Robust Cervical Abnormal Cell Detection via Distillation from Local-Scale Consistency Refinement

Manman Fei[1], Xin Zhang[1], Maosong Cao[2], Zhenrong Shen[1], Xiangyu Zhao[1], Zhiyun Song[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

Abstract. Automated detection of cervical abnormal cells from Thinprep cytologic test (TCT) images is essential for efficient cervical abnormal screening by computer-aided diagnosis system. However, the detection performance is influenced by noise samples in the training dataset, mainly due to the subjective differences among cytologists in annotating the training samples. Besides, existing detection methods often neglect visual feature correlation information between cells, which can also be utilized to aid the detection model. In this paper, we propose a cervical abnormal cell detection method optimized by a novel distillation strategy based on local-scale consistency refinement. Firstly, we use a vanilla RetinaNet to detect top-K suspicious cells and extract region-of-interest (ROI) features. Then, a pre-trained Patch Correction Network (PCN) is leveraged to obtain local-scale features and conduct further refinement for these suspicious cell patches. We design a classification ranking loss to utilize refined scores for reducing the effects of the noisy label. Furthermore, the proposed ROI-correlation consistency loss is computed between extracted ROI features and local-scale features to exploit correlation information and optimize RetinaNet. Our experiments demonstrate that our distillation method can greatly optimize the performance of cervical abnormal cell detection without changing the detector's network structure in the inference. The code is publicly available at https://github.com/feimanman/Cervical-Abnormal-Cell-Detection.

Keywords: Cervical abnormal cell detection · Consistency learning · Cervical cytologic images

1 Introduction

Cervical cancer is the second most common cancer among adult women. If diagnosed early, it can be effectively treated and cured [19]. Nevertheless, delayed diagnosis of cervical cancer until an advanced stage will have a negative impact on patient prognosis and consume medical resources. Currently, early screening of cervical cancer is recommended worldwide as an effective method to prevent

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 652–661, 2023.
https://doi.org/10.1007/978-3-031-43987-2_63

and treat cervical cancer. Thin-prep cytologic test (TCT) is the most common and effective screening method for detecting cervical abnormal and premalignant cervical lesions [5]. Conventionally it is performed by visually examining the stained cells collected through smearing on a glass slide, and generating a diagnosis report using the descriptive diagnosis method of the Bethesda system (TBS) [15]. Although TCT has been widely used in clinical applications and has significantly reduced the mortality rates caused by cervical cancer, it is still unavailable for population-wide screening [18]. This is partly due to its labor-intensive, time-consuming, and high cost [1]. Therefore, there is a high demand for automated cervical abnormality screening to facilitate efficient and accurate identification of cervical abnormalities.

With the development of deep learning [10], several attempts have been made to identify cervical abnormal cells using convolutional neural networks (CNNs). For example, Cao et al. [2] developed an attention feature pyramid network (AttFPN) for automatic abnormal cervical cell detection in cervical cytopathological images to assist pathologists in making more accurate diagnoses. Chen et al. [3] proposed a new framework that decomposes tasks and compares cells for cervical lesion cell detection. Liang et al. [11] proposed to explore contextual relationships to boost the performance of cervical abnormal cell detection. Lin et al. [22] presented an automatic cervical cell detection approach based on the Dense-Cascade R-CNN. It is worth mentioning that all of the aforementioned detection methods inevitably produce false positive results, which should be further refined by pathologists for manual checking or classification models established for automatic screening. To solve this problem, Zhou et al. [23] proposed a three-stage method including cell-level detection, image-level classification, and case-level diagnosis obtained by an SVM classifier. Zhu et al. [24] developed an artificial intelligence assistive diagnostic solution, which integrated YOLOv3 [16] for detection, Xception, and Patch-based models to boost classification.

Although the above-mentioned attempts can improve the screening performance significantly, there are several issues that need to be addressed: 1) Object detection methods often require accurate annotated data to guarantee performance with robustness and generalization. However, due to legal limitations, the scarcity of positive samples, and especially the subjectivity differences between cytopathologists for manual annotations [20], it is likely to generate noisy samples that affect the performance of the detection model. 2) Conventional object detection methods intend to directly extract the feature from the object area to locate and classify the object simultaneously. However, in clinical practice pathologists usually examine the target cells by comparing them to the surrounding cells to determine whether they are abnormal. Therefore, the visual feature correlations between the target cells and their surroundings can provide valuable information to aid the screening process, which also needs to be utilized when designing the cervical abnormal cell detection network.

