

Master's Thesis : Cytomine modules for multispectral data analysis

Auteur : Mathy, Lionel

Promoteur(s) : Maree, Raphael

Faculté : Faculté des Sciences appliquées

Diplôme : Master en science des données, à finalité spécialisée

Année académique : 2020-2021

URI/URL : <http://hdl.handle.net/2268.2/11241>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

MASTER'S THESIS

Cytomine modules for multispectral data
analysis

Author:
Lionel Mathy

Supervisor:
Raphaël Marée

*Thesis performed in order to obtain the grade of master in
data science*

University of Liège
Faculty of Applied Science

Academic year 2020-2021



Abstract

Computer vision has always been a challenging fields. Many success have been obtained with sophisticated techniques. However, the vast majority of the algorithms developed last decades only uses images with 1 channel (black and white image) or 3 channels (colored images). However, the light spectrum is larger than this and all these other frequencies may contain very useful information for some machine learning applications. The goal of this master thesis is to exploit machine learning algorithms applied to multispectral images. Some random forests algorithms have been successfully applied for histology classification. We show that some pre-processing steps are required, but others are not. Also, some hyper-parameters were tested in a grid search way in order to find the optimal values of these parameters. We also show that having a lot of frequencies is good, but for some applications, a good selection of frequencies provided to machine learning algorithm allows to obtain high accuracy, which allows for improving the throughput of the whole process in the future.

Illustration from the master's thesis

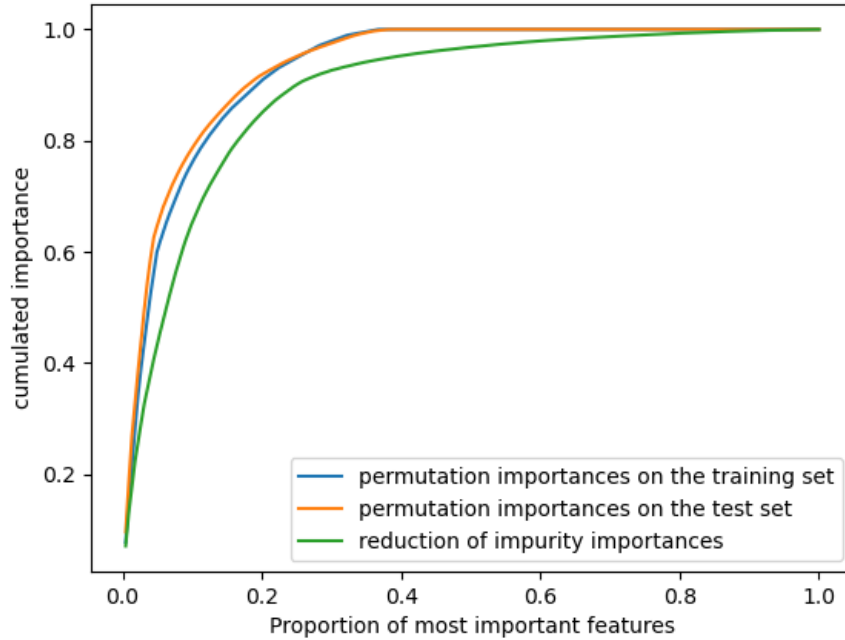


Figure 1: importance of features cumulated according to features importance sorted from highest to lowest

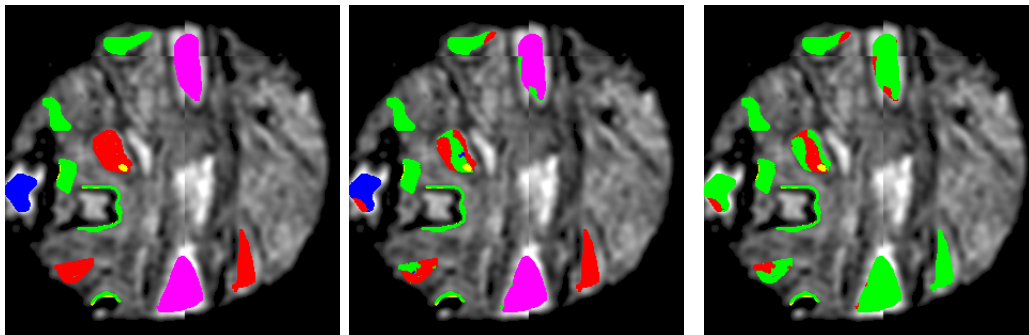


Figure 2: label core = B4. Left image = real classes. Centre image = predicted classes. Right image = accuracy classes. For left and centre image: green = epithelium, red = stroma, purple = blood, blue = necrosis, yellow = pixel rejected by the quality test. For the right image: green = pixel correctly classified, red = pixel wrongly classified