Weakly Semi-Supervised Cervical Lesion Cell Detection via Twin-Memory Augmented Multiple Instance Learning

Manman Fei[1†], Zhiyun Song[1†], Zhenrong Shen[2], Mengjun Liu[1], Qian Wang[2(✉)], and Lichi Zhang[1(✉)]

[1] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
lichizhang@sjtu.edu.cn
[2] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
qianwang@shanghaitech.edu.cn

Abstract. Deep learning methods have demonstrated promising results in cervical lesion cell detection. Training detection models that generalize well typically require a large amount of cell-level annotations that are expensive and time-consuming to obtain. Instead, weak slide-level annotations, which entail assigning a gigapixel whole slide image (WSI) with a single label, are easier to acquire. However, due to significant differences in annotation scales, they cannot be directly utilized to assist in the training of cervical cell detectors. To address this challenge, we propose a Twin-memory augmented Multiple Instance Learning (Twin-MIL) framework to refine cervical lesion cell detection. Firstly, we utilize the multiple instance learning to bridge the gap between cell-level and slide-level tasks. Then, we reduce false positives in conventional MIL by introducing a twin-memory module, which improves the classification capability by capturing more discriminative patterns of positive and negative cells. We also propose uncertainty-regulated negative instance learning to enhance the robustness of negative latent space against noisy instances and its separability from the positive one. Experiments indicate that our method is effective in enhancing different detection models trained on the datasets with varying annotation levels.

Keywords: Cervical lesion cell detection · Whole slide images · Weakly semi-supervised learning.

1 Introduction

Cervical cancer ranks as the fourth most common cancer among females, with an annual occurrence of over half a million new cases and 300,000 deaths [22]. Early diagnosis of cervical cancer through formalized screening programs and HPV testing allows for fertility-sparing treatment and contributes to a decrease in both incidence and mortality rates [5]. Among various cervical screening methods, the thin-prep cytology test (TCT) stands out as one of the most frequently

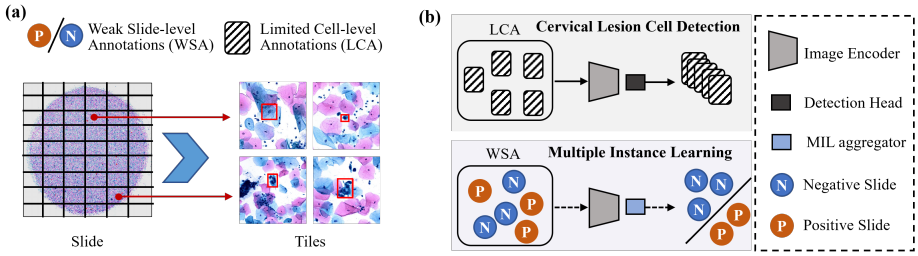

Fig. 1. (a) Two annotation types. (b) Our method bridges the gap between lesion cell detection trained with LCA and slide classification trained with WSA by sharing the detector's backbone with the MIL strategy.

used techniques due to its high effectiveness in detecting abnormal and premalignant cervical lesions [12]. Traditionally, TCT is performed by visually examining stained cells collected on a glass slide under a microscope, followed by the generation of the diagnosis report according to The Bethesda System (TBS) [15]. However, the examination of gigapixel whole slide images (WSIs) in TCT is often time-consuming, prone to errors, and demonstrates a high degree of intra/inter-observer variability [1]. Therefore, there is a pressing need for automated cervical cell analysis methods to aid cytologists in effectively analyzing cervical cytology images and delivering accurate, objective, and efficient diagnoses.

The detection of cervical lesion cells plays an important role in the automated analysis of cervical cytology images [28, 27, 3, 26]. Current studies employ fully supervised deep learning technology to detect cervical lesion cells in TCT screening [2, 4, 9, 24]. However, these methods heavily rely on abundant cell-level annotated data, which is often hard to collect due to the time-consuming and labor-intensive procedure for annotating each cell [16]. One possible approach is to utilize extra tile-level annotations to assist cell-level detection, which has been previously studied in the field of cervical cytology. For example, Shi et al. [19] leveraged cross-attention scores learned from tile classification as pseudo labels for cell detection. Fei et al. [6] proposed a method to distill knowledge from a tile classification network to enhance the robustness of cervical lesion cell detection. However, tile-level annotations are still not easily available in the scenario of high-resolution cervical cancer analysis. Instead, weak slide-level annotations, which involve associating a single label with hundreds of tiles, are significantly more accessible. Although several studies have explored the usage of slide-level annotations to assist tile-level classification [13, 20], the effective incorporation of slides to enhance cell-level detection remains an open challenge.

