
49	$\mathbf{99.56 \pm 0.3}$	$\mathbf{99.45 \pm 0.4}$	$\mathbf{99.60 \pm 0.3}$	$\mathbf{99.45 \pm 0.5}$	$\mathbf{99.57 \pm 0.4}$
64	99.28 ± 0.2	99.32 ± 0.1	99.22 ± 0.3	99.18 ± 0.4	99.24 ± 0.3
81	99.22 ± 0.6	99.14 ± 0.7	99.25 ± 0.6	99.13 ± 0.7	99.23 ± 0.8

Methods	Accuracy	Precision	Recall	F1	AUC
SegPrompt	99.56 ± 0.3	99.45 ± 0.4	99.60 ± 0.3	99.45 ± 0.5	99.57 ± 0.4
SegPrompt w/o r	99.07 ± 0.6	98.90 ± 0.7	99.11 ± 0.5	98.87 ± 0.8	99.08 ± 0.4
SegPrompt w/o z_e	99.38 ± 0.5	99.34 ± 0.5	99.36 ± 0.8	99.30 ± 0.3	99.42 ± 0.5

- Indicator token r
- Extra learnable token z_e