Multiple Instance Learning with Center Embeddings for Histopathology Classification

Philip Chikontwe[1], Meejeong Kim[2], Soo Jeong Nam[2], Heounjeong Go[2,3], and Sang Hyun Park[1(✉)]

[1] Department of Robotics Engineering, DGIST, Daegu, South Korea
{philipchicco,shpark13135}@dgist.ac.kr
[2] Department of Pathology, Asan Medical Center, Seoul, South Korea
[3] Department of Pathology, University of Ulsan College of Medicine, Seoul, South Korea

Abstract. Histopathology image analysis plays an important role in the treatment and diagnosis of cancer. However, analysis of whole slide images (WSI) with deep learning is challenging given that the duration of pixel-level annotations is laborious and time consuming. To address this, recent methods have considered WSI classification as a Multiple Instance Learning (MIL) problem often with a multi-stage process for learning instance and slide level features. Currently, most methods focus on either instance-selection or instance prediction-aggregation that often fails to generalize and ignores instance relations. In this work, we propose a MIL-based method to jointly learn both instance- and bag-level embeddings in a single framework. In addition, we propose a center loss that maps embeddings of instances from the same bag to a single centroid and reduces intra-class variations. Consequently, our model can accurately predict instance labels and leverages robust hierarchical pooling of features to obtain bag-level features without sacrificing accuracy. Experimental results on curated colon datasets show the effectiveness of the proposed methods against recent state-of-the-art methods.

1 Introduction

In the current age of digitizing glass slides into histopathology images using whole-slide image (WSI) scanners, histopathology images play a vital role as a gold standard for cancer diagnosis in clinical settings. However, given the sheer volume of a single WSI (i.e. $100k$ pixels), effective analysis is considered both laborious and time consuming. Moreover, due to the high computational costs and the bias of subjective judgments among observers, automated and accurate analysis of WSIs promises improved diagnostics and better treatment strategies.

In particular, deep learning has become a popular solution and has yielded remarkable results when sufficiently labeled training data is provided. However, pixel level annotations are difficult and costly to acquire. Thus, multiple instance learning (MIL) based training of neural networks [2,5,9,12,13,20] presents a

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12265, pp. 519–528, 2020.
https://doi.org/10.1007/978-3-030-59722-1_50

unique solution to alleviate these challenges towards final WSI diagnosis without precise annotations. Though, due to the ambiguity of instance labels in MIL, learning robust instance embeddings is also challenging.

To address this, recent methods [4,7,10,19,23,25] often adopt a two-stage approach that includes (1) learning an instance encoder based on sampled regions from the WSI to obtain a prediction score or low dimensional feature, and (2) learning an aggregation model that uses the learned instance encoder to integrate instance level information for slide-level prediction. Though successful in some problem settings, this approach often fails when learning based on noisy labels and is worsened in the second stage of aggregation since features are not representative of the true labels. Therefore, an end-to-end framework that exploits both instance- and bag-level representations is beneficial. Also, providing interpretable results that follow standard pathology workflows can further aid experts to make informed decisions on diagnosis and treatment options. In this work, we propose a model that consists of top-k instance selection, instance- and bag-representation learning, as well as a center embedding module that reduces intra-class variations during learning via a center loss [21].

In addition, instance- and bag-losses with respect to the slide labels are also used to update the entire model. By this process, our model can improve feature representation and also use the instance level information to regularize bag level features. Moreover, instead of a decoupled feature aggregation step as in previous methods [2,9], our model consists of a bag module that creates a bag level feature via pyramidal feature aggregation. We consider a bag of k instances as a single input with k channels. Thus, we apply several convolution operations via the proposed bag module to obtain a single representation that considers key semantic information from earlier layers of our model. Consequently, we obtain a robust embedding that is effective for classification. On the other hand, despite using noisy labels in a MIL scenario, both instance and bag loss in our method enable features from different classes to be apart while center loss clusters features of the same class towards corresponding centers. Lastly, apart from the use of a bag classifier for diagnosis prediction, we also propose a *soft*-assignment method [22] to assign aggregated slide-level embeddings to the learned centers for final diagnosis. The benefit of using a *soft*-assignment is empirically shown in our experiments. Our main contributions are summarised as follows:

- We propose an end-to-end model for weakly supervised histopathology classification able to assign correct instance and bag labels.
- We show that center loss can improve bag feature learning and also reduce ambiguity in instance labels.
- By jointly considering instance based-MIL (i.e. accurate instance level predictions) and embedding-based MIL, our method shows improved classification performance with reduced false positives.
- Experimental results on two datasets further show that our approach is comparable in performance to methods that learn separate bag-level classifiers and use attention mechanisms for instance selection.

