Task-Aware Active Learning for Endoscopic Polyp Segmentation

Pranav Poudel[2*], Shrawan Kumar Thapa[2*], Sudarshan Regmi[2], Binod Bhattarai[1,4], and Danail Stoyanov[1]

[1] University College London, UK
[2] NAAMII, Nepal
[3] University of Aberdeen, Abeerden, UK

Abstract. Semantic segmentation of polyps is one of the most important research problems in endoscopic image analysis. One of the main obstacles to researching such a problem is the lack of annotated data. Endoscopic annotations necessitate the specialist knowledge of expert endoscopists, and hence the difficulty of organizing arises along with tremendous costs in time and budget. To address this problem, we investigate an active learning paradigm to reduce the requirement of massive labelled training examples by selecting the most discriminative and diverse unlabeled examples for the task taken into consideration. To this end, we propose a task-aware active learning pipeline that considers not only the uncertainty that the current task model exhibits for a given unlabelled example but also the diversity in the composition of the acquired pool in the feature space of the model. We compare our method with the competitive baselines on two publicly available polyps segmentation benchmark datasets. We observe a significant performance improvement over the compared baselines from the experimental results. The code and implementation details are available at: https://github.com/bhattarailab/endo-active-learn

Keywords: Active Learning · Computer Assisted Interventions · Semantic Segmentation · Surgical AI

1 Introduction

Polyp segmentation [7,3] is a fundamental research problem in endoscopic image analysis. Automated polyp segmentation can help in the early diagnosis, detection, and treatment of colorectal disease by supporting endoscopists with computer-assisted detection and characterization systems. Such capabilities are needed to advance the toolkit available to endoscopists, enable standardization of adenoma detection rates, and potentially link to future robotic systems and automation [2]. The effectiveness of deep networks for such tasks has already been demonstrated. However, most solutions demand a large number of training examples. Annotating such a large volume of endoscopic data needs domain

[*] These authors made equal contributions.

experts, which incurs an immense cost in time and budget. Therefore, label-efficient methods are of utmost importance. Recently, to address the problem of annotated examples, several self-supervised learning algorithms are proposed [21,14,13,11]. However, the performance of these approaches depends upon the overlap of the pre-text task with the downstream task. As a consequence, it demands a careful design of the pre-text task. Similarly, data augmentations with different geometric transformations are another option to populate the training examples. However, Such approaches do not effectively add true distribution variability because the body organs such as Colons are tubular and rationally invariant.

Active Learning (AL) [24,26,9] has shown a lot of promise in becoming a viable solution to subsample the datasets by discarding redundant and less informative examples in computer vision. In AL, we repeatedly acquire labels using an acquisition function for a subset of an unlabelled set where the acquisition of labels is constrained by budget. Its task is to select the optimal subset of examples that enhance the model's performance when added to the training set. AL methods are emerging gradually in biomedical image analysis [8,19,6,29,15]. NVIDIA's open-source platform MONAI [4] has launched an intelligent interactive data annotation tool called MONAI Label. The workshop on the theme "interpretable and label-efficient learning" [10] was organised in conjunction with MICCAI 2020. PathAL [20] and [22] are some of the recent work on Active Learning for Histopathology Image Analysis.

In active learning, acquisition functions fall mainly into three categories: 1. Uncertainty-based [12,4,31,17], 2. Distribution-based [26,24], and 3. Combining uncertainty and distribution [22,25]. Relying only on uncertainty as a selection criterion helps us to choose examples from the region of the manifolds of the image where the model is less confident. However, it cannot avoid selecting redundant images from the same manifold region, limiting the diversity. On the other hand, distribution-based approaches address this issue by considering the selected samples' diversity. However, it is possible to miss the selection of difficult examples. So, the best bet is to combine the best of both worlds. Yet another difference between these two groups of methods is that uncertainty-based methods are aware of downstream tasks. In contrast, representative-based methods are task-agnostic.

