Comments for Reviewer 4Xqy

Qiehe Sun

March 18, 2024

1 Comments

Thank you for recognizing our work and for your valuable suggestions. And we quite agree with you, our point-to-point response is as follows:

- **Writing improvement.** In response to your suggestion, we have described Fig. 1 in more detail as follows. Furthermore, exhaustive presentations of the sources and dimensions of embeddings, such as $\boldsymbol{u}$ and $\boldsymbol{v}$, will be provided. For example, ”where $\boldsymbol{u} = ||_{\boldsymbol{x}_k \in \boldsymbol{X}'} f_\theta(\boldsymbol{x}_k) \in \mathbb{R}^{2K \times d}$ denotes the bag, obtained by concatenating instance embeddings, with d being the embedding dimension”, ”$\boldsymbol{v} \in \mathbb{R}^{1 \times d}$ is a learnable embedding that represents a virtual instance used for classification” and ”Then the aggregated $\boldsymbol{v}'$ is extracted from the spatial dimension of $\boldsymbol{u}' \in \mathbb{R}^{(1+2K)\times d}$”. These changes will be updated in the final version.

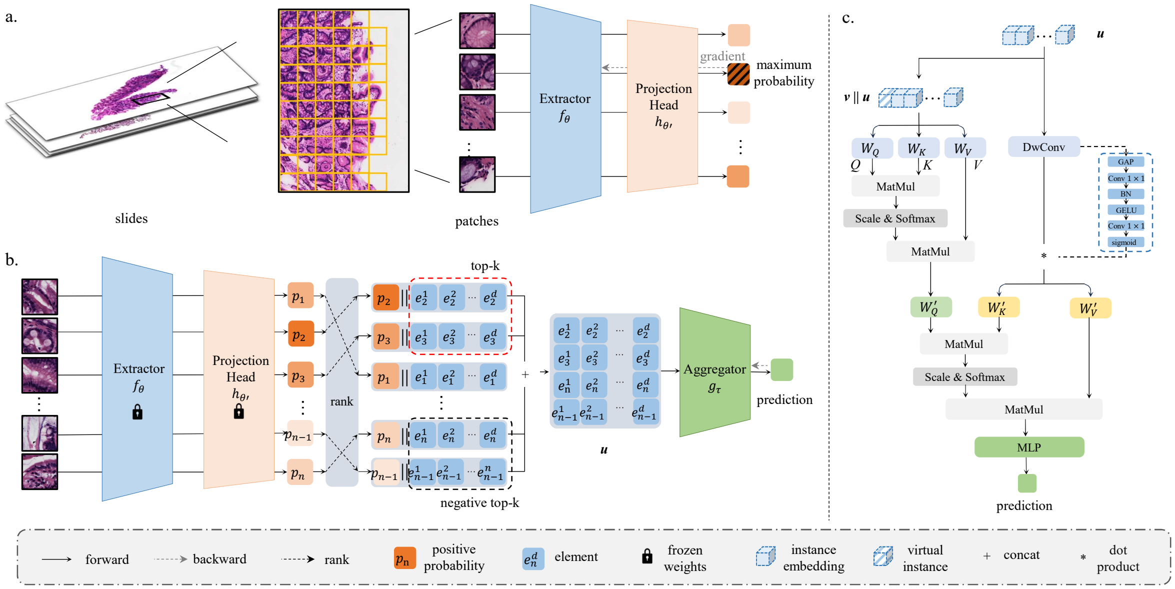

Figure 1: Overview of NcIEMIL. (a) Instances are fed into the extractor and projection head on a bag basis, and the instance with the highest probability is involved as a representative of its corresponding bag. (b) Instances within a bag are re-ranked, and bi-directional sampling is employed to generate bag embeddings. (c) A hybrid attention-based aggregator, comprising parallel spatial and channel attention mechanisms, is employed to mitigate information redundancy.

- **Additional comparison experiments.** Following your suggestions, we have incorporated two additional methods into our comparative experiment. The naive image-wise approach involves directly classifying downsampled WSI thumbnails using a classification network, akin to the approach in [CCY+21]. The fully-supervised approach, on the other hand, uses a batch of patches with real labels to train a classification network. In CAMELYON16, we extracted positive and negative patches from within and without cancerous regions respectively, according to the pixel-level annotations provided officially. Regarding BgIM, we randomly selected 2,000 preprocessed patches and annotated them for fully supervised training. The results of the slides were obtained by the mean value of all instances. The results are presented in the following table and will be incorporated into Tab. 1 of the original paper. To be fair, swin-tiny was chosen for all the above classification networks to maintain consistency.

Method	CAMELYON16			BgIM		
	ACC	AUC	F1-Score	ACC	AUC	F1-Score
Naive	$62.40_{1.42}$	$62.16_{1.40}$	$45.15_{5.28}$	$52.98_{3.32}$	$76.58_{1.28}$	$44.27_{4.44}$
Fully-supervised	$91.32_{1.99}$	$94.93_{0.87}$	$90.73_{2.12}$	$91.91_{1.17}$	$98.72_{0.33}$	$91.57_{2.61}$
NcIEMIL	$\mathbf{86.05}_{\mathbf{1.55}}$	$\mathbf{89.68}_{\mathbf{2.10}}$	$\mathbf{85.26}_{\mathbf{1.54}}$	$\mathbf{85.23}_{\mathbf{0.95}}$	$\mathbf{95.87}_{\mathbf{0.60}}$	$\mathbf{81.20}_{\mathbf{0.94}}$

- **Additional ablation experiments.** We also performed ablation experiments on K. The results are shown in the following table. Specifically, we introduced two values for K: a small K ($K = 128$ for CAMELYON16 and $K = 32$ for BgIM) and a medium K ($K = 288$ for CAMELYON16 and $K = 72$ for BgIM), and assessed their impacts on the results. The reason for not selecting a value larger than K in the original article ($K = 512$ for CAMELYON16 and $K = 128$ for BgIM) is because of the restricted number of instances in the smallest bag within the dataset.

Ablation item	CAMELYON16			BgIM		
	ACC	AUC	F1-score	ACC	AUC	F1-score
w/ small K	$85.75_{1.84}$	$\mathbf{90.06}_{\mathbf{1.87}}$	$84.87_{1.68}$	$83.33_{1.51}$	$95.21_{0.50}$	$80.09_{2.14}$
w/ medium K	$85.36_{1.49}$	$89.46_{1.30}$	$84.01_{1.30}$	$84.28_{1.90}$	$94.90_{0.83}$	$80.00_{1.95}$
NcIEMIL	$\mathbf{86.05}_{\mathbf{1.55}}$	$89.68_{2.10}$	$\mathbf{85.26}_{\mathbf{1.54}}$	$\mathbf{85.23}_{\mathbf{0.95}}$	$\mathbf{95.87}_{\mathbf{0.60}}$	$\mathbf{81.20}_{\mathbf{0.94}}$

References

[CCY+21] Chi-Long Chen, Chi-Chung Chen, Wei-Hsiang Yu, Szu-Hua Chen, Yu-Chan Chang, Tai-I Hsu, Michael Hsiao, Chao-Yuan Yeh, and Cheng-Yu Chen. An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning. *Nature communications*, 12(1):1193, 2021.