Dual Space Multiple Instance Representative Learning for Medical Image Classification

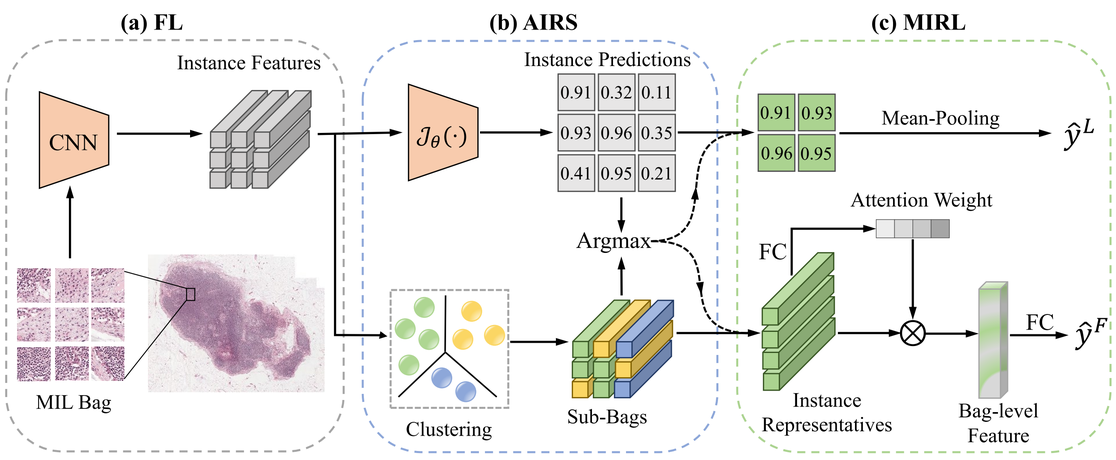
Xiaoxian Zhang, Sheng Huang✉, Yi Zhang,
Mingchen Gao, Chen Liu, Xiaohong Zhang

Abstract

Medical image classification plays a vital role in AI-aided medical diagnosis, which is usually addressed as a Multiple Instance Learning (MIL) issue (i.e., each sample is a bag of instances). The disease or discriminative area for medical images is usually smaller than the whole tissue. In other words, most of the instances in a bag are irrelevant and could interfere with the bag label inference. To address this issue, we add an instance representative selection process before MIL and propose a novel MIL approach named Dual Space Multiple Instance Representative Learning (DSMIRL). Extensive experiments on five standard MIL benchmarks and two medical image datasets demonstrate the promising performance of DSMIRL over the state-of-the-art MIL approaches.

Overview

This paper proposes Dual Space Multiple Instance Representative Learning (DSMIRL) to solve medical image classification. DSMIRL consists of three modules, namely Feature Learning (FL), Adaptive Instance Representative Selection (AIRS), and Multiple Instance Representative Learning (MIRL).



In this approach, we first employ AIRS to filter out irrelevant instances based on the instance features extracted by FL. Then we use MIRL to conduct the instance aggregation in both feature and label spaces to exploit their complementary information.

Methodology

Adaptive Instance Representative Selection

We treat image patches from the same patient as a bag B. We first divide each bag into multiple clusters by clustering methods $M = \pi\left(\{f_i\}_{i=1}^n\right), f_i \in B$, where $\pi(\cdot)$ can be any clustering method. Each cluster can be retrieved based on the indicator matrix M,

$$b_k \leftarrow \Phi(B, M_{:k}),\ \text{s.t.} \bigcup_{k=1}^{K} b_k = B \text{ and } \bigcap_{k=1}^{K} b_k = \oslash. \tag{1}$$

Then, each cluster is considered a sub-bag and scored by the maximum value of instance scores obtained by the instance label prediction network $\mathcal{J}_\theta(\cdot)$,

$$p_k = \max(\{\hat{y}_i\}_{f_i \in b_k}) = \max(\{\mathcal{J}_\theta(f_i)\}_{f_i \in b_k}). \tag{2}$$

Finally, we select the instances of the sub-bag with the highest score as the instance representatives of the corresponding bag.

Multiple Instance Representative Learning

• Aggregation in Feature Space

After obtaining the instance representative embedding, we employ attention-based MIL pooling to aggregate instance features. The features of instance representatives are weighted and summed to produce a final bag-level feature: $\hat{f} = \sum_{f_i \in b_{\hat{k}}} a_i f_i$. Then we introduce a classifier to infer the label with this bag-level feature $\hat{\mathcal{Y}}^F = \mathcal{Q}_\psi(\hat{f})$.

• Aggregation in Label Space

With regard to the instance aggregation in label space, we directly conduct the mean-pooling on the label predictions of instance representatives for achieving another bag-level label prediction,

$$\hat{\mathcal{Y}}^L = \frac{1}{|b_{\hat{k}}|} \sum_{x_i \in b_{\hat{k}}} \hat{y}_i. \tag{3}$$

We combine the above two pieces of information to optimize the model parameters. We only adopt the instance feature space information in the inference stage to infer the bag label.

Experiments

We conduct extensive experiments on five MIL benchmarks and two medical image datasets.

