

Revisiting Active Learning in Histopathology Segmentation: Representativeness and Cost-Effective Active Learning



Hyeseong Lee¹, Jonghyun Lee², Eunsu Kim¹, Sangjeong Ahn³, Sung Hak Lee¹

¹Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Republic of Korea

²Department of Medical and Digital Engineering, Hanyang University College of Engineering, Seoul, Republic of Korea

³Department of Pathology, Korea University Anam Hospital, College of Medicine, Korea University, Republic of Korea

gotjd709@catholic.ac.kr

Motivation

- Image **segmentation** is crucial in **histopathology** for tumor detection, but the process is resource-intensive. To mitigate this, active learning methods, including **uncertainty measurement**, **representativeness** and **Cost-Effective Active Learning (CEAL)**, have been explored.
- We conduct experiments to understand how these methods impact histopathology segmentation **under varying uncertainty levels**.

Data and Setting

- Architecture** :: DeepLabV3Plus with ResNet18 for binary segmentation to detect the tumor region.
- Dataset** :: An **in-house dataset** consisting of a total of 125 gastric cancer whole slide images (WSIs) obtained from five different institutions in South Korea and **Camelyon 16** consisting of 155 WSIs.

- Patch Generation**

- 0.5 Micron Per Pixel
- 512X512 pixel

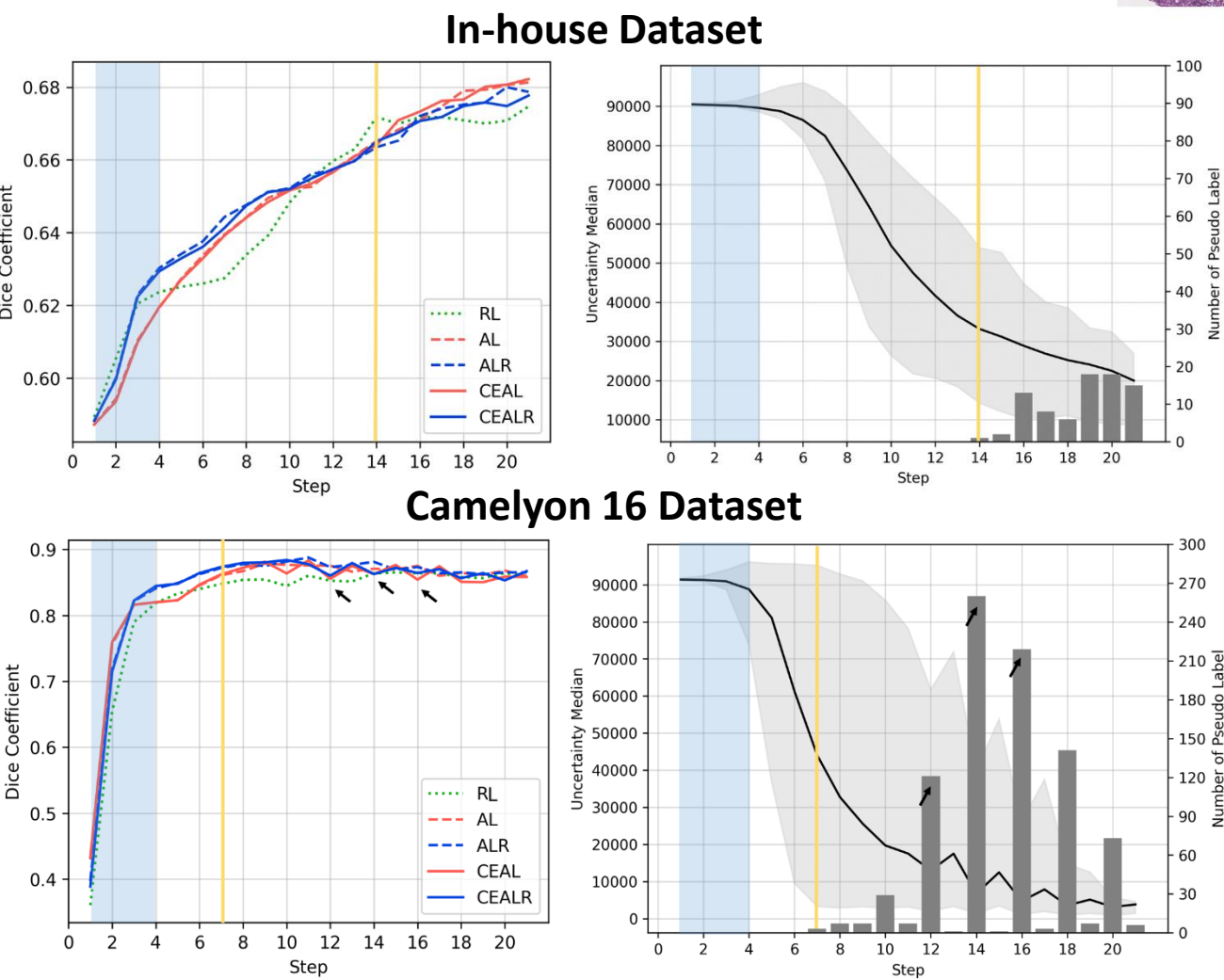
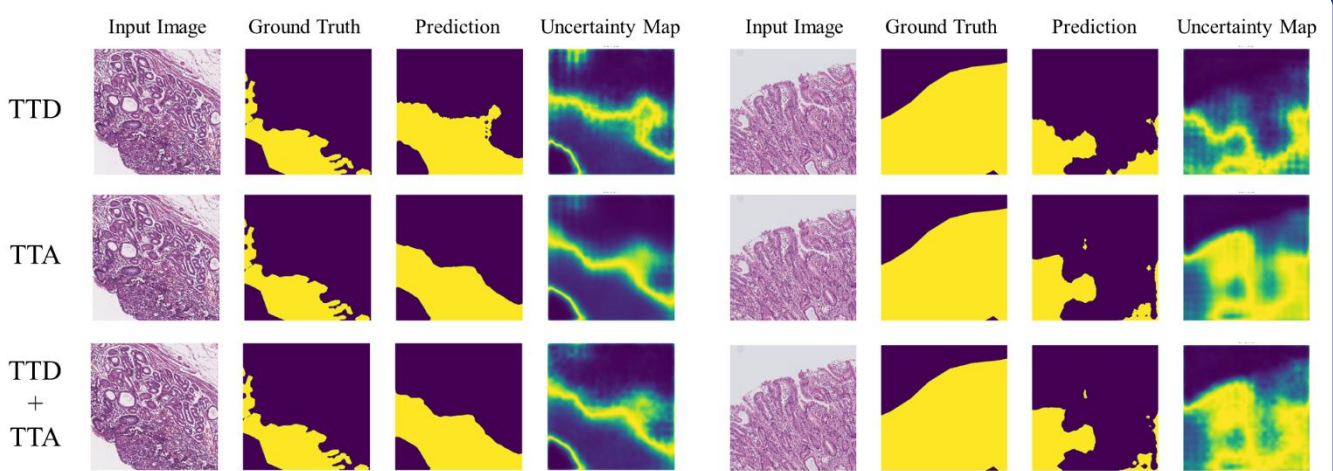
	In-House	Camelyon 16
Labeled data	323	226
Unlabeled data	2908	2034
Query data size	117	99