In this paper, we propose a **Twin**-memory augmented **M**ultiple **I**nstance **L**earning (Twin-MIL) framework, which enhances cervical lesion cell detection using a dataset with limited cell-level annotations (LCA) and a dataset with weak slide-level annotations (WSA). As shown in Fig. 1, Twin-MIL bridges the gap between cervical lesion cell detection trained on LCA and slide classification using WSA by sharing the detector with top-K ranking-based MIL. Furthermore,

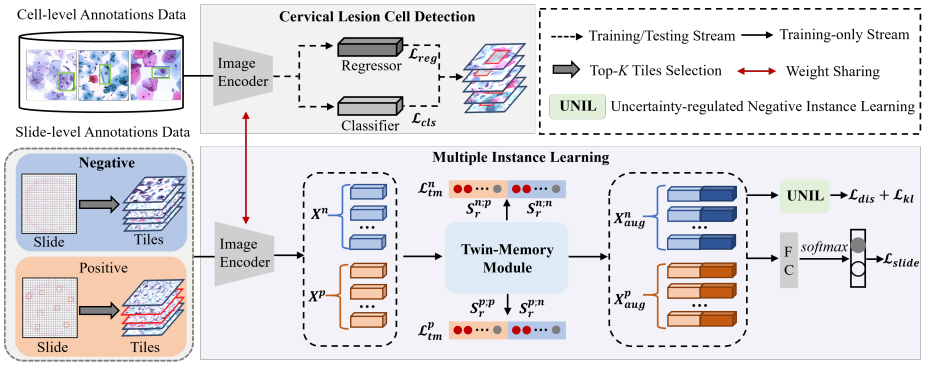

Fig. 2. The overview of Twin-MIL. The detection model is initially trained using LCA and further refined using WSA. The refinement process commences with top-K selection for relevant tiles, which are further encoded into positive features X^p and negative features X^n by the detector's encoder. The twin-memory module processes these features to yield the similarity scores, and augmented features. The UNIL further constrains the negative augmented features and enlarges the margin between negative and positive.

a twin-memory module is proposed to enhance the MIL process by capturing more discriminative patterns from both positive and negative slides. Additionally, uncertainty-regulated negative instance learning is designed to obtain more separable features of negative latent space and further reduce the false positives. Our main contributions are as follows: 1) We propose a novel framework to enhance the detector's performance by training on hybrid datasets with cell-level and weak slide-level annotations. 2) We introduce a novel multiple-instance learning strategy using twin memory augmentation to ensure effective guidance from both the positive and negative slides. 3) We design an uncertainty-regulated negative instance learning to improve the separability of latent space and the robustness against fluctuation. 4) We conduct experiments using various detectors and different dataset conditions, all of which consistently demonstrate the effectiveness and robustness of the proposed method.

2 Methods

Our purpose is to utilize the weak slide-level annotations (WSA) to enhance cell-level detection initially trained by limited cell-level annotations (LCA). As shown in Fig. 2, we present a top-K ranking-based MIL to learn effective representations for the detectors. Moreover, we introduce a twin memory module to generate more discriminative features for MIL using the selected tiles. Furthermore, we propose uncertainty-regulated negative learning to further improve the robustness and separability of negative latent space from the positive one.