Our contribution lies in developing a novel *task-aware* method for selecting diverse and difficult examples for a downstream task and applying it to novel biomedical image analysis tasks. To this end, we employ Coreset [1] sampling method in the downstream *task-aware* feature space in our active learning pipeline for endoscopic polyp segmentation. Previous work on active learning [24] based on Coreset-based sampling was evaluated on classification problems. Unlike previous work, we also combine the uncertainty-based method to sample the unlabelled data. Our pipeline is as shown in figure 1. We project all the data on the learner's feature space and apply the K-Centre Greedy Algorithm similar to that in [24] to select the diverse examples from the unlabelled dataset that

[4] https:monai.io

comprises a fraction of the total budget. Similarly, we acquire predictions for unlabeled examples and select the top uncertain examples using the Best vs. Second Best Strategy (BvSB)[17]. This would make up the remaining fraction of the total budget. This approach helps us to focus on difficult examples. The sum of examples from both of these approaches equals the total available budget and the trade-off is adjusted by empirical validation.

We summarize our contributions in the following points.

- We propose a novel task-aware coreset-based selection method in an active learning pipeline where we combine the uncertainty-based sampling technique with the task-aware coreset-based sampling technique.
- We compare the proposed method with multiple task-agnostic approaches based on Coreset, and Variational Autoencoder on challenging datasets for endoscopic polyp segmentation.
- We perform extensive quantitative and qualitative experiments to validate our approach.

2 Method

Active Learning is an iterative process to select a subset of examples (X^s) from a large pool of unlabelled sets (X) to query their labels (Y). We label the examples (x, y) $\subset$ ($X \times Y$) incrementally and add them to a set of the labelled examples (X^l). The labeled examples are used to train a network minimizing the objective of the end task ($\mathcal{L}$). Eqn 1 summarises the Active Learning pipeline. Given any sampling function $\mathcal{A}$, the main objective of AL is to minimize the no. of selection stages n to reduce the no. of examples for which labels need to be queried.

$$\min_{n} \min_{\mathcal{L}} \mathcal{A}(\mathcal{L}(x, y; \theta) | X_0^s \subset \cdots \subset X_n^s \subset X). \tag{1}$$

To begin with annotation, we select the first batch X_0^s randomly, where subscript 0 denotes the first selection stage and superscript s indicates a selected set of examples to query their labels. Once Oracle queries their labels, we add those examples to the pool of labeled examples $X^l = \{X_0^s \ U \ \emptyset\}$. These labeled examples act as seed annotations to guide the next selection stages. Figure 1 depicts the proposed method. There are three major components in the pipeline A) Learner, B) Sampler, and C) Oracle. We discuss these in detail below.

2.1 Learner (A)

The role of a learner in the AL pipeline is to learn the parameters for a downstream task from the labeled set of examples. In our case, we are dealing with polyp segmentation. Hence, we choose U-Net [23], a widely-used semantic segmentation architecture for bio-medical image segmentation, to implement the learner. Suppose x represents an image with its corresponding ground-truth label y from labeled set X^l. When we feed in x to the model, the encoder projects the image into a low-dimensional vector, z. And the decoder reconstructs z back

