A Joint Framework for Scalable Retrieval and Detection of Mammographic Masses

Menglin Jiang
Rutgers University
12/10/2015 at 03:30 pm
CBIM 22

Abstract
For years, mammography has served as the gold standard for diagnosis of breast cancer, which is the second leading cause of cancer-related death among women. Nevertheless, as a major indicator of breast cancer, mammographic masses are very difficult to diagnose due to their large variance in shape, margin, size and their obscure boundaries. To facilitate the interpretation of mammographic masses, a vast number of computer-aided diagnosis (CAD) methods have been presented during the past half century. Most of them are based on either machine learning or content-based image retrieval (CBIR) techniques. However, either category has its limitations.

In this talk, we introduce a joint framework to solve mammographic mass retrieval and detection together. Our framework integrates scalable image retrieval and discriminative learning. Specifically, a large number of previously diagnosed masses with annotated boundaries form a training set, and a simple classifier is learned for each of them to distinguish true masses from similar regions. Given a query mammogram, it is matched with the training masses through a local bag-of-words (BoW) model, which simultaneously localizes suspicious regions in the query and retrieves their similar training masses. Then, using the classifiers corresponding to retrieved masses, the query masses, if any, are detected from the suspicious regions. Compared with learning-based methods, our approach could provide radiologists with not only a detection result, but also similar diagnosed cases. Radiologists could use the retrieved cases as decision support and make a correct diagnosis for the query mammogram. Compared with traditional CBIR-based methods, our approach doesn’t need radiologists to label suspicious regions in the query mammogram. Therefore, it could serve as a fully automated double reading aid. Furthermore, the proposed approach is computationally efficient. In fact, it’s among the first few attempts to tackle the large-scale medical image analysis problem.