Mueller polarimetric imaging and Deep Learning:
How to handle inaccurate annotations and low data volume ?

M2 Applied Math Internship - Spring/Summer 2026

Centre de Mathématiques Appliquées – École Polytechnique

Erwan Le Pennec (CMAP) / Angelo Pierangelo (LPICM)

Although the development of deep learning has fueled considerable progress in the last decade in the field of computer vision, natural language processing, and reinforcement learning, decision support systems still pose a challenge in healthcare. Indeed, data is often scarce, and there is a need for strong control (guarantees) for the results. This project is focused on a new measurement device using Mueller polarimetric imaging and its application in detecting the risk of premature birth.

The Laboratory of Physics of Interfaces and Thin Films (LPICM, CNRS Ecole Polytechnique) has developed a high performance Mueller Polarimetric Colposcope for the examination of cervical tissue in vivo (https://lpicm.cnrs.fr/research/digital-multispectral-mueller-colposcopy-for-exploring-cervical-microstructure-in-vivo/) [1]. It has been used at the patient's bedside to measure the state of the cervix during pregnancy, in low- and high-risk pregnancies, in the framework of the clinical trial COLPOTERME.

We wish to determine whether polarimetric imaging allows for a better prediction of the risk of premature birth and/or to explain it through the evolution of tissue polarimetry. The specific challenges we aim to address include the low volume of data and the quality of spatial annotations.

Spatial annotations

Preliminary experiments have shown the importance of spatial structure in diagnosis. A first prototype of structure detection has been implemented using a U-Net-type network, providing satisfactory results. However, these results are significantly limited by the difficulty of annotating images. Pixel-by-pixel "labels'' do not seem reliable. One possible approach to mitigate this issue is to modify the labels for training with smoothed labels. The most commonly used technique involves mixing annotators' labels with random labels. This approach does not consider the geometric aspect of the problem and its connection to annotators' uncertainty. Spatial smoothing of labels is an option we consider, as proposed by Islam and Brocker in 2021 [3]. An extension of this model considering image geometry was suggested by Adiga Vasudeva et al. in 2022 [4], but it does not account for annotator-related uncertainty. Therefore, we aim to propose a model that combines image geometry and annotator uncertainty estimation to achieve the most realistic label smoothing.

Low Data Volume

Although in-vivo clinical image collection is ongoing, the final volume of images is expected to be a few hundred at best. Furthermore, these images are specific, as they are polarimetric rather than colorimetric. Conventional techniques using feature extraction via pre-trained networks may not necessarily yield good results. It is likely unrealistic to expect these hundreds of images to be sufficient for training from scratch. Therefore, we are considering data augmentation techniques to enable the extraction of interesting features for prediction. Because deforming acquired images may not be sufficient, we are exploring an approach based on generating artificial data using either physical models, generative models, or a combination of both [5].

The goal of this internship is to advance on those two topics to pave the road for a PhD thesis in Applied Math that is already funded. The intern will be the PhD candidate, if the internship goes well

Candidate

The candidate will be in the second year on an applied math master. She or he will be interested in applying its knowledge to a real world setting, while working on theoretical justifications of their methodology. Ideally, she or he will have studied Deep Learning, but we will value the curiosity and the interest of the candidate.

Team

Erwan LE PENNEC, Professor, CMAP, École polytechnique
erwan.le-pennec@polytechnique.edu

Angelo PIERANGELO, Professor, LPICM, École polytechnique
angelo.pierangelo@polytechnique.edu

Bibliography

1. J. Rehbinder, J. Vizet, J. Park, R. Ossikovski, J.-C. Vanel, et al. "Depolarization imaging for fast and non-invasive monitoring of cervical microstructure remodeling in vivo during pregnancy. Sci. Rep. 12, 12321 (2022)".
2. J. Park "Digital multispectral Mueller colposcopy for exploring cervical microstructure in vivo" (https://theses.fr/2022IPPAX141) (2022)
3. Mobarakol Islam and Ben Glocker. "Spatially Varying Label Smoothing: Capturing Uncertainty from Expert Annotations". In Information Processing in Medical Imaging: 27th International Conference, IPMI 2021 (2021)
4. Sukesh Adiga Vasudeva, Jose Dolz, Herve Lombaert "GeoLS: Geodesic Label Smoothing for Image Segmentation". Medical Imaging with Deep Learning, PMLR 227:468-478, - (2024)
5. N. A. A. Braham, C. M. Albrecht, J. Mairal, J. Chanussot, et al., "SpectralEarth: Training hyperspectral foundation models at scale." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, (2025)