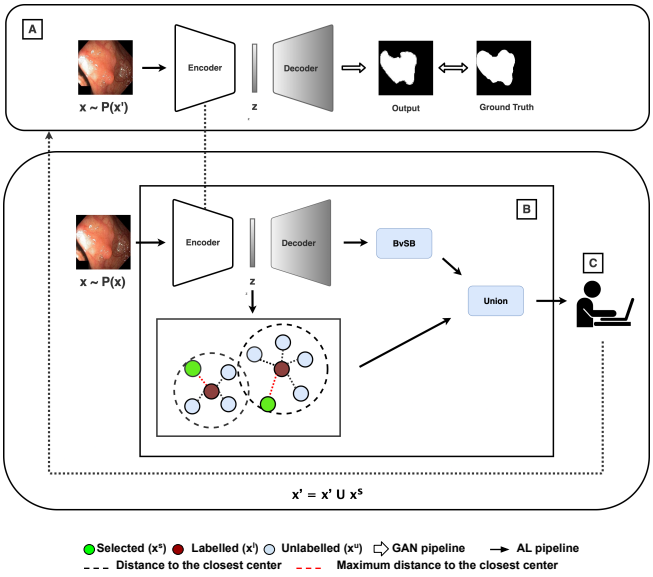

Fig. 1: Illustration of the proposed Active Learning Pipeline. Here, (A) denotes the downstream task model, referred to as the Learner, a semantic segmentation network in this instance. (B) represents a sampler that shares the parameters of the learner, which makes our sampling technique task-aware. We leverage latent representations from both labeled and unlabeled data, utilizing Coreset and BvSB for sampling examples that are diverse and uncertain with respect to the task. These selected samples are subsequently labeled (C). The Learner is then trained on these along with the pre-existing labeled data, and this process iterates until the budget is exhausted.

to the output $\hat{y}$, along with using different levels of features from the encoder. We minimize the objective given in Equation 2 to train the network.

$$\mathcal{L}(y, \hat{y}) = \mathcal{L}_{\mathcal{CE}}(y, \hat{y}) + \mathcal{L}_{dice}(y, \hat{y}) \tag{2}$$

Here, $\mathcal{L}_{\mathcal{CE}}$ is a binary cross-entropy loss [30] and $\mathcal{L}_{dice}$ is dice loss[27]. Once we learn the parameters of U-Net from the available labeled examples, component B of the pipeline, the sampler, comes into play.

2.2 Sampler (B)

Uncertainty Based Sampling: In this stage, we feed unlabeled images to the model trained in the first stage and obtain the respective segmentation mask. Then, we use the BvSB method[17] to compute the uncertainty of the current model on given unlabeled images. This method uses the difference between the highest and the second-highest probability score predicted by the model. In the

segmentation task, for each pixel (i, j), where $i \in H$, $j \in W$, the model predicts a categorical distribution denoted by a vector $\hat{y}(i, j) \in [0, 1]^C$, where C is the total number of distinct classes in the task. H and W are the height and the width of the input image/segmentation mask respectively. Then, the Best vs Second Best Score for each pixel is calculated as follows:

$$\mathcal{BvSB}(\hat{y}(i,j)) = 1 - [\max_{k \in \hat{y}(i,j)} \hat{y}(i,j) - \max_{l \in \hat{y}(i,j) \setminus k} \hat{y}(i,j)] \tag{3}$$

The smaller the difference in top-2 class prediction, the higher the uncertainty of the example for the model. Since polyps segmentation is a binary segmentation task, the score calculation for each image can be simplified as shown in the following equation:

$$\mathcal{BvSB}(\hat{y}) = 1 - \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} [\max(\hat{y}(i,j)) - \min(\hat{y}(i,j))] \tag{4}$$

We sample top B_u uncertain examples using the uncertainty score given by eqn 4. The uncertainty is estimated from the predictions of the learner that models our downstream task; thus making this component task-aware.

Task-Aware Core-set: We sample B_d diverse examples using parameters from stage A to project labeled and unlabeled images into a 512-dimensional latent space z. The latent representations, optimized for the downstream task, facilitate the selection of a core-set [1] from the dataset for label querying. The core-set is computed using the K-Center Greedy algorithm in this space. The diversity selection process is illustrated in Figure 1 block B (lower sub-block), where nodes represent images with latent features from the learner's encoder: sky-blue for unlabeled, red for labeled, and green for newly selected examples. Post-selection, green nodes transition to red. Edges indicate Euclidean distances, with length proportional to distance magnitude. The selection algorithm identifies the farthest nodes from each red node's nearest neighbors, avoiding duplicates that offer redundant information to the downstream task. This Task-Aware Coreset (TA-Coreset) approach ensures a representative subset by excluding redundant examples, optimizing for the downstream task.

