

## Deep Learning for Content-Based Image Retrieval in Biomedical applications

**Auteur :** Schyns, Axelle

**Promoteur(s) :** Maree, Raphael; Geurts, Pierre

**Faculté :** Faculté des Sciences appliquées

**Diplôme :** Master en ingénieur civil en informatique, à finalité spécialisée en "intelligent systems"

**Année académique :** 2022-2023

**URI/URL :** <http://hdl.handle.net/2268.2/17731>

---

### *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.*

---

# Deep Learning for Content-based Image Retrieval in Biomedical Applications

*Author:* Axelle Schyns

*Supervisors:* Pierre Geurts and Raphaël Marée

*Academic Year:* 2022 - 2023

*Place:* University of Liège

*Master in*

Computer Science and Engineering,  
Intelligent Systems

Due to advances in the digital field, the number of images being generated every day grows exponentially. The field of histopathology is no exception and witnesses the emergence of an increasing number of Whole Slide Images that need to be treated, analyzed and diagnosed. One way to facilitate the diagnostic process is by comparing a particular case with other similar cases. This implies, first, the accessibility to other cases, as well as the ability to retrieve the most useful ones, i.e., the most similar cases. To achieve the latter goal, the technique of Content-Based Image Retrieval (CBIR) was conceived. CBIR involves retrieving the most similar images in a database to a given query image.

The goal of this thesis is to study the different elements that compose a CBIR framework and the options available for them, with a specific focus on the feature extraction part of the framework. It offers an open-source implementation that allows the combination of the researched options to create a fully operational CBIR framework. It provides both supervised and self-supervised models as a way to accommodate all situations and datasets.

All feature extraction models are trained on a single dataset containing over 600,000 histopathological images and evaluated on approximately 200,000 different images from the same dataset. Extensive experiments are conducted to analyze the resilience of the frameworks in different situations, such as when dealing with new data or handling class imbalance.

While the supervised models have displayed great results and the self-supervised methods have demonstrated great potential, the scope of what could have been achieved is limited by the lack of evaluation by trained pathologists and by the few remaining untested combinations.