TRAIL Summer Workshop' 25
Project Proposal

Full Name of Team Leader	Clément Fuchs* (UCLouvain) - Maxence Wynen (UCLouvain) Benoît Gérin (UCLouvain) - Laura Galvez Jimenez (ULB) – Arthur Lefebvre (Imperial College London) * Corresponding leader: clement.fuchs@uclouvain.be
Project Title	Adapting foundational vision models to instance segmentation and data quality assessment with few annotations: application in histopathology.
Profile of the Team Leader(s)	Clément Fuchs* - Maxence Wynen – Benoît Gérin – Laura Galvez Jimenez * Corresponding leader: clement.fuchs@uclouvain.be Clément Fuchs: PhD student at UCLouvain working on the adaptation of foundational models for natural image classification and medical image segmentation. Maxence Wynen: PhD student at UCLouvain working on instance segmentation applied to MRI and multiple sclerosis. Benoît Gérin: PhD student at UCLouvain currently working on self-supervised and weakly-supervised learning for drug discovery and repurposing. Laura Galvez Jimenez: PhD student at ULB working on the study of robust deep learning in the presence of imperfect annotations in digital pathology applications. Arthur Lefebvre: PhD student at Imperial College London in artificial intelligence with a focus on computational cardiology.
Abstract	Instance segmentation enables both class-level and instance-level analysis, making it invaluable for biomedical imaging applications such as histopathology. However, obtaining annotations for instance segmentation is significantly more labor-intensive than for semantic segmentation, leading to a scarcity of labeled data and specialized models for this task. At the same time, several self-supervised foundation models using large quantities of unlabeled data have been proposed to be used as general-purpose feature extractors. Adapting those models to instance segmentation with only a few annotated patches offers a promising path to reduce annotation burden while leveraging their robustness and performance. In this context, recent studies have demonstrated the feasibility of repurposing these models for instance segmentation in histopathology. However, the potential of domain-specialized semantic segmentation models and other, more recent, foundation models remains underexplored. Thus, we propose broadening the scope of these studies to newly released foundational models for histopathology, as well as specialist models, e.g. models trained for semantic segmentation. Additionally, we will provide a more comprehensive benchmark for these approaches on a wider range of datasets. This work aims to reduce the cost of instance-level annotation, accelerate the deployment of AI tools in clinical and research workflows, and enable broader access to instance segmentation in resource-constrained settings where large-scale annotation is not feasible. Moreover, we will study the impact of imperfect training data on downstream performances by leveraging a previously collected re-annotation of the MoNuSac challenge

	dataset as well as artificial corruptions, and propose a data quality assessment tool, in the form of an annotation quality predictor. Although our work will focus on histopathology, it could apply to other domains with similar needs, such as multiple sclerosis lesion instance segmentation in brain MRI.
Project Objectives	Leverage foundational models and specialized segmentation models to provide accurate instance segmentation in H&E-stained whole slide images using only a few patches with instance annotations.
Project Dataset	Lizard: https://www.kaggle.com/datasets/aadimator/lizard-dataset (495000 nuclei / Colon only). PanNuke: https://huggingface.co/datasets/RationAI/PanNuke (205000 nuclei from different tissues). NuCLS: https://sites.google.com/view/nucls/home (220000 nuclei / Breast only). MoNuSac: https://monusac-2020.grand-challenge.org/ (46 000 nuclei from 37 centers and 71 patients). ConSEP: https://paperswithcode.com/dataset/consep (41 1000 X 1000 images with instance annotations for 24 319 nuclei — Colon only). Ocelot: https://ocelot2023.grand-challenge.org/datasets/ (113 026 nuclei from 663 patches of size 1024 X 1024, 6 organs) Panoptils: https://sites.google.com/view/panoptils/ (814 886 nuclei from 1709 patches, breast cancer) PUMA: https://puma.grand-challenge.org/ (310 patches of size 1024 x 1024 with instance annotations, about 100 000 nuclei in total, 8 different organs)
Background Information	Histopathology analysis tackles automatic processing of very high-resolution images (up to 100k x 100k pixels) obtained from organic tissue, with numerous applications in oncology. Currently, there is a growing interest in foundational models in histopathology [1 – 7,10]. These models are trained from large scale histopathology data, following self-supervised learning paradigms, largely bypassing the need for human annotation. They can later be used as powerful features extractors, at different patches scale or at the whole slide level. They have shown impressive performance for classification or regression, both for patch and slide centric tasks. Additionally, a few works investigate their performance on semantic segmentation [4, 6, 7]. More recently, adaptation methods have been proposed to leverage pre-trained models for instance segmentation tasks [8, 9, 11], but they remain limited in the number of investigated downstream tasks or pre-trained models. <u>Overview of adaptation methods:</u> - CellViT [8] pioneered adapting foundational histopathology to nuclei instance segmentation. They employ an encoder-decoder architecture where the encoder is initialized with a foundational model, i.e. HIPT or SAM. The decoder predicts three maps: a binary semantic segmentation mask, a radial distance to instance center, and a nuclei type probability. Post-processing then assigns an instance label to each foreground pixel, and a nuclei type to each instance. They train and evaluate their approach using the PanNuke dataset. They further investigate the performance of the PanNuke trained model on MoNuSeg for nuclei detection, and analyze cell embeddings on the ConSeP dataset.