Combining TA-Coreset with Uncertainty Here, B_u and B_d are functions of γ such that $B_u = \gamma * B$ and $B_d = (1 - \gamma) * B$, where B is the total number of examples at a selection stage. We obtain X^s from the union of uncertain and diverse sets having B_u and B_d examples in each set, respectively. Note that both sets are disjoint. In fact, it is done to show the complementary nature of these methods. If in case, the same example(s) is selected by both methods, the union of the two sets does not result in a total of B samples. In such cases, we continue the same sampling scheme for the remaining number to be sampled from the budget. Before proceeding to the next cycle, we add the selected examples in the current cycle to the existing pool of labeled examples. This iterative sampling process is summarized in algorithm 1.

Algorithm 1: Combining TA-Coreset with Uncertainty

Input: latent representation of data $\mathbf{Z}$, existing pool $\mathbf{s}^0$, unlabelled set X^U, a budget B, sampling ratio γ

Initialize $\mathbf{s} = \phi$

repeat

- $B_i = B - |\mathbf{s}|$
- $B_u = \gamma * B_i$
- $B_d = (1 - \gamma) * B_i$
- $\mathbf{s_u} = \{B_u$ samples from X^U using equation 4$\}$
- $\mathbf{s_d} = \{B_d$ samples from $X^U \cup \mathbf{s}^0$ using k-center greedy$\}$
- $\mathbf{s} = \mathbf{s} \cup \mathbf{s_c} \cup \mathbf{s_d}$
- $\mathbf{s}^0 = \mathbf{s}^0 \cup \mathbf{s}$
- $X^U = X^U \setminus \mathbf{s}^0$

until $|\mathbf{s}| = B$

return s

2.3 Oracle (C)

We query the labels of the selected set, X^s from the Oracle. After retrieving their label, the selected set is appended to the labelled set ($X^l = X^l \cup X^s$), and the selected set ($X^s = \emptyset$) is emptied. This cycle is repeated till the budget limit is reached.

3 Experiments and Results

Datasets: We perform extensive experiments on Kvasir-SEG [16] and CVC-ClinicDB[5]. Kvasir-SEG consists of 1,000 colonoscopy images with polyp masks. We used 900 of them for training and the rest for validation. We reported our performance on a smaller test dataset provided by the same project, identified as sessile-Kvasir-SEG consisting of 196 images. Clinic-DB is a similar dataset, but data points amounting only to 612 images. We randomly selected 112 images as a test set and 100 from the remaining 500 images as a validation set. The validation set is used to select the best model during training. The model is then evaluated on the respective test sets.

Baselines: We compared our method with a wide range of competitive baselines. **Random** is the technique most commonly used to sub-sample the training examples. We applied **Principal Component Analysis (PCA)** [28], and compressed the images to the dimension of 512, which is equal to that of the latent representations in our method. We applied Coreset [24] on PCA-compressed features, which we denote as **PCA-Coreset**. **Uncertainty** [17] is another sampling technique to find the most informative examples. Finally, we compared our performance with **VAAL** [26], one of the most popular task-agnostic active learning methods. Although our method is not task-agnostic, we made a comparison with VAAL to shed light on the importance of task-aware active learning.

Implementation Details: We utilized U-Net for polyp segmentation, although our modular approach allows for straightforward integration of alternative architectures into the pipeline. We used Adam optimizer [18] ($\beta_1 = 0.9, \beta_2 = 0.999$) with a learning rate of 2×10^{-4}. The model was trained over 100 epochs per cycle with a batch size of 8, using images resized to 256×256 and normalized by dividing pixel values by 255, then centering at a mean of 0.5 with a standard deviation of 0.5. For AL experiments, the labeled pool for the Kvasir-SEG dataset was initialized with 100 random examples, and for the Clinic-DB dataset with 40, both maintaining a sampling budget of 100. Model performance was evaluated using mean Intersection Over Union (mIOU) across different selection stages, averaged over 5 random seeds.

