様式８

論　文　内　容　要　旨

報告番号	甲　先　第　419　号	氏　名	王　陸
学位論文題目	Research and Practice on Visual Attention Module Design for Saliency Detection		

内容要旨

Visual attention is an important mechanism in the human visual system. When human observe images and videos, they usually do not describe all the content in them. Instead, they tend to talk about the semantically important regions and objects in the image. The human eye is usually attracted by some regions of interest rather than the entire scene. These regions of interest that can present mainly meaningful or semantic content are called saliency region. Visual saliency detection refers to the use of intelligent algorithms to simulate human visual attention mechanism, extract both the low-level features and high-level semantic information and localize the salient object regions in images and videos. The generated saliency map indicates the regions that are likely to attract human attention.

As a fundamental problem of image processing and computer vision, visual saliency detection algorithms have been extensively studied by researchers to solve practical tasks, such as image and video compression, image retargeting, object detection, etc. The visual attention mechanism adopted by saliency detection in general are divided into two categories, namely the bottom-up model and top-down model. The bottom-up attention algorithm focuses on utilizing the low-level visual features such as colour and edges to locate the salient objects. While the top-down attention utilizes the supervised learning to detect saliency. In recent years, more and more research tends to design deep neural networks with attention mechanisms to improve the accuracy of saliency detection. The design of deep attention neural network is inspired by human visual attention. The main goal is to enable the network automatically to capture the information that is critical to the target tasks and suppress irrelevant information, shift the attention from focusing

on all to local. Currently various domain' s attention has been developed for saliency detection and semantic segmentation, such as the spatial attention module in convolution network, it generates a spatial attention map by utilizing the inter-spatial relationship of features; the channel attention module produces a attention by exploring the inter-channel relationship of features. All these well-designed attentions have been proven to be effective in improving saliency detection.

This paper studies on the visual attention mechanism for saliency object detection and applies it to digital histopathology image analytics for breast cancer metastases detection and classification. As shown in following contents, the main research contents include three parts:

First, we propose a semantic attention mechanism to accurately localize the salient objects in complex scenarios. The proposed semantic attention uses Faster-RCNN to capture high-level deep features and replaces the last layer of Faster-RCNN by a FC layer and sigmoid function for visual saliency detection; it calculates proposals' attention probabilities by comparing their feature distances with the possible salient object. The proposed method introduces a re-weighting mechanism to reduce the influence of the complexity background, and a proposal selection mechanism to remove the background noise to obtain objects with accurate shape and contour. The simulation result shows that the semantic attention mechanism is robust to images with complex background due to the consideration of high-level object concept, the algorithm outperforms 13 saliency detection methods.

Second, we designed a deep segmentation network (DSNet) for saliency object prediction. In the network, Pyramidal Attentional ASPP (PA-ASPP) module is explored to provide pixel level attention. DSNet extracts multi-level features with dilated ResNet-101 and the multiscale contextual information was locally weighted with the proposed PA-ASPP. The pyramid feature aggregation encodes the multi-level features from three different scales. This feature fusion incorporates neighboring scales of context features more precisely to produce better pixel-level attention. Finally, we use a scale-aware selection (SAS) module to locally weight multi-scale contextual features, capture important contexts of

ASPP for the accurate and consistent dense prediction. The simulation results demonstrated that the proposed PA-ASPP is effective and can generate more coherent results. Besides, with the SAS, the model can adaptively capture the regions with different scales effectively.

At last, we verified the deep attention mechanism in digital pathology image analytics. A novel Deep Regional Metastases Segmentation (DRMS) framework for breast cancer metastases detection and classification has been proposed. As we know, the digitalized whole slide image has high-resolution, however the size of lesions is often relatively small, and most of the slide region are normal. The highly trained pathologists usually localize the regions of interest first in the whole slide, then perform precise examination in the selected regions. Even though the process is time-consuming and prone to miss diagnosis. The visual attention mechanism should be one of the efficient ways to solve the problem. With the proposed framework, we first utilize the proposed DSNet to detect the regional metastases in patch-level. Then, adopt the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to predict the whole metastases from individual slides. Finally, determine patient-level pN-stages by aggregating each individual slide-level prediction. In combination with the above techniques, the framework can make better use of the multi-grained information in histological lymph node section of whole-slice images. Experiments on large-scale clinical datasets demonstrate that our method delivers advanced performance and provides consistent and accurate metastasis detection.