A Quantitative Analysis Platform for PD-L1 Immunohistochemistry Based on Point-Level Supervision Model

Haibo Mi¹, Kele Xu¹, Yang Xiang¹, Yulin He¹, Dawei Feng¹, Huaimin Wang¹, Chun Wu², Yanming Song² and Xiaolei Sun²

¹School of Computer, National University of Defense Technology, Changsha, China
²Abcarta Inc., Suzhou, China
haibo_mihb@126.com

Abstract

Recently, deep learning has witnessed dramatic progress in the medical image analysis field. In the precise treatment of cancer immunotherapy, the quantitative analysis of PD-L1 immunohistochemistry is of great importance. It is quite common that pathologists manually quantify the cell nuclei. This process is very time-consuming and error-prone. In this paper, we describe the development of a platform for PD-L1 pathological image quantitative analysis using deep learning approaches. As point-level annotations can provide a rough estimate of the object locations and classifications, this platform adopts a point-level supervision model to classify, localize, and count the PD-L1 cells nuclei. Presently, this platform has achieved an accurate quantitative analysis of PD-L1 for two types of carcinoma, and it is deployed in one of the first-class hospitals in China.

1 Introduction

Programmed death-ligand 1 (PD-L1) is known as a kind of proteins. The quantitative analysis of PD-L1 immunohistochemistry plays a very important role in the precise treatment of tumor therapy [Brahmer et al., 2012; Reck et al., 2016]. By quantitative analysis of the expression of PD-L1 in the tumor tissue of patients, it is beneficial to design a personalized medication strategy. The quantitative analysis of PD-L1 requires classifying the cell nuclei into the positive tumor nuclei and negative tumor ones, as well as needing to accurately calculate the proportion of positive and negative ones respectively. Currently, in the process of quantitative analysis, pathologist (even in the first-class Hospital) manually calculate the cell nucleus. However, the sizes of whole slide PD-L1 images ranges from several hundred megabytes to more than one gigabyte, in which the number of cell nuclei ranges from several thousands to even millions. Manual analysis not only takes a lot of time and effort, but is also extremely error-prone. Therefore, it is desirable to design an automatic and accurate quantitative analysis tool.

The quantitative analysis of PD-L1 does not require predicting the exact shape of the cell nucleus. As point-level annotations for the cell nucleus is relatively easy for pathologist to mark, and also can provide a rough estimate of the object locations, this kind of annotations is utilized to localize, classify and count the cell nuclei. Point supervision is widely adopted to object counting [Bearman et al., 2016; Laradji et al., 2018]. The solutions for object counting analysis can be classified into three categories: detection-based methods, regression-based methods, and density estimation-based methods [Loy et al., 2013]. Although the regression-based methods, and density estimation-based methods are widely-used to predict the number of the detected objects, they cannot estimate the locations of the objects [Xue et al., 2016]. On the contrary, detection-based methods [Arteta et al., 2016] can be refined to precisely count the cell nuclei with their locations and classifications. In our platform, we choose the detection-based methods as the baseline.

With the point-level annotations of the cell nuclei, we adopt a fully convolutional networks (FCN) [Long et al., 2015] to quantitatively classify, localize, and count the PD-L1 cell nuclei. To reduce the counting error caused by cell nuclear stacking and deformation, HOG (histogram oriented gradients) [Dalal and Triggs, 2005] and NMS (Non-maximum suppression) [Hosang et al., 2017] are also applied to extract low-level features and filter negative positives. Using the aforementioned algorithms, we implement a platform for pathologist to automatically and quantitatively analyze the cell nuclei.

Figure 1: The PD-L1 point-level annotation tool. The red dots represent the positive nuclei (i.e., tumor nuclei) and the yellow dots represent the negative nuclei (i.e., normal nuclei).
of PD-L1. Currently, this platform is used by pathologist in one of the first-class hospitals in China.

2 Implementation

2.1 The PD-L1 Annotation Tool

We have developed a PD-L1 annotation tool for pathologists, as shown in Figure 1. With this tool, the pathologist can quickly classify and locate the cell nuclei. The red dots represent the positive nuclei (i.e., tumor nuclei) and the yellow dots represent the negative nuclei (i.e., normal nuclei). In order to be able to make the analysis more accurate, we recommend that the pathologists mark the nuclei in the middle of its shapes. Empirically, it takes several minutes to finish a size of 1024 × 1204 images. The efficiency of the point-level samples collection is greatly improved compared to that of the mask-level samples collection.

2.2 Overview of the Platform

As shown in Figure 2, the platform mainly includes three parts: a front-end interactive UI, an image processing engine, and detecting agents.

- The interactive UI is primarily used to show the results of quantitative analysis, including the location, number, and proportion of negative and positive cell nuclei. Pathologists can directly browse the full-scale PD-L1 profiles of quantitative analysis results through the interactive UI.
- The image processing engine is mainly used to split the whole slide, distribute the divided sub-pictures, and merge the detecting results. In order to improve the efficiency of the inference, the processing engine splits the original PD-L1 image into several sub-pictures and uses the map-reduce computing framework to send those sub-pictures to multiple agents for the detection. Finally, the detecting results on different agents are combined together and sent to the interactive UI.
- The detecting agent contains a quantitative analysis model based on the FCN network [Long et al., 2015]. The backbone architecture uses a ResNet101 [He et al., 2016] network, and a dilation convolution kernel [Yu and Koltun, 2015] is also utilized to enhance the receptive field. We define an image-level loss to classify the nuclei into the negative ones and positive ones. Since the degree of nuclear staining, extrusion, and stacking affects the accuracy of counting and localizing, we use the HOG algorithm to calculate the nuclear split loss. Further, in the phase of the model inference, we adopt the NMS algorithm to filter the predicted point-level annotations to ensure that each cell nucleus contains only one annotation.

As the annotation tools provide an easy way for the pathologist to annotate pictures, thousands of PD-L1 pictures are collected for two kinds of carcinoma (i.e., lung squamous carcinoma and ductal breast carcinoma) in a short time. In the choice of detection network structure, we compare the performance between DUNet [Jin et al., 2018], DetNet [Li et al., 2018], DeepLab-v3 [Chen et al., 2017] and other models in detail. In our experiments, we find that the FCN network (with a ResNet101 as the backbone architecture) is more suitable for our scenario as the model has the highest accuracy and the lowest false positive rate.

3 Conclusion and Future Work

Deep learning-based quantitative analysis of PD-L1 immunohistochemistry can not only greatly improve the diagnostic efficiency of pathologists, but also greatly improve the diagnostic accuracy. In this paper, we introduced a PD-L1 quantitative analysis platform based on the point supervision model. PD-L1 quantitative analysis is a good application of AI-based technology in the field of cancer therapy. Currently, we have achieved an accurate quantitative analysis of PD-L1 for two types of carcinoma. In the follow-up work, we will continue to expand the scopes of the platform among the hospitals.

Acknowledgments

This work was supported by the National Grand R&D Plan (Grant No. 2016YFB1000101).
References