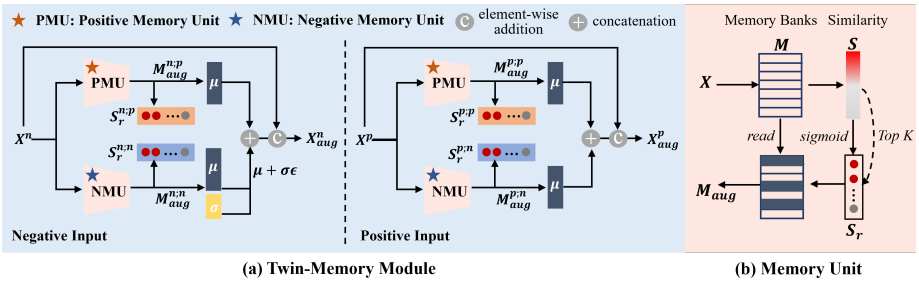

Fig. 3. (a) The structure of the Twin-Memory Module, which consists of a Positive Memory Unit and a Negative Memory Unit. (b) The mechanism of the Memory Unit.

2.1 Multiple Instance Learning with Twin-Memory Module

We adopt a top-K ranking-based MIL [8] as the basic framework, and introduce a cell-level memory-based mechanism to effectively utilize fine-grained priors from weakly annotated slides. Memory mechanisms have been widely used to retain key prototypes, but most existing methods rely on a single memory unit, leading to an overemphasis on positive instances. To address this issue, we introduce a twin-memory module (TMM), designed to learn discriminative cell-level features by capturing both positive and negative prototypes. The structure of TMM is shown in Fig. 3 (a). On the left side, the negative tile features pass through the twin memory units to form relevant tile output scores and augmented features. Similarly, the positive tile features follow the same process on the right side.

The mechanism of the Memory Unit is shown in Fig. 3 (b). Similar to previous work [14], a memory bank is designed as a matrix $\mathbf{M} \in \mathbb{R}^{M \times C}$, where M and C are numbers and dimensions of prototypes, respectively. The cell-level features $\mathbf{X} \in \mathbb{R}^{NHW \times C}$ are obtained by encoding the selected N tiles in a slide, where H and W are the spatial dimensions of the feature map. The similarity scores $\mathbf{S} \in \mathbb{R}^{NHW \times M}$ between each instance and the prototypes are calculated as:

$$\mathbf{S} = \sigma\left(\frac{\mathbf{X}\mathbf{M}^T}{\sqrt{C}}\right), \quad \mathbf{M}_{\text{aug}} = \mathbf{S}\mathbf{M}, \tag{1}$$

where σ is the sigmoid operation. Then the memory augmentation feature generated by a read operation is denoted as M_{aug}. The top-K most similar prototypes are selected to produce the rectified similarity scores via

$$\mathbf{S}_r = \frac{1}{K}\sum_{j=1}^{K} top\text{-}K(\mathbf{S})(:, j), \tag{2}$$

which means the rectified scores are obtained by averaging the top-K sorted original scores along the second dimension. When sending the negative embedding $\mathbf{X}^n$ to the positive memory bank, the positive query scores denoted as $\mathbf{S}_r^{n;p}$ should be $\mathbf{0} \in \mathbb{R}^{NHW}$. Conversely, when X^n is sent to the negative memory bank,

the negative query scores $S_r^{n;n}$ are constrained to $1 \in \mathbb{R}^{NHW}$. The memory loss function for negative instances is:

$$L_{tm}^n = \text{BCE}(\mathbf{S}_r^{n;p}, \mathbf{0}) + \text{BCE}(\mathbf{S}_r^{n;n}, \mathbf{1}), \tag{3}$$

where BCE means binary cross-entropy.

Different from the negative tiles that only contain negative cells, the positive tiles might consist of a combination of negative and positive cells. Therefore, the memory loss function for the positive tiles is:

$$L_{tm}^p = \text{BCE}(\frac{1}{K}\sum_{i=1}^{K} \textit{top-K}(\mathbf{S}_r^{p;p})(i), \mathbf{1}) + \text{BCE}(\frac{1}{K}\sum_{i=1}^{K} \textit{top-K}(\mathbf{S}_r^{p;n})(i), \mathbf{0}), \tag{4}$$

which means the loss function only constrains top-K instances for positive tiles.