1.1 Related Works

We consider two categories of approaches in recent literature that use MIL for medical image analysis i.e. (a) methods that use MIL for tile selection [2,3,7,11,24], and (b) those that use unsupervised learning with MIL to learn instance embeddings [13,15,18].

MIL in Histology. MIL methods have been proposed for histopathology [2,4,5,7,8,10,24,25]. Li et al. [9] proposed a key instance learning using hard negative mining based on adaptive attention. However, attention based methods such as [8] may overfit when the number of slides is limited and fail to learn useful representations due to the bias of negative samples. More recently, Campanella et al. [2] proposed a two-step MIL method with impressive results on a large dataset using a recurrent neural network (RNN) for feature aggregation. However, RNNs are inherently difficult to train and have limited applicability. Although previous studies [8,9] show considerable improvements by using attention and its related variants, most methods propose multi-stage learning schemes that decouple instance level learning and instance aggregation, which inturn makes clinical deployment non trivial. To address these challenges, we propose a MIL method that jointly learns instance and bag embeddings in a single framework, showing the benefit of robust representative and separable features for classification.

Unsupervised Methods with MIL. In this setting, recent methods [13,15,18,22] aim to learn representative features across tissue heterogeneity enabling WSI regression such as survival analysis [24,25] since MIL may fail when instances in a slide are assigned noisy labels. Tellez et al. [18] proposed neural compression for WSIs based on unsupervisedly learnt instance encoders using contrastive self pre-training before classification. Lu et al. [13] proposed contrastive predicting coding (CPC) as pre-training for histology and combined it with MIL to achieve improved results. To address survival analysis, Muhammad et al. [15] proposed to cluster instances with a reconstruction based objective. However, in most cases, to learn effective slide clusters from instance embeddings, iterative methods such as K-means are required to initialize centers and require complex cluster re-assignment strategies. In this work, we address this with a center loss capable to clustering features based on a mini-batch without the requirement of iterating the entire dataset.

2 Methods

Given a WSI dataset $\mathcal{D} = \{S_1, \ldots, S_n\}$ where each S_i is a whole slide image with its label $y_i \in \{0, 1\}$. For each $S_i \in \mathcal{D}$ with m instances $S_i = \{s_{i1}, s_{i2}, \ldots, s_{im}\}$ s.t $m \leq M$, sampled from non-background regions with M denoting the total number of instances in a slide. Our goal is to assign the correct label y_i to each slide. In this setting, only slide level labels are available during training, where

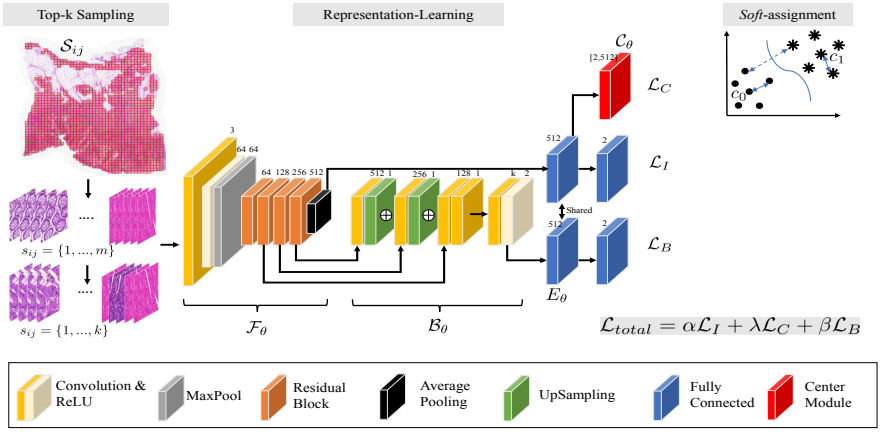

Fig. 1. Overview of the proposed framework. For a given slide, we first sample k instances at the start of each training epoch based on top predicted probabilities per slide via the instance module (i.e. $\mathcal{F}_\theta$ through $\mathcal{L}_I$ branch with shared embedding module E_θ). Top-k instances are feed through $\mathcal{F}_\theta$ to obtain embeddings with k channels via the bag module $\mathcal{B}_\theta$. Following, we feed the bag- and instance-embeddings to E_θ then the classification modules (bag and instance) to obtain final predictions and update the model.

