Nuclei Segmentation and Color Normalization of Histological Images:

Rough-Fuzzy Circular Clustering to Deep Generative Modeling

A thesis submitted to Indian Statistical Institute
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science

by

Suman Mahapatra
Senior Research Fellow

Under the supervision of
Dr. Pradipta Maji, Professor

Machine Intelligence Unit
Indian Statistical Institute, Kolkata

July 2024

Abstract

Histological image analysis deals with microscopic examination of tissue samples, which are stained with multiple histochemical reagents to highlight different cellular structures. The most important characteristic of histological images is the enormous amount of information, which makes computer-aided diagnosis more accurate than other imaging modalities. One of the important tasks in digital pathology is nuclei segmentation as it provides significant morphological information, which helps in therapeutic diagnosis and treatment of cancer. However, nuclei segmentation is a challenging task as nuclei structures can exhibit different morphologies, color, texture or can be occluded partially by other cellular components. Segmentation becomes more difficult when histological images suffer from inadmissible inter and intra-specimen disagreement in the appearance of stained tissue color due to inconsistency in staining routine, orientation of lens aperture, and so on. So, color normalization without hampering histological information is important for accurate nuclei segmentation. One of the main problems associated with stain color normalization is uncertainty due to incompleteness and vagueness in stain class definition, and overlapping characteristics of stains.

In this regard, the thesis introduces the concept of rough-fuzzy circular clustering for stain color normalization. It judiciously integrates the merits of both fuzzy and rough sets. While the theory of rough sets deals with uncertainty due to vagueness and incompleteness in stain class definition, fuzzy set handles overlapping nature of histochemical stains. The proposed rough-fuzzy circular clustering works on a weighted hue histogram, which considers both saturation and local neighborhood information of an image. A new dissimilarity measure is introduced to deal with the circular nature of the hue values. Being a generalization of existing algorithms, rough-fuzzy circular clustering facilitates accurate color normalization of histological images, while its integration with von Mises distribution provides better modeling of circular data and enables proper analysis of stained histological images.

In order to normalize the color without hampering the histological information of the image, a new deep generative model is introduced to capture the disentangled color appearance and stain bound information. To deal with the overlapping nature of histochemical stains, the proposed model assumes that the latent color appearance code is sampled from a mixture of truncated normal distributions. To segment cell nuclei from histological images, a new deep generative model is introduced, which considers an embedding space for handling information-asymmetry between histological image space and segmentation map. Judiciously integrating the concepts of optimal transport and measure theory, the model develops an invertible generator, which provides an efficient optimization framework with lower network complexity. Since segmentation and color normalization are two intertwined procedures, a novel simultaneous

segmentation and color normalization model is finally introduced, integrating the merits of spatial attention and truncated normal distribution. The latent color appearance information is assumed to be pairwise independent of nuclei segmentation map and other tissue-level details.