2.2 Uncertainty-regulated Negative Instance Learning

We further regulate the negative features to improve the robustness against fluctuations in negative latent space and its separability from positive ones. Specifically, we update the memory bank to ensure that augmented negative memory features $\mathbf{M}_{aug}^n = \mathbf{S}\mathbf{M}^n$, derived from negative inputs via the negative memory bank, are distinct from augmented positive memory features $\mathbf{M}_{aug}^p$, obtained from positive inputs via the positive memory bank. Similar to the previous method [17], we achieve this by constraining the negative features with Gaussian distribution and derivating the positive features out of this distribution. We use μ^n and σ^n to denote the encoded mean and variance (uncertainty) of $\mathbf{M}_{aug}^n$, and use the reparameterization trick to obtain the negative regulated features $z^n = \mu^n + \sigma^n \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$. We implement Kullback-Leibler loss to ensure the stability of σ via:

$$L_{kl} = \text{KL}(\mathcal{N}(z^n|\mu^n, \sigma^n)||\mathcal{N}(0, 1)). \tag{5}$$

Similar to the negative slides, we can also obtain the encoded mean μ^p of $\mathbf{M}_{aug}^p$. Based on the presumption that the positive slides are out of distribution, we use a distance loss to separate them in the latent space:

$$L_{dis} = \text{ReLU}(d - (||\mu^p||_2^2 - ||z^n||_2^2)), \tag{6}$$

where d is a constant that measures the desired discrepancy between features. The regulated features are combined with the original embeddings $\mathbf{X}$ before feeding to the slide classifier. The multiple instance learning loss function is:

$$L_{MIL} = L_{slide} + \lambda_1(L_{tm}^n + L_{tm}^p) + \lambda_2 L_{kl} + \lambda_3 L_{dis}, \tag{7}$$

where L_{slide} is the BCE loss for the classification of cervical cytology images. We employ augmented features processed through a fully connected (FC) layer for slide-level classification. $\lambda_1, \lambda_2, \lambda_3$ are used to balance the auxiliary losses.

2.3 Optimization

To better utilize MIL and optimize the detector, we adopt the following training strategy. In the first stage, we train the detector using LCA until it converges. In the second stage, we optimize the image encoder within the detector using WSA. At this stage, we freeze the detection head to maintain the learned detection capabilities. The goal is to improve the representation learning of the image encoder by leveraging WSA. The total objective function is:

$$L_{total} = L_{cls} + L_{reg} + L_{MIL}, \tag{8}$$

where L_{cls} and L_{reg} are the ordinary detection loss for each detection head. During inference, only the optimized detector is used to output the final detection results without the need for any additional modules or components.

3 Experiments

3.1 Dataset and Experimental Setup

We evaluate the proposed method using two datasets: one with limited cell-level annotations (LCA) and the other with weak slide-level annotations (WSA). The LCA dataset comprises a total of 5035 tiles with 7,321 annotated cells, each with a resolution of 1024×1024 pixels. These images have been meticulously annotated by three experienced cytologists. The annotation process involved two cytologists responsible for the initial annotations, and the accuracy of these annotations was further verified by a third cytologist. The dataset is divided into training, validation, and test sets with a ratio of 8:1:1, respectively. The WSA dataset consists of 1000 positive slides and 1000 negative slides. These slides do not have cell-level annotations by cytologists; instead, they are labeled solely with positive or negative tags. All samples were obtained from Shanghai Medical College Hospital, Shanghai Cancer Hospital, and Suzhou Dushu Lake Hospital. The tiles in the LCA dataset are cropped from the WSI. We divided the datasets based on patient case IDs, ensuring there is no overlap.

In the first stage, we train the detection network using the LCA dataset. The initial learning rate is set to 1e-4, and we use the Adam optimizer with a weight decay of 1e-4 and batch size of 4. In the second stage, we finetune the image encoder using the WSA dataset. We utilize the pre-trained detection model to perform detection on these slides, selecting the top N=8 tiles based on their detection scores. The top 8 tiles per slide are fixed. Additionally, we set the batch size to 2 during this process, which consists of one positive case and one negative case, with 8 tiles per case. In our implementation, we set the hyperparameters λ_1 to 0.1, λ_2 to 0.001, and λ_3 to 0.0001. Additionally, we set K to 64 and define the memory bank size as 60. We perform a grid search for key hyperparameters and select the parameter set that shows the best performance on the validation set. We conduct a quantitative evaluation using three types of metrics: the COCO-style [11] average precision (AP), average recall (AR), and F1-score.