3.1 Quantitative Evaluations

Table 1 shows IOU of all baselines and our method across various datasets. We can observe that using task-aware features to select diverse examples helps improve performance from respective task-independent counterparts. TA-Coreset outperforms other baseline methods most of the time in Kvasir-SEG and all the time in CVC-ClinicDB dataset. In contrast, task agnostic methods PCA-coreset and VAAL[26], which employs the sampling method on latent representations of the images, demonstrate lower performance. This performance gap highlights the importance of making image representations task-aware in active learning.

Table 1: Performance Comparison on image features

Method	Kvasir-Seg					CVC-ClinicDB				
	200	300	400	500	600	80	120	160	200	240
Random	68.14	73.67	77.48	81.63	85.61	78.92	82.38	84.83	85.82	86.77
VAAL	**68.38**	73.58	**78.41**	81.82	85.59	79.02	82.42	84.40	85.21	87.69
PCA-Coreset	64.71	70.20	75.43	78.48	81.96	79.65	83.67	86.22	87.47	88.42
TA-Coreset(ours)	67.99	**74.02**	78.28	**81.92**	**86.89**	**79.94**	**84.33**	**87.07**	**88.14**	**88.78**

Combining Uncertainty and TA-Coreset: Tuning γ: We performed experiments on combining samples from the Uncertainty-based acquisition function and TA-Coreset. We determine the optimal value of $\gamma = 0.5$ in both datasets when sampling fromγ from 0 to 1, where $\gamma = 0$ corresponds to TA-Coreset and $\gamma = 1$ to Uncertainty. Table 2 shows the impact of adding examples based on the distribution of unlabeled images to the uncertain set. This approach boosts performance, especially in the early stages in the Kvasir-SEG dataset. In later stages, performance plateaus as the pool of unique, informative examples shrinks. In particular, this method complements the uncertainty information and significantly outperforms the diverse sampling of the image/feature space (Tables 1 and 2). The metrics show a marked improvement over PCA and TA-Coreset, particularly in the later stages. In the case of the CVC-ClinicDB dataset, the performance improved much more than with other baseline methods. This could

be due to a smaller set of 400 images compared to 900 in Kvasir, which is similar to the observation in later stages of the Kvasir-Seg dataset.

Table 2: Performance Comparison on Combination with Uncertainty.

Method	Kvasir-Seg					CVC-ClinicDB				
	200	300	400	500	600	80	120	160	200	240
Uncertainty (Unc)	66.69	73.96	81.08	**85.97**	89.03	79.18	84.01	85.96	87.93	88.94
Unc + PCA	65.92	71.73	77.73	83.46	87.79	**80.25**	83.46	87.05	87.78	**88.98**
Unc + TA-Corset	**67.93**	**75.35**	**82.68**	85.40	**89.24**	79.33	**84.60**	**87.12**	**88.02**	88.61

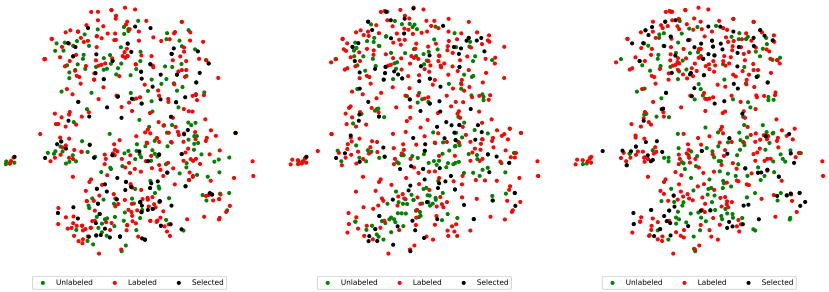

Fig. 2: T-SNE plots showing the comparison of a selection of unlabelled examples at the third selection stage on Kvasir-SEG. Left, middle, and right plots show selection by Random, PCA-Coreset, and TA-Coreset, respectively (Zoom in for a better view).