- CellVTA [11] uses the same core concepts as CellViT but additionally incorporates high-resolution spatial features in the query of each transformer layer in the encoder. They train and evaluate their approach on CoNIC and PanNuke.
- CellViT++ [9] follows the same core principle of CellViT, but investigates Virchow2, UNI, HIPT and SAM. They first train their model on PanNuke, and later investigate performance on Ocelot, ConSeP, Lizard, NuCLS and PanopTILs. Notably, they fine-tune their model on CoNSeP and present results with a varying amount of finetuning data.

Overview of foundational models for histopathology

- Prov-GigaPath [1] was pretrained on a proprietary dataset of more than 1.3 billion 256 X 256 patches from 171 189 whole slide images. The model uses both a tile encoder, trained following DinoV2, and a slide level encoder, trained following LongNet. It was tested on a task of gene mutation prediction, cancer subtyping and image-text alignment.
- Virchow2 [2] was trained using a proprietary dataset of more than 1.5 million whole slide images. It uses a self-supervised training scheme adapted from DinoV2 and was tested on 7 tiles classification benchmarks.
- REMEDIS [3] uses convolutional networks pre-trained on natural images to adapt them to the medical domain using an annotation free contrastive learning scheme. For histopathology, they pre-train the model on 50 million unlabeled patches from TCGA.
- CTransPath [4] trains a hybrid CNN-Transformer architecture using a training algorithm adapted from the MocoV3 contrastive learning scheme, with 15 million patches extracted from 30 000 whole slide images. They evaluate their model for 5 downstream tasks, i.e. patch retrieval, patch classification, whole-slide image classification, mitosis detection, and colorectal adenocarcinoma gland segmentation.
- HIPT [5] is trained with 10 000 whole slide images. They train three encoders sequentially, at respective scales of 256 X 256, 4096 X 4096, and whole slide. They use a pretraining scheme derived from DINO. They test their model on two downstream tasks, cancer subtyping and cancer survival prediction.
- UNI [6] is trained using more than 100M whole slide images, following the DinoV2 pretraining scheme. They evaluate the model on ROI-level classification, segmentation, retrieval and prototyping, and slide-level classification tasks.
- CONCH [7] uses 1.17 million image-caption pairs to pre-train both a visual and a textual encoder with a vision-language contrastive pretraining scheme. It is tested on histology image classification, segmentation, captioning, and text-to-image and image-to-text retrieval. Notably, the pretraining data include several different staining, compared to the standard H&E staining.
- Quilt-1M [10] is a CLIP model fine-tuned on 1M of image-text pairs obtained from educational youtube videos about histopathology, research articles and twitter.

Data Quality Assessment

An often overlooked challenge in histopathology images analysis is the quality of the ground truth annotations. For instance, [12] has shown how the MoNuSac contains significant noise in its labeling, leading to degraded performances of models trained on the dataset or

	incorrect evaluation of models' effectiveness. Subsequently, they presented a method for extracting subsets with reduced noise from the original dataset. Similarly, [13] investigated how corruptions of the training data could impact the performance of CNNs trained from a random initialization. Over the course of the workshop, we will investigate how imperfect annotations may influence the adaptation of foundational models to the task of instance segmentation. In addition, we will seek to develop a data quality assessment tool in the form of a neural network capable of predicting annotations quality, thereby improving the reliability of performance evaluation in histopathology images analysis. For this, we will leverage a cleaned MoNuSac dataset provided by the ULB, as well as artificial corruptions of patches in the datasets listed above. The neural network will be trained by leveraging initialization with foundational models.
Bibliographic References	[1] Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., ... & Poon, H. (2024). A whole-slide foundation model for digital pathology from real-world data. *Nature*, *630*(8015), 181-188. [2] Zimmermann, E., Vorontsov, E., Viret, J., Casson, A., Zelechowski, M., Shaikovski, G., ... & Severson, K. (2024). Virchow2: Scaling self-supervised mixed magnification models in pathology. *arXiv preprint arXiv:2408.00738*. [3] Azizi, S., Culp, L., Freyberg, J., Mustafa, B., Baur, S., Kornblith, S., ... & Natarajan, V. (2023). Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. *Nature Biomedical Engineering*, *7*(6), 756-779. (project page : https://arxiv.org/abs/2205.09723v2) [4] Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., ... & Han, X. (2022). Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis*, *81*, 102559. [5] Chen, R. J., Chen, C., Li, Y., Chen, T. Y., Trister, A. D., Krishnan, R. G., & Mahmood, F. (2022). Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16144-16155). [6] Chen, R. J., Ding, T., Lu, M. Y., Williamson, D. F., Jaume, G., Song, A. H., ... & Mahmood, F. (2024). Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, *30*(3), 850-862. [7] Lu, M. Y., Chen, B., Williamson, D. F., Chen, R. J., Liang, I., Ding, T., ... & Mahmood, F. (2024). A visual-language foundation model for computational pathology. *Nature Medicine*, *30*(3), 863-874. [8] Yang, Y., Xu, X., Zhou, Y., & Zheng, J. (2025). CellVTA: Enhancing Vision Foundation Models for Accurate Cell Segmentation and Classification. *arXiv preprint arXiv:2504.00784*. [9] Hörst, F., Rempe, M., Heine, L., Seibold, C., Keyl, J., Baldini, G., ... & Kleesiek, J. (2024). Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, *94*, 103143. [10] Ikezogwo, W., Seyfioglu, S., Ghezloo, F., Geva, D., Sheikh Mohammed, F., Anand, P. K., ... & Shapiro, L. (2023). Quilt-1m: One million image-text pairs for histopathology. *Advances in neural information processing systems*, *36*, 37995-38017. [11] Hörst, F., Rempe, M., Becker, H., Heine, L., Keyl, J., & Kleesiek, J. (2025). CellViT++: Energy-Efficient and Adaptive Cell Segmentation and Classification Using Foundation Models. *arXiv preprint arXiv:2501.05269*. [12] Galvez Jiménez, L., Franzin, A., & Decaestecker, C. (2023, September). Training data selection to improve multi-class instance segmentation in digital pathology. In *Proceedings*