Methodology

- Uncertainty Estimation** :: Using Test Time Dropout (TTD) and Test Time Augmentation (TTA) methods to jointly estimate **Epistemic uncertainty** and **aleatoric uncertainty**.
- Representativeness** :: representing the similarity between images using a **cosine similarity metric** and then transforming it into a **minimum cover set problem** based on a **threshold**.
- CEAL** :: Using the predictions of samples with low estimated uncertainty as pseudo labels.

Results

- When the prediction results are good (left), the method that considers both TTA and TTD estimates lower uncertainty compared to the method that considers only TTD.
- Conversely, in the opposite case (right), it estimates higher uncertainty.



- RL :: Random Learning
- AL :: Active Learning
- ALR :: Active Learning with Representativeness
- CEAL :: Cost-Effective Active Learning
- CEALR :: CEAL with Representativeness
- black line :: median of uncertainty estimate
- light gray region :: range between the max and min value.
- When there is no difference in uncertainty estimation intervals (light blue region), AL and CEAL methods lag behind ALR and CEALR methods, but they catch up in performance when intervals diverge.
- As the estimation of uncertainty decreases and the use of pseudo labels begins (yellow line), differences arise depending on the presence of CEAL.

Take Home Message

- Representativeness** can prevent performance degradation in methods that consider only Uncertainty when there is little difference in Uncertainty at the beginning.
- CEAL** may not always yield favorable results because even when uncertainty is low, there can still be cases with suboptimal predictions.

Code