3.2 Qualitative Evaluations

Task Aware vs Task Agnostic Features: Fig 2 demonstrates the importance of task-aware features for selecting an optimal set. The diagram depicts green dots as unlabeled examples, red dots as previously labeled examples, and black dots as currently selected examples, which were green prior to selection. The plots illustrate that random sampling (left) uniformly selects across the manifold, whereas PCA-Coreset (middle) achieves more dispersed coverage. This illustrates the effectiveness of Coreset in reflecting the input data distribution. Conversely, TA-Coreset (right) selectively concentrates examples in the upper and lower regions, leaving the middle sparse. These findings suggest that task-aware latent representations outperform task-agnostic ones in sampling efficiency.
Common Examples between diversity and uncertainty: To demonstrate the importance of combination, we calculate the number of common examples between TA-Coreset and Uncertainty. We observe very few examples are common

selection from both methods across various selection stages i.e. less than 10. This shows the complementary nature of Uncertainty and diversity.

4 Conclusions

In this paper, we present a novel task-aware active learning framework for endoscopic image analysis. We combined diversity-based sampling on task-aware feature space with uncertainty-based sampling and tested the proposed method on two publicly available polyp segmentation datasets. We observed a superior performance from the extensive experiments compared to the multiple competitive baselines, validating our hypothesis that the features required for sampling should be task-aware. Furthermore, we also noted that the addition of model uncertainty information proved to be complementary though the performance starts getting competitive with the availability of a smaller pool of unlabeled sets.

Acknowledgments. This work was supported by Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS)[203145Z/16/Z]; Engineering and Physical Sciences Research Council(EPSRC)[EP/P027938/1, EP/R004080/1, EP/P012841/1]; The Royal Academy of Engineering Chair in Emerging Technologies scheme; and the EndoMapper project by Horizon 2020 FET (GA 863146).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Agarwal, P.K., Har-Peled, S., Varadarajan, K.R.: Geometric approximation via coresets survey. Current Trends in Combinatorial and Computational Geometry, E. Welzl, ed., Cambridge University Press, Cambridge (2006)
2. Ahmad, O.F., Mori, Y., Misawa, M., Kudo, S.e., Anderson, J.T., Bernal, J., Berzin, T.M., Bisschops, R., Byrne, M.F., Chen, P.J., et al.: Establishing key research questions for the implementation of artificial intelligence in colonoscopy: a modified delphi method. Endoscopy **53**(09), 893–901 (2021)
3. Ali, S., Dmitrieva, M., Ghatwary, N., Bano, S., Polat, G., Temizel, A., Krenzer, A., Hekalo, A., Guo, Y.B., Matuszewski, B., et al.: Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy. Medical image analysis **70**, 102002 (2021)
4. Beluch, W.H., Genewein, T., Nürnberger, A., Köhler, J.M.: The power of ensembles for active learning in image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9368–9377 (2018)
5. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized medical imaging and graphics **43**, 99–111 (2015)
6. Bernhardt, M., Castro, D.C., Tanno, R., Schwaighofer, A., Tezcan, K.C., Monteiro, M., Bannur, S., Lungren, M.P., Nori, A., Glocker, B., et al.: Active label cleaning for improved dataset quality under resource constraints. Nature communications **13**(1), 1–11 (2022)