To address these issues, we propose a novel method for cervical abnormal cell detection using distillation from local-scale consistency refinement. Inspired by knowledge distillation, we construct a pre-trained Patch Correction Network

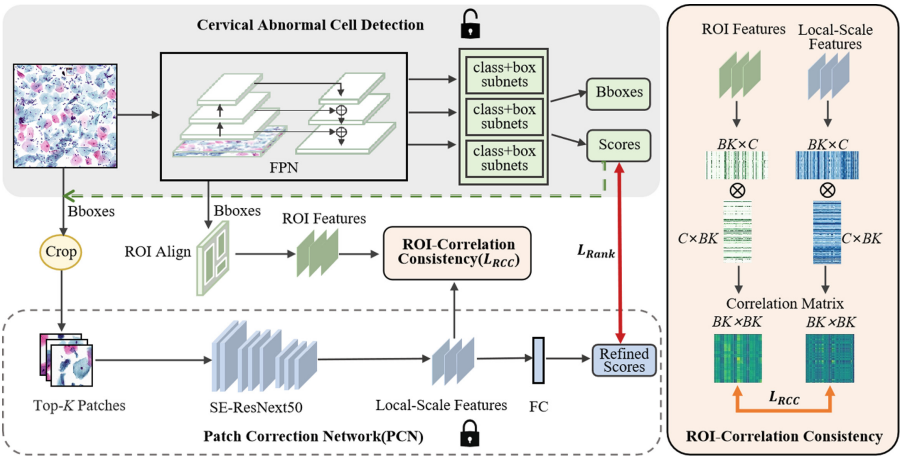

Fig. 1. The overview of our proposed framework, where PCN provides refined scores and local-scale features. Ranking Loss(L_{Rank}) is proposed to optimize the RetinaNet proposal classifier on detection scores and refined scores. And we incorporate consistency learning between ROI features and Local-scale features by RCC Loss(L_{RCC}). Note that PCN is frozen and the cervical abnormal cell detection is updated by L_{Rank} and L_{RCC} during training.

(PCN), which is designed to exploit the supervised information from the PCN to reduce the impact of noisy labels and utilize the contextual relationships between cells. In our approach, we begin by utilizing RetinaNet [12] to locate suspicious cells and crop the top-K suspicious cells into patches. Then we feed them into the PCN to obtain classification scores and propose a ranking loss to refine the classifier of the detection network by correcting the score of the detection model. In addition, we propose an ROI-Correlation Consistency (RCC) loss between ROI features and local-scale features from the PCN, which encourages the detector to explore the feature correlations of the suspicious cells. Our proposed method achieves improved performance during inference without changing the detector structure.

2 Method

The proposed framework is shown in Fig. 1, which includes cervical abnormal cell detection and the PCN. Concerning the huge size of the Whole Slide Image (WSI) and the infeasibility to handle a WSI scan for detection, we crop the WSI into images with the size of 1024 × 1024 as input to the detection. Firstly, We choose RetinaNet as our cervical abnormal cell detection, which uses a Feature Pyramid Network (FPN) backbone and attaches two subnetworks to obtain bounding boxes and classification scores. We implement the detection to locate the suspicious lesion cervical cells and extract the top K patches from the original image. Besides, we add the ROI Align layer [17] to the output of the FPN

and generate ROI features. Then these patches are fed into the PCN to obtain refined scores and local-scale features. Subsequently, our ranking loss is employed to correct the score of the detection, followed by the RCC loss to capture the contextual relationships between the extracted cells for further optimizing the detection model. The distillation process involves leveraging the learned knowledge and expertise from the PCN to refine the detection results of RetinaNet.

2.1 Patch Correction Network(PCN)

In Fig. 1, the detection can automatically locate the suspicious cervical abnormal cells by providing their bounding boxes with the confidence scores. Due to the intrinsic architecture limitation of the detection and incomplete annotations, the confidence scores output by the RetinaNet may not be accurate, so we need another classification model to regrade the representative patches. Our framework leverages a local-scale classification refinement mechanism to guide the training of the detection model. We adopt SE-ResNext-50 [8] as the PCN, which has demonstrated its effectiveness in this field. The PCN is employed to refine and enhance the RetinaNet proposal classifier, which is trained from a large number of patches collected in advance with more excellent classification performance.