1 means positive and 0 negative. Formally, MIL must satisfy the constraints: if S_i is negative, then all instances in S_i should be negative i.e. $y_i = 0, \forall(y_{ij}) = 0$, where y_{ij} are instance labels. If S_i is positive, then at least one instance should be positive, i.e. $y_i = 1$, $\sum y_{ij} \geq 1$.

To this end, we develop a convolutional neural network (CNN) model able to perform both instance level discrimination and slide level classification in an end-to-end framework. In particular, our method consists of (1) top-k instance selection (2) instance- and bag-representation learning, and (3) *soft*-assignment based inference with learned centers. Figure 1 shows the overall framework.

Top-k Instance Selection. We implement a neural network $\mathcal{F}_\theta(\cdot)$ to transform an instance from i-th slide(bag) into a low dimensional embedding $g_{ij} = \mathcal{F}_\theta(s_{ij})$ via the instance branch ($\mathcal{L}_I$ path in Fig. 1) with a shared embedding module E_θ, followed by a classification module that outputs the positiveness of an instance $p_{ij} = \mathcal{H}_\mathcal{I}(g_{ij})$ i.e. $\mathcal{H}_\mathcal{I}$ is binary classifier. Then, instance level probabilities $\{p_{i1}, \ldots, p_{im}\}$ from all bags are sorted to obtain top-k instances per slide for training i.e. we form a $\mathcal{D}_k \subset \mathcal{D}$. Note that in this procedure, parameters θ of $\mathcal{F}$, $\mathcal{B}$, and E as well as other modules are not updated or stored.

Instance Level Learning. In each training phase, for a given instance input $s_{ij} \in \mathcal{D}_k$ we obtain embeddings g_{ij} after global average pooling via $\mathcal{F}_\theta$. Each instance is assigned its corresponding bag-level label y and is used to compute the instance loss $\mathcal{L}_I$ given the prediction $\hat{y}$ after g_{ij} is feed to E_θ and an instance

classifier $\mathcal{H}_{\mathcal{I}}$ using cross-entropy. Formally, instance loss is presented as follows.

$$\mathcal{L}_I = -\sum y_{ij} \log p_{ij} \tag{1}$$

Pyramidal Bag-Level Learning. In addition to g_{ij}, three features maps from the earlier blocks of $\mathcal{F}_\theta$ are used for bag-level learning in $\mathcal{B}_\theta$. We feed each feature map to a series of convolutional blocks with input sizes $\{512, 256, 128\}$ corresponding to the feature map size and obtain reduced features each with a single channel. Next, we upsample the maps using linear interpolation to match the size at the previous block and concatenate with the previous feature map to form a single spatial map with k channels that is used to obtain the final flattened feature z_i. Following, z_i is passed to our shared embedding module E then the bag classifier $\mathcal{H}_{\mathcal{B}}(z_i)$. Bag predictions $\hat{y}$ are used to evaluate bag loss $\mathcal{L}_B(\hat{y}, y_i)$ using cross-entropy, formally:

$$\mathcal{L}_B = -\sum y_i \log \hat{y}. \tag{2}$$

In addition, to improve the discriminative power of the deep features; a center loss is proposed. It characterizes intra-class variations by learning embeddings that minimize the distance of instances from the same bag. Formally, center loss is expressed as:

$$\mathcal{L}_C = \frac{1}{2} \sum_{i=1}^{u} \parallel g_{ij} - c_{y_i} \parallel_2^2 . \tag{3}$$

In Eq.(3), c_{y_i} denotes the y_{ij}th class center's deep features with same dimension as g_{ij} and u is the mini-batch size. In this work, class centers are parameterized by the center module $\mathcal{C}_\theta(\cdot)$ initialized from a standard normal distribution and trained jointly with $\mathcal{L}_C$ and $\mathcal{L}_B$. Intuitively, instance-embeddings from the same bag should have similar features that can cluster to similar points in the embedding space. The class centers are updated based on instance-embeddings in a mini-batch rather than the entire training set which would be impractical for large datasets. The final loss function is defined as:

$$\mathcal{L} = \alpha \mathcal{L}_I + \lambda \mathcal{L}_C + \beta \mathcal{L}_B, \tag{4}$$

where λ, α and β balance the losses.