7. Brandao, P., Mazomenos, E., Ciuti, G., Caliò, R., Bianchi, F., Menciassi, A., Dario, P., Koulaouzidis, A., Arezzo, A., Stoyanov, D.: Fully convolutional neural networks for polyp segmentation in colonoscopy. In: Medical Imaging 2017: Computer-Aided Diagnosis. vol. 10134, pp. 101–107. SPIE (2017)
8. Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. Medical Image Analysis **71**, 102062 (2021)
9. Caramalau, R., Bhattarai, B., Kim, T.K.: Sequential graph convolutional network for active learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9583–9592 (2021)
10. Cardoso, J., Van Nguyen, H., Heller, N., Abreu, P.H., Isgum, I., Silva, W., Cruz, R., Amorim, J.P., Patel, V., Roysam, B., et al.: Interpretable and Annotation-Efficient Learning for Medical Image Computing: Third International Workshop, iMIMIC 2020, Second International Workshop, MIL3ID 2020, and 5th International Workshop, LABELS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings, vol. 12446. Springer Nature (2020)
11. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
12. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: International Conference on Machine Learning. pp. 1183–1192. PMLR (2017)
13. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al.: Bootstrap your own latent-a new approach to self-supervised learning. Advances in neural information processing systems **33**, 21271–21284 (2020)
14. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
15. Hoi, S.C., Jin, R., Zhu, J., Lyu, M.R.: Batch mode active learning and its application to medical image classification. In: Proceedings of the 23rd international conference on Machine learning. pp. 417–424 (2006)
16. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., Lange, T.d., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: International Conference on Multimedia Modeling. pp. 451–462. Springer (2020)
17. Joshi, A.J., Porikli, F., Papanikolopoulos, N.: Multi-class active learning for image classification. In: 2009 ieee conference on computer vision and pattern recognition. pp. 2372–2379. IEEE (2009)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
19. Li, C.T., Tsai, H.W., Yang, T.L., Lin, J.C., Chow, N.H., Hu, Y.H., Cheng, K.S., Chung, P.C.: Imbalance-effective active learning in nucleus, lymphocyte and plasma cell detection. In: Interpretable and Annotation-Efficient Learning for Medical Image Computing, pp. 223–232. Springer (2020)
20. Li, W., Li, J., Wang, Z., Polson, J., Sisk, A.E., Sajed, D.P., Speier, W., Arnold, C.W.: Pathal: An active learning framework for histopathology image analysis. IEEE Transactions on Medical Imaging (2021)
21. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: NeurIPS (2017)
22. Mahapatra, D., Poellinger, A., Shao, L., Reyes, M.: Interpretability-driven sample selection using self supervised learning for disease classification and segmentation. IEEE transactions on medical imaging **40**(10), 2548–2562 (2021)

23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
24. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations (2018)
25. Shi, X., Dou, Q., Xue, C., Qin, J., Chen, H., Heng, P.A.: An active learning approach for reducing annotation cost in skin lesion analysis. In: International Workshop on Machine Learning in Medical Imaging. pp. 628–636. Springer (2019)
26. Sinha, S., Ebrahimi, S., Darrell, T.: Variational adversarial active learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5972–5981 (2019)
27. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 240–248. Springer (2017)
28. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometrics and intelligent laboratory systems **2**(1-3), 37–52 (1987)
29. Wu, J., Ruan, S., Lian, C., Mutic, S., Anastasio, M.A., Li, H.: Active learning with noise modeling for medical image annotation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 298–301. IEEE (2018)
30. Yi-de, M., Qing, L., Zhi-Bai, Q.: Automated image segmentation using improved pcnn model based on cross-entropy. In: Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004. pp. 743–746. IEEE (2004)
31. Yoo, D., Kweon, I.S.: Learning loss for active learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 93–102 (2019)

5 Supplementary Material

Algorithm 2: k-Center-Greedy

Input: latent representation of data $\mathbf{Z}$, existing pool $\mathbf{s}^0$ and a budget b
Initialize $\mathbf{s} = \mathbf{s}^0$
repeat
  $u = \arg\max_{i \in [n] \setminus s} \min_{j \in s} \Delta(\mathbf{z}_i, \mathbf{z}_j)$
  $\mathbf{s} = \mathbf{s} \cup \{u\}$
until $|\mathbf{s}| = b + |\mathbf{s}^0|$
return $\mathbf{s} \setminus \mathbf{s}^0$

