Topological clustering for biological images

Internship Position
Biological Image Analysis Unit, Institut Pasteur, Paris, France

Topological learning. Topological analysis of tissues consists in detecting and characterizing the biological structures constituting the tissue architecture. From a medical point of view, it allows to represent the visual criteria used by physicians to distinguish the grades of a disease.

Since several years, the topological characterization of the image content is a central tool for several applications. Topological descriptors allow to define efficient and, above all, interpretable methods for clustering or classification. Justifying the model prediction by arguments referring to the tissue morphology improves both the understanding of the studied phenomena and the physicians confidence. Finally, a robust detection of the tissue architecture allows the definition of image-based biomarkers to study the evolution of a disease.

Nowadays, a large literature exists about methods taking into account topological constraints and their application to microscopy or histopathology [6, 9, 3, 5]. Most of these works are based on the theory of persistent homology allowing to define images signatures encoding the topological and geometrical structure of the tissue [1, 7, 2, 8, 4].

Project. The goal of the internship is the development of a clustering method to distinguish different tissues based on their topological morphologies. This can be particularly useful, when exploring a new large datasets, to detect different subgroups of tissue architectures. An other application could be the discover of intermediate grades for a given disease.

In particular, we would like to study the role of a topological loss in the training of autoencoders. The main idea is to enrich the images representation in lower dimension, performed by autoencoding, with a constraint on morphological consistency.

This approach should force the algorithm to encode the tissue architecture independently of its acquisition conditions (orientation, artifacts) facilitating a semantic representation of the image. The method will be applied to different datasets of histology, cytology, or microscopy images.

Expected work. A good knowledge in mathematics and deep learning theory is need for this topic. After a review of the literature on topological learning via persistent homology, a new method of topological autoencoding will be developed. The method validation will be performed on different datasets.

Contacts :
Giacomo Nardi, Biological Image Analysis Unit, Institut Pasteur, giacomo.nardi@pasteur.fr.

References

[1] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[2] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[3] Peter Lawson, Andrew Sholl, J. Brown, Brittany Fasy, and Carola Wenk. Persistent homology for the quantitative evaluation of architectural features in prostate cancer histology. *Scientific Reports*, 9:1139, 02 2019.

[4] Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological autoencoders. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7045–7054. PMLR, 13–18 Jul 2020.

[5] Talha Qaiser, Yee Tsang, Daiki Taniyama, Naoya Sakamoto, Kazuaki Nakane, David Epstein, and Nasir Rajpoot. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Medical Image Analysis*, 55, 04 2019.

[6] Lucille Quénéhervé, Grégoire David, Arnaud Bourreille, Jean-Benoit Hardouin, Gabriel Rahmi, N. Neunlist, Jeremy Bregeon, and Emmanuel Coron. Quantitative assessment of mucosal architecture using computer-based analysis of confocal laser endomicroscopy in inflammatory bowel diseases. *Gastrointestinal Endoscopy*, 89, 08 2018.

[7] Fan Wang, Huidong Liu, Dimitris Samaras, and Chao Chen. Topogan: A topology-aware generative adversarial network. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 118–136, Cham, 2020. Springer International Publishing.

[8] Jiaqi Yang, Xiaoling Hu, Chao Chen, and Chialing Tsai. A topological-attention ConvLSTM network and its application to EM images. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 217–228, Cham, 2021. Springer International Publishing.

[9] Daniela Ştefănescu, Costin Streba, Elena Cârţână, Adrian Saftoiu, Gabriel Gruionu, and Lucian Gruionu. Computer aided diagnosis for confocal laser endomicroscopy in advanced colorectal adenocarcinoma. *PLOS ONE*, 11:e0154863, 05 2016.