***Soft*-Assignment Based Inference**. In the testing stage, given a bag embedding obtained via $\mathcal{B}$, our goal is to assign the correct label $\hat{y}$ as the final diagnosis. To achieve this, we infer that instance-embeddings of the same bag should be matched to a single centroid that represents the bag-label. Given a bag-level embedding $z_i = \mathcal{B}\{g_{i1}, \ldots, g_{ik}\}$ for slide $\mathcal{S}_i$, we assign z_i to the learned centroid c_{y_i} using the students t-distribution [14] as a kernel to measure the similarity between the two points as:

$$q_i = \frac{(1 + \parallel z_i - c_{y_i} \parallel^2 / \varphi)^{-\frac{\varphi+1}{2}}}{\sum_{i'} (1 + \parallel z_i - c_{y_{i'}} \parallel^2 / \varphi)^{-\frac{\varphi+1}{2}}} \tag{5}$$

where q_i can be considered as the probability of assigning z_i to a class center and φ are the degrees of freedom of the Students distribution.

3 Experiments and Results

Dataset and Settings. We conduct experiments on two colectoral cancer (CRC) WSI datasets collected at Anonymous medical center with different scanning conditions i.e. Dataset I and II. The datasets involve Hematoxylin and Eosin stained (H&E) slides of normal and malignant tissues scanned at x40 magnification using different scanners. CRC is the third most prevalent cancer in humans and common cause of death in both male and females. In this work, malignant slides involve microsatellite instable (MSI) [1] CRC which is a molecular phenotype due to defective DNA mismatch repair system. Expert pathologists curated and detected the presence of MSI using Immunohistochemical analysis (IHC) and PCR-based amplication. Determination of MSI status in CRC has prognostic and therapeutic implications.

Dataset I consists of 173 WSI (59 normal & 114 MSI) whereas set II has 193 (85 normal & 108 MSI) WSIs. We follow a 40%-10%-50% split to create non-overlapping sets for training, validation and testing, respectively. For each WSI, we applied the Otsu method [17] to filter-out non tissue regions after conversion to HSV color space. Patch candidate locations were selected randomly for extraction per slide during training and validation, respectively. During training and inference, we set top-k instance selection to 50 and used instances with 256×256 size. A ResNet-34 [6] was finetuned and used as the feature extraction model $\mathcal{F}_\theta(\cdot)$. In addition, Fully Connected (FC) layers was employed for both the instance and bag classifiers $\mathcal{H}_\mathcal{I}$ and $\mathcal{H}_\mathcal{B}$. The feature dimension of the embeddings and center module was set to 512. The entire framework was trained end-to-end with a learning rate of $1e{-}4$ for 40 epochs. We empirically set the loss hyperparameters $\lambda = 0.01$, $\alpha = 1.0$ and $\beta = 0.01$, respectively.

Comparison Methods. In this work, we compare the proposed method with the recent state-of-the-art method in [2][1], we term this method as 'CampanellaMILRNN'. In addition, we also evaluated against DeepAttenMIL [8][2] and TwoStageMIL [16][3] methods. For a fair comparison, the same backbone $\mathcal{F}_\theta(\cdot)$ is used in all cases. Both CampanellaMILRNN and TwoStageMIL use a two-stage learning procedure i.e. (1) instance level learning and (2) slide-level aggregation. DeepAttenMIL uses an end-to-end approach with permutation invariant pooling based on the attention mechanism.

[1] We used the publicly available implementations: https://github.com/MSKCC-Computational-Pathology/MIL-nature-medicine-2019.
[2] https://github.com/AMLab-Amsterdam/AttentionDeepMIL.
[3] https://github.com/ImagingLab/ICIAR2018.

Table 1. Evaluation of the proposed methods on Dataset I with $k = 50$. 'bag' implies the use of the bag classifier and 'soft' means with soft-assignment during testing.