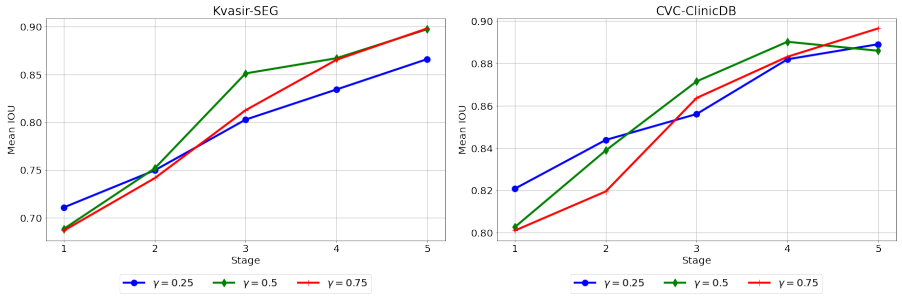

Fig. 3: Performance study on different values of γ for combining Uncertainty and TA-Coreset $(1 - \gamma)$ on Kvasir (left) and CVC (right). When $\gamma = 0$, it is equivalent to TA-Coreset. We uniformly vary the weight from 0.25 to 0.75. This graph shows that our method is complementary to Uncertainty based Active Learning methods.

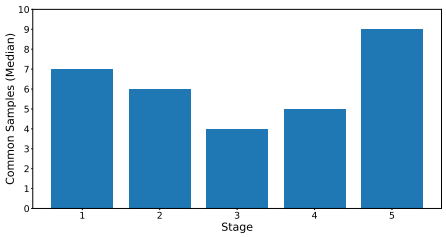

Fig. 4: Median Number of Common Examples sampled by TA-Coreset and Uncertainty at each stage of sampling.

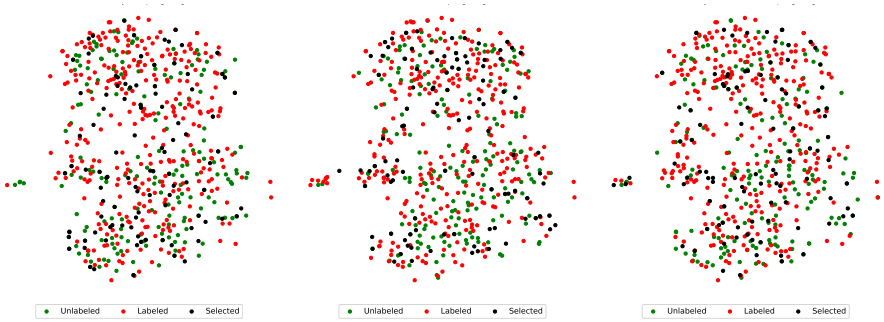

Fig. 5: T-SNE plots showing the comparison of selection of unlabelled examples at the third selection stage on Kvasir-SEG. Left, middle, and right plots show selection by Uncertainty, TA-Coreset, and Uncertainty + TA-Coreset, respectively (Zoom in for the better view). We can see that TA-Coreset (Middle) misses the uncertain/informative example, as recognized by the uncertainty-based sampling method (Left), from the lower-middle part of the image space. When their sampling is combined (Right), we can see it covers regions with both uncertain examples and diverse task-aware features.

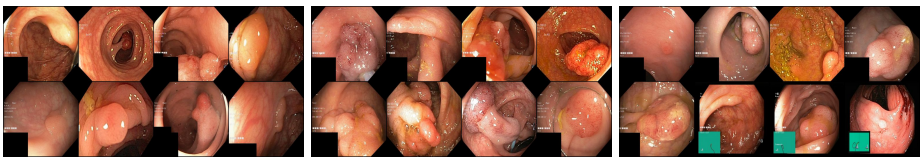

Fig. 6: Examples sampled by Uncertainty+TA-Coreset method in third stage of selection which are A) also sampled by Uncertainty-based method, but not Coreset B) also sampled by Coreset, but not Uncertainty-based method C) not sampled by both (unique to the combination)