Attention Based Multi-Instance Thyroid Cytopathological Diagnosis with Multi-Scale Feature Fusion

Shuhao Qiu[1], Yao Guo[1], Chuang Zhu[1] *, Wenli Zhou[1], Huang Chen[2]
[1] Beijing University of Posts and Telecommunications, Beijing, China
[2] China-Japan Friendship Hospital Department of Pathology, Beijing, China

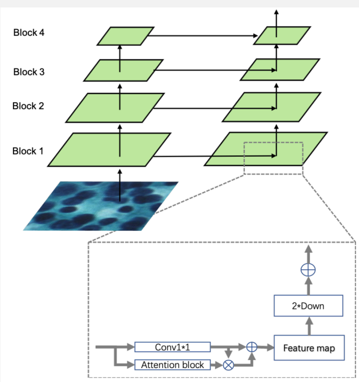
ABSTRACT

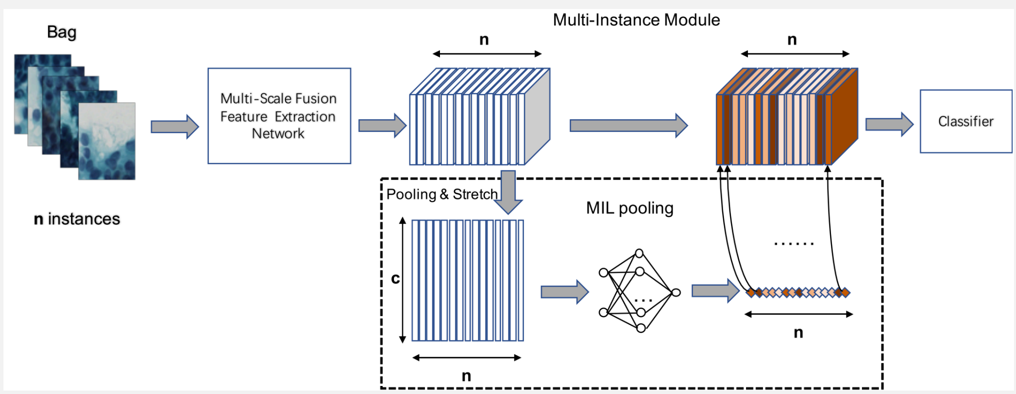
In recent years, deep learning has been popular in combining with cytopathology diagnosis. Using the whole slide images (WSI) scanned by electronic scanners at clinics, researchers have developed many algorithms to classify the slide. However, the key area that support the diagnosis result can be relatively small in a thyroid WSI, and only the global label can be acquired, which make the direct use of the strongly supervised learning framework infeasible. What's more, because the clinical diagnosis of the thyroid cells requires the use of visual features in different scales, a generic feature extraction way may not achieve good performance. In this paper, we propose a weakly supervised multi-instance learning framework based on attention mechanism with multi-scale feature fusion (MSF) using convolutional neural network (CNN) for thyroid cytopathological diagnosis. We take each WSI as a bag, each bag contains multiple instances which are the different regions of the WSI, our framework is trained to learn the key area automatically and make the classification. We also propose a feature fusion structure, merge the low-level features into the final feature map and add an instance-level attention module in it, which improves the classification accuracy.

METHODOLOGY

- Multi-Scale Feature Fusion: We propose a MSF module with attention mechanism to merge the feature maps of different scales together. We use ResNet-18 and ResNet-34 as the backbone of the feature extraction network. Both of these two networks are divided into 4 convolution blocks, the 4 output layers of the blocks are used in the MSF structure. Each layer is transformed by a lateral connection module proposed in FPN before merging. In this paper, we add an attention block into the lateral connection to weigh different level of feature map, for the importance of different scale of features are different among instances. What's more, a residual connection is used in the structure, which will make the training converges faster and avoid gradient vanishing. Finally, the output layer of block 4 will be fed into the multi-instance module for the follow-up classification.

- Multi-Instance Framework: MIL is used to solve the problem when only a bag label is available, which matches the scene of Thyroid diagnosis. For a malignant WSI, there maybe only several patches indicate the malignancy, which brings difficulties for the model to focus on the useful patches. In this paper, we use attention based MIL pooling to weigh the importance of different instances. Specially, we use fully connected layer structure to implement the attention mechanism which is widely used. The instances in a bag will be pooled and stretched into N C-dimensional vectors, then sent into two fully connected layer and reshaped into N × 1 as N weights of instances. . During this process, the model will learn automatically to give the important instance a higher weight, which leads to a better classification result.





Experiment Results and DISCUSSION

Our experiments were conducted on two datasets. We conduct 6 group of experiments on the thyroid dataset by adding additional architecture into two backbone networks (ResNet-18 and ResNet-34). It can be seen that the network using ResNet-18 performs better than the network using ResNet-34 overall, which may because of the overfitting caused by the deeper network. After adding MSF structure to the ResNet-18 network, the accuracy decreased 1.6%, which indicates adding different scale of feature map directly together may not help for all situations. By adding the instance-level attention mechanism before the merging, the accuracy increased by 5.1%, which proves that weighing different scale of feature is necessary for feature fusion in image classification problem.

Experiment Results on Thyroid Dataset Compared with Other Methods

	Acc	Recall	Pre	F1
WELDON+ResNet-18 [14]	0.839	0.807	0.823	0.815
WELDON+ResNet-34 [14]	0.879	0.871	0.862	0.864
mi-net [3]	0.831	0.742	0.852	0.793
MI-net [3]	0.844	0.841	0.841	0.842
MI-Net+DS [3]	0.886	0.868	0.877	0.87
MI-Net+RC [3]	0.859	0.774	0.889	0.828
MIMS [9]	0.873	0.774	0.923	0.842
ResNet-18 + MSFwA (Ours)	**0.932**	**0.910**	**0.935**	**0.921**

Experiment Results on Thyroid Dataset

	Acc	Recall	Pre	F1
ResNet-18 Baseline	0.881	0.890	0.860	0.869
ResNet-18 + MSF	0.865	0.780	0.910	0.828
ResNet-18 + MSFwA	**0.932**	**0.910**	**0.935**	**0.921**
ResNet-34 Baseline	0.842	0.710	0.930	0.781
ResNet-34 + MSF	0.896	0.852	0.914	0.879
ResNet-34 + MSFwA	0.851	0.697	0.952	0.800

Experiment Results on Breast Cancer Dataset

	Acc	Recall	Pre	F1
WELDON+ResNet-18 [14]	0.696	0.808	0.612	0.696
WELDON+ResNet-34 [14]	0.746	0.883	0.651	0.750
mi-net [3]	0.714	0.875	0.618	0.724
MI-net [3]	0.721	0.925	0.623	0.743
MI-Net+DS [3]	0.739	0.891	0.661	0.752
MI-Net+RC [3]	0.732	0.875	0.636	0.737
MIMS [9]	0.750	0.792	0.679	0.731
ResNet-18 Baseline (Ours)	0.772	0.975	0.658	0.786
ResNet-18 + MSF (Ours)	0.768	0.883	0.687	0.768
ResNet-18 + MSFwA (Ours)	**0.796**	**0.967**	**0.689**	**0.804**

We also use our method to make comparisons with other state-of-the-art multi-instance methods on both thyroid dataset and a public breast cancer dataset. It can be seen that, our method achieves better result than others.