Machine-learning framework aided by stochastic geometry and mathematical morphology for classifying RNA cell features

Six-months internship at Mines Paris, Spring 2023

Context and purpose of the internship:

Gene expression control strongly depends on the cells' RNA abundance and localization. A team of researchers from the CBIO laboratory of Mines Paris and Institut Curie [1] has developed techniques based on image analysis of RNA fluorescence in-situ hybridization (FISH) that quantify key features of cells, such as mRNA spots, sought to play an important role for the spatial control of gene expression at the single cell level.

The dataset, obtained from a large number of experiments, provides information on populations of cells, and may be described as a *points patterns* with various *features* attached to each point. The main purpose of this internship is the development of original methods for describing and classifying cell populations, based on a machine-learning framework. Here we propose to investigate the newly-developed "morphological layers" deep-learning framework [2-3] developed by Santiago Velasco-Forero. In contrast with classical machine-learning pipelines based on convolutional layers, the morpho-layer networks use mathematical morphology transformations. When trained, the method amounts to learn "structural elements" that are at the basis of morphological operators (e.g. dilations, erosions). In doing so, the shape of these structural elements discriminate between different types of point processes and their spatial dispersion, for instance clusters and repulsion effects. From a theoretical point of view, these structural elements can be regarded as "input values" of the so-called Choquet-capacity functional [4].

Rigorous reconstruction theorems from random sets theory state how such functional provides, in theory, complete information on random sets, or in the present case, point processes. Although, in practice, the functional can not generally be computed beyond elementary characteristics (typically covariances) it has long been hypothesized that higher-order statistics could provide robust and accurate descriptors. The internship therefore aims to explore whether one may use machine-learning methods to provide, in practice, relevant shape-descriptors, that are able to describe the spatial structure of random point processes with a good accuracy.

The figure below (from Ref. [1]) shows examples of point patterns that will be the focus of the internship. These patterns will be treated as images on which analysis will be performed.
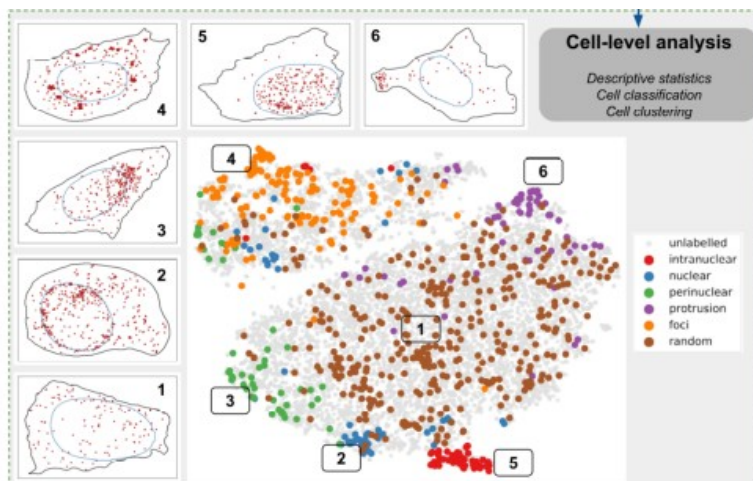
Program:
The internship work will consist in (i) analyzing and assessing cell data provided by CBIO, (ii) generating ideal point structures obtained in exact models (for instance, Poisson, Matern etc.), (iii) defining a machine-learning architecture that uses the morpho-layers library for point patterns, (iv) applying the method to classify point ptterns, first on ideal point structures, second on biological data.

References:
[1] A Imbert, W Ouyang, A Safieddine, E Coleno, C Zimmer, E Bertrand, T Walter, F Mueller. FISH-quant v2: a scalable and modular analysis tool for smFISH image analysis. bioRxiv 2021.07.20.453024; DOI: https://doi.org/10.1101/2021.07.20.453024
[2] S Velasco-Forero, A Rhim, J Angulo. Fixed Point Layers for Geodesic Morphological Operations, British Machine Vision Conference
[3] S Velasco-Forero, R Pagès, J Angulo. Learnable Empirical Mode Decomposition based on Mathematical Morphology SIAM Journal on Imaging Sciences 15 (1), 23-44. See also https://people.cmm.minesparis.psl.eu/users/velasco/morpholayers/intro.html
[4] Molchanov, Ilya, and Ilya S. Molchanov. *Theory of random sets*. Vol. 19. No. 2. London: Springer, 2005.

Duration: 6 months, April to September 2023.

Location: the internship will be carried out at the Centre of Mathematical Morphology (CMM) laboratory of Mines Paris, located in Fontainebleau, south of Paris. Regular meetings will be held in Paris and Fontainebleau. Part-time remote working is possible.

Supervision: the internship will be jointly supervised by François Willot, (CMM) Santiago Velasco-Forero (CMM), Jesus Angulo (CMM) and Thomas Walter (CBIO).

How to apply: we are seeking Master 2 students in machine learning, data science, image analysis or biology with excellent mathematical background and a focus on research. Please send your CV, master scores, exam marks, and any other useful document including university reports, recommendation letters, previous internships, contact of Professors, etc. to francois.willot@minesparis.psl.eu.