• MIL benchmarks

The experimental results are taken from the original paper.

Methods	Datasets				
	Musk1	Musk2	Fox	Tiger	Elephant
mi-SVM	.874±N/A	.836±N/A	.582±N/A	.784±N/A	.822±N/A
mi-Graph	.889±.03	.903±.04	.620±.04	.860±.04	.869±.04
MI-Kernel	.880±.03	.893±.02	.603±.03	.842±.01	.843±.02
mi-Net	.889±.04	.858±.05	.613±.04	.824±.03	.858±.04
AttMIL	.892±.04	.858±.05	.615±.04	.839±.02	.868±.02
DSMIL	.932±.02	.930±.02	.729±.02	.869±.01	.925±.01
DSMIRL(ours)	**.966±.05**	**.960±.04**	**.785±.07**	**.921±.07**	**.935±.05**

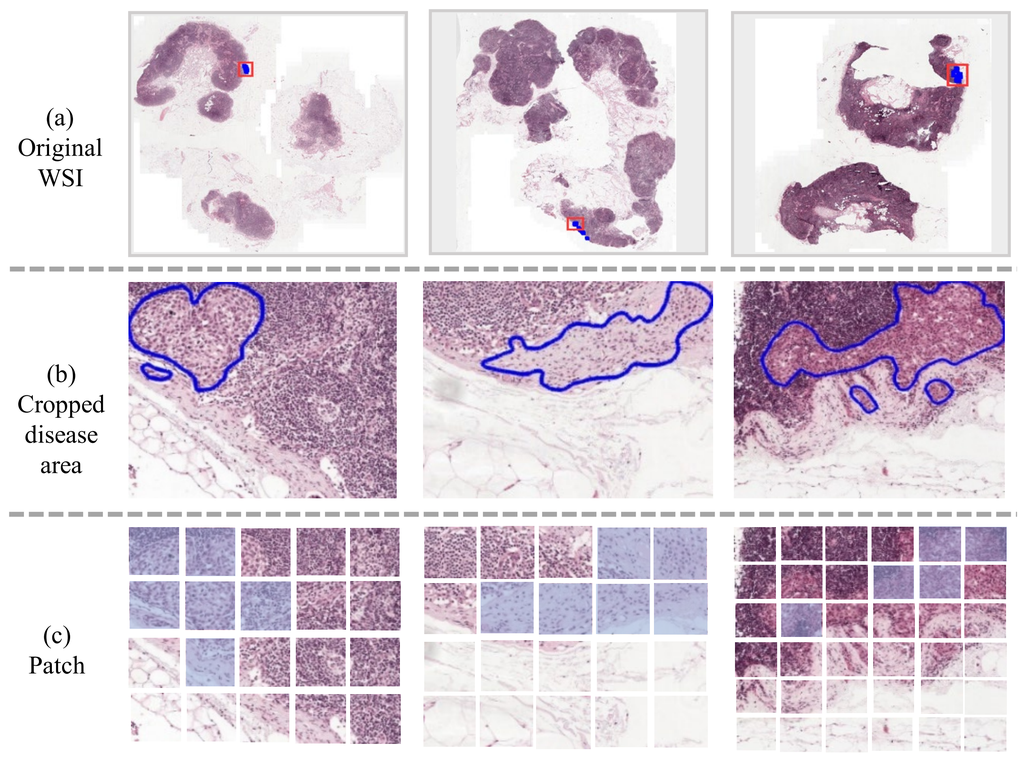
• Camelyon16 and Pneumonia CT datasets

The experimental results are reproduced through the code provided by the author.

Methods	Camelyon16			Pneumonia CT		
	Accuracy	AUC	F1-score	Accuracy	AUC	F1-score
Max-pooling	.864±.02	.920±.03	.821±.03	.835±.05	.895±.05	.834±.05
Mean-pooling	.859±.03	.917±.03	.836±.03	.849±.01	.903±.01	.851±.01
AttMIL	.862±.02	.937±.01	.839±.01	.897±.02	.957±.01	.895±.01
DSMIL	.862±.02	.930±.01	.839±.02	.911±.01	.956±.01	.907±.01
CLAM-SB	.869±.03	.936±.02	.819±.04	.903±.01	.958±.01	.900±.01
CLAM-MB	.852±.04	.934±.01	.807±.06	.885±.02	.947±.02	.886±.02
TransMIL	.857±.03	.945±.02	.800±.06	.866±.05	.943±.02	.876±.05
DSMIRL(ours)	**.889±.01**	**.953±.01**	**.866±.02**	**.930±.01**	**.967±.01**	**.930±.01**

Qualitative Results

We conduct the visualization experiments on the Camelyon16 dataset. In Figure(c), the patches covered in light blue are the selected instance representatives.



Conclusion

- We present a novel MIL approach (DSMIRL) for medical image classification, introducing an instance representative selection before MIL;
- We elaborate a simple but effective instance representative method (AIRS) to solve redundant information;
- We conduct a dual space aggregation strategy (MIRL) to fully exploit the complementary information of the feature and label space.