Method	F1Score	Precision	Recall	Accuracy
CampanellaMIL [2]	76.8	84.46	79.49	79.49
CampanellaMIL+RNN [2]	80.16	85.98	82.05	82.05
DeepAttenMIL [8]	85.95	86.03	85.90	85.90
TwoStageMIL [16]	65.21	80.41	71.79	71.79
Ours (w/o center loss)	86.93	87.2	87.18	87.18
Ours (w/all loss + bag)	86.77	87.58	87.18	87.18
Ours (w/all loss + soft)	**92.36**	**92.54**	**92.31**	**92.31**

Quantitative Results. Tables 1 and 2 shows the performance of the proposed method against recent approaches. Among the comparison methods, the proposed solution with a *soft*-assignment based inference achieved the best result. Since 'CampanellaMIL' considers the probability of the top-most instance as the final slide classification, it performed poorly. Moreover, when our method was evaluated using bag classification only, we note consistent improvements over other methods i.e. +9.97% over CampanellaMIL's method, whereas our assignment based approach showed +15.56%. Notably, DeepAttenMIL was the best among the compared methods showing the benefit of attention based aggregation. Though no attention was used in our method, we report better performance. This leads us to believe performance of our model can further increase if attention is applied in our bag module.

In most cases, soft-assignment shows to be a good alternative over bag classifiers. We argue that since the learned centers capture the maximum information among similar instance embeddings; soft inference can be more robust than using a bag classifier in some settings. Results on dataset II also show that our method has the best performance. Though the RNN-based aggregation was the best among others on this set, our method showed a +1.28% improvement. Given previous methods emphasize decoupling of instance and bag level learning, evidence serves to show that there is no clear advantage in separating the two procedures and supports our initial objective of a single framework. Further, the majority of methods showed high performance on dataset II compared to the first. Due to the difference in scanning protocols used in preparation, performance may vary. However, note that no color normalization methods were used in pre-processing since we follow the standard protocol introduced in [2] that did use recent normalization techniques.

Qualitative Results. In Fig. 2, we present qualitative results of our method in two aspects (1) top-k patches that the model samples per slide-class, and (2) effectiveness of the learned model in interpretability via segmentation. In clinical workflows, it is beneficial to view which regions the model focuses on when

Table 2. Evaluation of the proposed methods on Dataset *II* with $k = 50$. 'bag' implies the use of the bag classifier and 'soft' means with soft-assignment during testing.

Method	F1Score	Precision	Recall	Accuracy
CampanellaMIL [2]	92.06	93.13	92.31	92.31
CampanellaMIL + RNN [2]	97.41	97.53	97.44	97.44
DeepAttenMIL [8]	93.65	94	93.59	93.59
TwoStageMIL [16]	89.55	89.89	89.74	89.74
Ours (w/o center loss)	93.43	94.17	93.59	93.59
Ours (w/all loss + bag)	97.41	97.53	97.44	97.44
Ours (w/all loss + soft)	**98.71**	**98.74**	**98.72**	**98.72**

making a decision to aid experts. Thus, we visually validate the effectiveness of the proposed methods by collecting the top-$k = 5$ patches with both high and low positiveness predicted by the instance classifier. Notably, the lowest predicted instances all correspond to actual normal tissues whereas those with high probability are regions of highly malignant nuclei clustered together. This shows that our model can accurately separate ambiguous labels in each slide. In addition, we performed patch-wise classification on entire slides using the trained model to obtain a heatmap that is thresholded and shows regions of high tumor probability. Notably, predictions correctly match expert annotations. This is attributed to the sampling step that accurately selects suspicious instances in each phase and avoids false negatives. The results validate our initial assertions and show that our method has a good trade-off between learning instance and bag features.

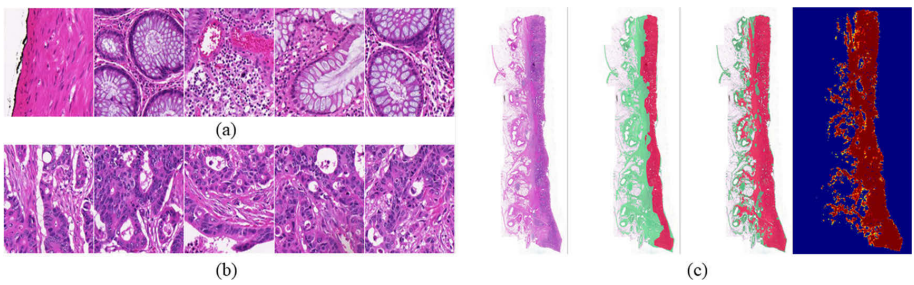

Fig. 2. Top-k examples and segmentation. (a) Instances with lowest probability (least-positiveness) (b) Instance with highest probability (positiveness) (c) Patch-based segmentation results of the proposed method: From left to right, WSI, WSI ground-truth annotation and predicted malignant regions. (Green) Normal regions w/o muscle and irrelevant tissues (Red) Suspicious tumor regions. (Color figure online)