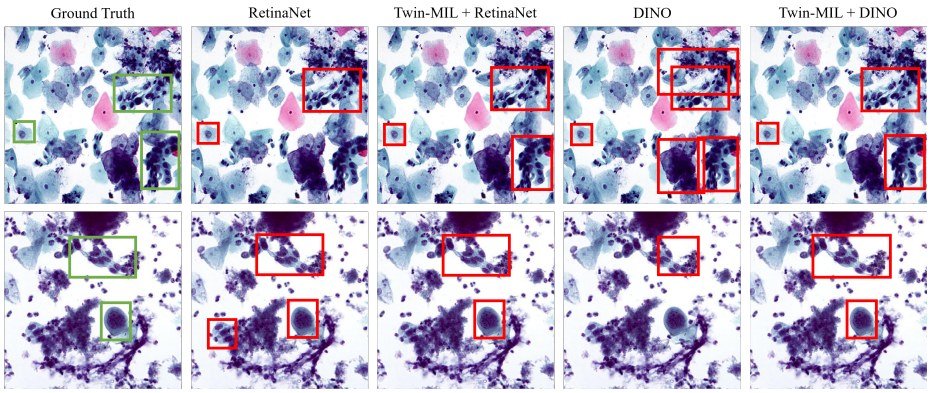

Fig. 4. Qualitative comparisons of our method on different detectors. The green bounding boxes represent the ground truth annotations, while the red bounding boxes indicate the predicted bounding boxes.

Table 1. Performance comparison with state-of-the-art detection methods.

Method	$AP_{0.5:0.95}$(%)	$AP_{0.5}$ (%)	AR (%)	F1-score (%)
Faster R-CNN [18]	30.3	70.4	41.3	42.3
Sparse R-CNN [21]	29.4	68.9	43.2	40.4
YOLOv8 [7]	33.9	71.0	43.6	45.8
AttFPN [23]	35.8	76.2	46.6	49.1
Cascade RRAM and GRAM [9]	32.4	73.8	44.3	45.0
RetinaNet [10]	29.1	70.2	44.8	41.1
Twin-MIL + RetinaNet	32.7	75.1	**49.4**	45.5
DINO [25]	36.5	76.0	45.7	49.3
Twin-MIL + DINO	**38.1**	**78.9**	47.0	**51.4**

3.2 Comparison with State-of-the-art Detection Methods

Detectors augmented with Twin-MIL are compared with other state-of-the-art detection methods for cervical lesion cells to demonstrate the benefits brought by the weak supervision from slides and the effectiveness of our design. In our experiments, we apply Twin-MIL to both RetinaNet and DINO detectors. As indicated in Table 1, leveraging WSA for weak supervision effectively improves the detection performance. When using RetinaNet as the detector, we observe a clear improvement in the detection performance, which matches that achieved by more sophisticated detectors [18, 21, 9]. Furthermore, by adopting Twin-MIL for DINO, we achieve a new state-of-the-art performance by a clear margin. We also visualize the detection results from different detectors. As shown in Fig. 4, it can be observed that Twin-MIL achieves good performance on RetinaNet and DINO due to improved identification of lesion cells and reduced false positives.

Table 2. Performance comparison for different amounts of cell-level annotated data.

	Method	$AP_{0.5:0.95}$ (%)	$AP_{0.5}$ (%)	AR (%)	F1-score (%)
40%	RetinaNet	21.3	52.8	40.2	30.3
	+Twin-MIL	**25.4**	**61.3**	**43.3**	**35.9**
60%	RetinaNet	24.2	58.3	39.6	34.2
	+Twin-MIL	**29.1**	**68.9**	**44.1**	**40.8**
80%	RetinaNet	28.2	65.9	44.0	39.5
	+Twin-MIL	**31.5**	**74.1**	**48.4**	**44.2**
100%	RetinaNet	29.1	70.2	44.8	41.1
	+Twin-MIL	**32.7**	**75.1**	**49.4**	**45.5**