	of the 2023 10th International Conference on Bioinformatics Research and Applications (pp. 27-33). [13] Jiménez, L. G., & Decaestecker, C. (2024). Impact of imperfect annotations on CNN training and performance for instance segmentation and classification in digital pathology. *Computers in biology and medicine*, *177*, 108586.
Detailed Work Plan	The team leaders will prepare both a dataset and instantiate the two adaptation methods with one foundational model before the start of the workshop. The first week will therefore be dedicated to casting the other datasets to this common format and instantiate the adaptation methods with all relevant pre-trained models. The second week will be dedicated to the generation of results with a varying number of annotations and training the data quality predictor. Computational resources will be provided by the CECI clusters, namely Manneback and Lyra. Additionally, a dedicated machine with 4 A10 40 Gb GPU can be reserved through the MedReSyst (UCLouvain) project for the workshop duration. From our references, we estimate the highest computational requirement would be about 30 GPU hours on a single A100 80 Gb (available on Manneback) for the largest model and largest dataset. For the other datasets, the computational requirements should be about 1 to 6 GPU hours. We will also need tables, chairs, markers, a whiteboard, power outlet for team member's laptops, and a stable and fast internet connection for access to clusters.
Other Remarks	

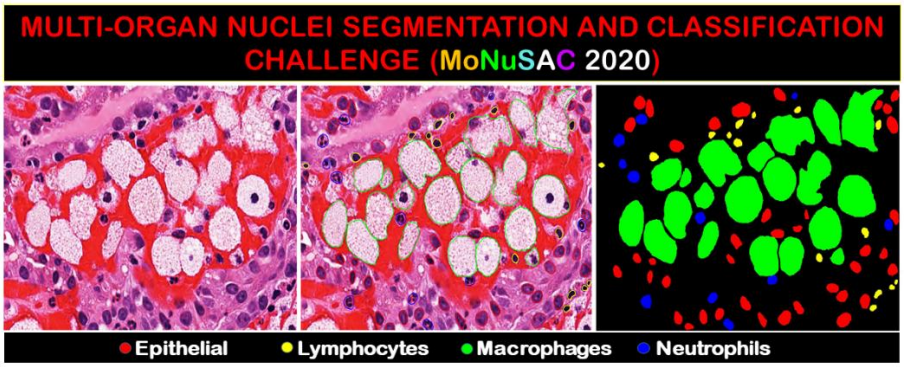

Figure 1 - Example data from MoNuSac taken from their website.

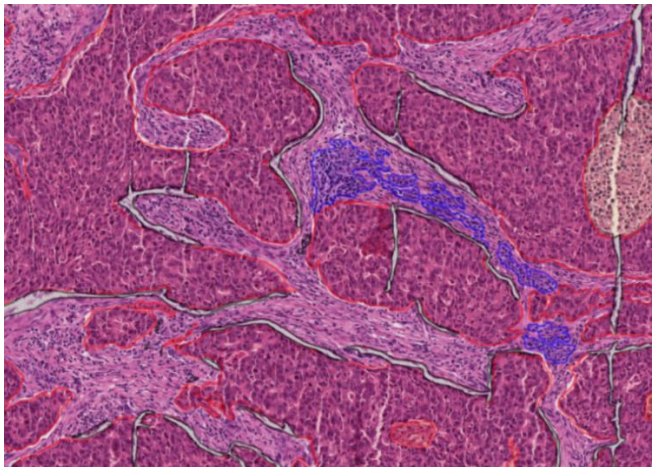

Figure 2 - Example of an annotated patch from BCSS. Note there are no instance-level annotations.