H2-MIL: Exploring Hierarchical Representation with Heterogeneous Multiple Instance Learning for Whole Slide Image Analysis

Abstract

Current representation learning methods for wholeslide image (WSI) with pyramidal resolutions are inherently homogeneous and flat, which cannot fully exploit the multi-scale and heterogeneous diagnostic information of different structures for comprehensive analysis. This paper presents a novel graph neural networkbased multiple instance learning framework (i.e., H2MIL) to learn hierarchical representation from a heterogeneous graph with different resolutions for WSI analysis. A heterogeneous graph with the "resolution" attribute is constructed to explicitly model the feature and spatial-scaling relationship of multi-resolution patches. We then design a novel resolution-aware attention convolution (RAConv) block to learn compact yet discriminative representation from the graph, which tackles the heterogeneity of node neighbors with different resolutions and yields more reliable message passing. More importantly, to explore the task-related structured information of WSI pyramid, we elaborately design a novel iterative hierarchical pooling (IHPool) module to progressively aggregate the heterogeneous graph based on scaling relationships of different nodes. We evaluated our method on two public WSI datasets from the TCGA project, i.e., esophageal cancer and kidney cancer. Experimental results show that our method clearly outperforms the state-of-the-art methods on both tumor typing and staging tasks. The code will be publicly available.