3.3 Weakly Semi-supervised Detection with Different Settings

In this section, we evaluate the effectiveness of the proposed method for different amounts of cell-level annotated data along with weak slide-level annotations. We compare the performance of our proposed method with that of RetinaNet, a baseline model trained solely on the same number of cell-level annotated images. Table 2 presents the evaluation results, showing the $AP_{0.5}$ and the F1-score for different amounts of LCA. Our proposed method achieves improvements in $AP_{0.5}$ of 3.4%, 8.5%, 10.6%, 8.2%, and 4.9%, and improvements in F1-score of 1.4%, 5.6%, 6.6%, 4.7%, and 4.4% compared to solely training on the same number of annotated cell images. It provides evidence that incorporating WSA can be a valuable strategy for boosting cell detection performance, especially in scenarios where the number of fully annotated cell images is limited.

3.4 Ablation Study

We further conduct ablation studies to evaluate the components and loss settings in the proposed method. We can learn from Table 3 that the incorporation of weakly annotated slides using the slide loss (SL) L_{slide} can effectively enhance the detector's performance. Furthermore, it is noteworthy that the twin-memory module (TMM) is more effective with the introduction of twin memory loss (TML), whose performance is further improved from 31.5% to 32.0% in $AP_{0.5:0.95}$ due to the improved discriminability of the negative and positive instances. In addition, superior results are achieved with the regularization of uncertainty learning (UNIL), which can be attributed to enhanced robustness and separability of latent space.

4 Conclusion

In conclusion, we propose a novel Twin-memory augmented Multiple Instance Learning (Twin-MIL) framework for weakly semi-supervised cervical lesion cell detection. By leveraging slide-level annotations, Twin-MIL improves the accuracy of cervical cell localization and classification. Experiments with different

Table 3. Quantitative results for the ablation study.

SL	TMM	TML	UNIL	$AP_{0.5:0.95}$ (%)	$AP_{0.5}$ (%)	AR (%)	F1-score (%)
				29.1	70.2	44.8	41.1
✓				31.6	72.8	48.5	44.0
✓	✓			31.5	74.1	48.4	44.2
✓	✓	✓		32.0	74.4	49.0	45.4
✓	✓	✓	✓	**32.7**	**75.1**	**49.4**	**45.5**

dataset settings demonstrate the effectiveness of Twin-MIL in localizing and classifying cervical cells in cytology images. We also demonstrate the universality of the proposed method by adopting Twin-MIL to various detection models. In summary, Twin-MIL provides an effective solution for balancing improved performance and reduced annotation costs in the field of cervical cytology. In future work, we plan to integrate Twin-MIL into existing clinical workflows and conduct comprehensive validation studies to assess its real-world impact on cervical cancer screening programs.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 62471288).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bengtsson, E., Malm, P.: Screening for cervical cancer using automated analysis of pap-smears. Computational and Mathematical Methods in Medicine **2014** (2014)
2. Cao, L., Yang, J., Rong, Z., Li, L., Xia, B., You, C., Lou, G., Jiang, L., Du, C., Meng, H., et al.: A novel attention-guided convolutional network for the detection of abnormal cervical cells in cervical cancer screening. Medical Image Analysis **73**, 102197 (2021)
3. Chai, S., Xin, J., Wu, J., Yu, H., Liang, Z., Ma, Y., Zheng, N.: Dpd-net: Dual-path proposal discriminative network for abnormal cell detection in cervical cytology images. Biomedical Signal Processing and Control **89**, 105887 (2024)
4. Chen, T., Zheng, W., Ying, H., Tan, X., Li, K., Li, X., Chen, D.Z., Wu, J.: A task decomposing and cell comparing method for cervical lesion cell detection. IEEE Transactions on Medical Imaging **41**(9), 2432–2442 (2022)
5. Cohen, P.A., Jhingran, A., Oaknin, A., Denny, L.: Cervical cancer. The Lancet **393**(10167), 169–182 (2019)
6. Fei, M., Zhang, X., Cao, M., Shen, Z., Zhao, X., Song, Z., Wang, Q., Zhang, L.: Robust cervical abnormal cell detection via distillation from local-scale consistency refinement. In: International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 652–661. Springer Nature Switzerland, Cham (2023)
7. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.: Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 (2021)