4 Conclusion and Future Work

In this work, we presented a novel framework for histopathology slide classification in a multiple instance learning (MIL) setting. Two key concepts are introduced: (a) an end-to-end framework for both instance and bag level learning with a center loss that minimizes intra-class distances in the embedding space. (b) we proposed a *soft*-assignment based inference method for bag-level prediction. The proposed method was validated on two Colectoral cancer (CRC) datasets with overall improved performance over recent state-of-the-art methods. More importantly, the benefit of using a center loss is evident in the discriminative and generalized nature of learnt embeddings. Future works will explore multi-class multi-instance settings as well as the applicability of this method in WSI tumor grading.

Acknowledgment. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2019R1C1C1008727), and the Grant of Artificial Intelligence Bio-Robot Medical Convergence Technology funded by the Ministry of Trade, Industry and Energy, the Ministry of Science and ICT, and the Ministry of Health and Welfare (No. 20001533).

References

1. Boland, C.R., Goel, A.: Microsatellite instability in colorectal cancer. Gastroenterology **138**(6), 2073–2087 (2010)
2. Campanella, G.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. **25**(8), 1301–1309 (2019)
3. Chen, H., et al.: Rectified cross-entropy and upper transition loss for weakly supervised whole slide image classifier. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 351–359. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_39
4. Cosatto, E., et al.: Automated gastric cancer diagnosis on h and e-stained sections; training a classifier on a large scale with multiple instance machine learning. In: Medical Imaging 2013: Digital Pathology, vol. 8676, p. 867605. International Society for Optics and Photonics (2013)
5. Hashimoto, N., et al.: Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with non-annotated histopathological images. arXiv preprint arXiv:2001.01599 (2020)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H.: Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2424–2433 (2016)
8. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning. arXiv preprint arXiv:1802.04712 (2018)

9. Li, M., Wu, L., Wiliem, A., Zhao, K., Zhang, T., Lovell, B.: Deep instance-level hard negative mining model for histopathology images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, pp. 514–522. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_57
10. Li, S.: Multi-instance multi-scale CNN for medical image classification. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11767, pp. 531–539. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32251-9_58
11. Li, W., Nguyen, V.-D., Liao, H., Wilder, M., Cheng, K., Luo, J.: Patch transformer for multi-tagging whole slide histopathology images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 532–540. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_59
12. Lin, Y., Zhang, H.: Regularized instance embedding for deep multi-instance learning. Appl. Sci. **10**(1), 64 (2020)
13. Lu, M.Y., Chen, R.J., Wang, J., Dillon, D., Mahmood, F.: Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding. arXiv preprint arXiv:1910.10825 (2019)
14. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
15. Muhammad, H., et al.: Unsupervised subtyping of cholangiocarcinoma using a deep clustering convolutional autoencoder. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 604–612. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_67
16. Nazeri, K., Aminpour, A., Ebrahimi, M.: Two-stage convolutional neural network for breast cancer histology image classification. In: Campilho, A., Karray, F., ter Haar Romeny, B. (eds.) ICIAR 2018. LNCS, vol. 10882, pp. 717–726. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93000-8_81
17. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
18. Tellez, D., Litjens, G., van der Laak, J., Ciompi, F.: Neural image compression for gigapixel histopathology image analysis. IEEE Trans. Pattern Anal. Mach. Intell. (2019)
19. Wang, X.: Weakly supervised deep learning for whole slide lung cancer image analysis. IEEE Trans. Cybern. **50**, 3950–3962 (2019)
20. Wang, X., Yan, Y., Tang, P., Bai, X., Liu, W.: Revisiting multiple instance neural networks. Pattern Recogn. **74**, 15–24 (2018)
21. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31
22. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: International Conference on Machine Learning, pp. 478–487 (2016)
23. Xu, G., et al.: CAMEL: a weakly supervised learning framework for histopathology image segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10682–10691 (2019)
24. Yao, J., Zhu, X., Huang, J.: Deep multi-instance learning for survival prediction from whole slide images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 496–504. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_55
25. Zhu, X., Yao, J., Zhu, F., Huang, J.: WSISA: making survival prediction from whole slide histopathological images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7234–7242 (2017)