8. Li, W., Vasconcelos, N.: Multiple instance learning for soft bags via top instances. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4277–4285 (2015)
9. Liang, Y., Feng, S., Liu, Q., Kuang, H., Liu, J., Liao, L., Du, Y., Wang, J.: Exploring contextual relationships for cervical abnormal cell detection. IEEE Journal of Biomedical and Health Informatics **27**(8), 4086–4097 (2023)
10. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2980–2988 (2017)
11. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European Conference on Computer Vision. pp. 740–755. Springer Nature Switzerland, Cham (2014)
12. Liu, Y., Zhang, L., Zhao, G., Che, L., Zhang, H., Fang, J.: The clinical research of thinprep cytology test (tct) combined with hpv-dna detection in screening cervical cancer. Cellular and Molecular Biology **63**(2), 92–95 (2017)
13. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021)
14. Luo, W., Zhang, T., Yang, W., Liu, J., Mei, T., Wu, F., Zhang, Y.: Action unit memory network for weakly supervised temporal action localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9969–9979 (2021)
15. Nayar, R., Wilbur, D.C.: The bethesda system for reporting cervical cytology: a historical perspective. Acta Cytologica **61**(4-5), 359–372 (2017)
16. Pirovano, A., Almeida, L.G., Ladjal, S., Bloch, I., Berlemont, S.: Computer-aided diagnosis tool for cervical cancer screening with weakly supervised localization and detection of abnormalities using adaptable and explainable classifier. Medical Image Analysis **73**, 102167 (2021)
17. Ramírez Rivera, A., Khan, A., Bekkouch, I.E.I., Sheikh, T.S.: Anomaly detection based on zero-shot outlier synthesis and hierarchical feature distillation. IEEE Transactions on Neural Networks and Learning Systems **33**(1), 281–291 (2022)
18. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015)
19. Shi, J., Wu, K., Zheng, Y., He, Y., Li, J., Jiang, Z., Yu, L.: Global-local attention network for weakly supervised cervical cytology roi analysis. In: 19th International Symposium on Biomedical Imaging. pp. 1–5 (2022)
20. Stegmüller, T., Abbet, C., Bozorgtabar, B., Clarke, H., Petignat, P., Vassilakos, P., Thiran, J.P.: Self-supervised learning-based cervical cytology for the triage of hpv-positive women in resource-limited settings and low-data regime. Computers in Biology and Medicine **169**, 107809 (2024)
21. Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., Tomizuka, M., Li, L., Yuan, Z., Wang, C., et al.: Sparse r-cnn: End-to-end object detection with learnable proposals. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14454–14463 (2021)
22. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians **71**(3), 209–249 (2021)

23. Xiang, Y., Sun, W., Pan, C., Yan, M., Yin, Z., Liang, Y.: A novel automation-assisted cervical cancer reading method based on convolutional neural network. Biocybernetics and Biomedical Engineering **40**(2), 611–623 (2020)
24. Yi, L., Lei, Y., Fan, Z., Zhou, Y., Chen, D., Liu, R.: Automatic detection of cervical cells using dense-cascade r-cnn. In: Pattern Recognition and Computer Vision. pp. 602–613. Springer (2020)
25. Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L.M., Shum, H.Y.: Dino: Detr with improved denoising anchor boxes for end-to-end object detection. arXiv preprint arXiv:2203.03605 (2022)
26. Zhang, Z., Yao, P., Chen, M., Zeng, L., Shao, P., Shen, S., Xu, R.X.: Scac: A semi-supervised learning approach for cervical abnormal cell detection. IEEE Journal of Biomedical and Health Informatics (2024)
27. Zhou, M., Zhang, L., Du, X., Ouyang, X., Zhang, X., Shen, Q., Luo, D., Fan, X., Wang, Q.: Hierarchical pathology screening for cervical abnormality. Computerized Medical Imaging and Graphics **89**, 101892 (2021)
28. Zhu, X., Li, X., Ong, K., Zhang, W., Li, W., Li, L., Young, D., Su, Y., Shang, B., Peng, L., et al.: Hybrid ai-assistive diagnostic model permits rapid tbs classification of cervical liquid-based thin-layer cell smears. Nature Communications **12**(1), 3541 (2021)