More specifically, the input image is processed by the base detector $F_d(\cdot)$ firstly to obtain the primary proposal information. The proposed PCN $F_c(\cdot)$ takes the top-K patches as inputs, which are cropped from original images according to the proposal location, denoted as $I_p = Cr(I, p)$, where $Cr(\cdot)$ denotes the crop function, I and p denote input image and proposal boxes predicted by $F_d(\cdot)$, respectively. Similar to the RetinaNet proposal classifier in $F_d(\cdot)$, the PCN $F_c(\cdot)$ outputs a classification distribution vector s_c. Therefore, the proposed PCN $F_c(\cdot)$ can be represented as:

$$s_c = F_c(I_p) = F_c(Cr(I, p)). \tag{1}$$

The key idea is to augment the base detector $F_d(\cdot)$ with the PCN $F_c(\cdot)$ in parallel to enhance the proposal classification capability.

2.2 Classification Ranking Loss

Due to the inaccurate confidence scores output by RetinaNet, false positive cells are inevitable after detection. Hence, a good correction network is required to generate more precise scores. In this work, the suspicious ranking of the detected patches is updated by applying PCN to them. The detector is optimized by inter-scale pairwise ranking loss. Specifically, the ranking loss is given by:

$$L_{Rank}(s_d, s_c) = max\left\{0, s_c - s_d + margin\right\}, \tag{2}$$

where s_c is the classification refinement score and s_d is the detection score, which enforces $s_d > s_c + margin$ in training. We set $margin = 0.05$. Such a

design can enable RetinaNet to take the prediction score as references, and utilize refined scores from PCN to obtain more confident predictions. The ranking loss optimizes the detection to generate higher confidence scores than the previous prediction, thereby suppressing false positives and enabling the detection network to better distinguish between positive and negative cells.

2.3 ROI-Correlation Consistency (RCC) Learning

In order to solve the problem of mismatched inputs to the detection and classification models, we add the ROI Align layer to the output of the FPN. However, for cervical abnormal cell detection, normal and abnormal cells may have very similar appearances, which might not be sufficient for conducting effective differentiation. In clinical practice, to determine whether a cervical cell is normal or abnormal, cytopathologists usually compare it to the surrounding reference cells. Therefore, we studied the correlation between the top K ROIs to help more accurate classification of abnormal cells.

Based on the consistency strategy [14], which enhances the consistency of the intrinsic relation among different models, we propose ROI-correlation consistency, which regularizes the network to maintain the consistency of the semantic relation between patches under ROI features and local-scale features, and thereby encourage the detector to explore the feature interaction between cells from the extracted patches to improve the network performance.

We model the structured relation among different patches with a case-level Gram Matrix [6]. Given an input mini-batch with B samples, where B denotes the batch size. And each sample undergoes the ROI Align layer to obtain the top K ROIs, we denote the activation map of ROIs as $F^R \in \mathbb{R}^{B\times K\times H\times W\times C}$, where H and W are the spatial dimension of the feature map, and C is the channel number. We set $K = 10$, $H = 7$, $W = 7$, $C = 256$. We average pooling the feature map F^R along the spatial dimension and reshape it into $A^R \in \mathbb{R}^{BK\times C}$, and then the Case-wise Gram Matrix $G^R \in \mathbb{R}^{BK\times BK}$ is computed as:

$$G^R = A^R \cdot (A^R)^T, \tag{3}$$

where G_{ij} is the inner product between the vectorized activation map A_i^R and A_j^R, whose intuitive meaning is the similarity between the activations of i_{th} ROI and j_{th} ROI within the input mini-batch. The final ROI relation matrix R^R is obtained by conducting the L2 normalization for each row G_i^R of G^R, which is expressed as:

$$R^R = \left[\frac{G_1^R}{\left\|G_1^R\right\|_2}, \cdots, \frac{G_{BK}^R}{\left\|G_{BK}^R\right\|_2}\right]^T. \tag{4}$$

The proposed PCN $F_c(\cdot)$ takes the $B\times K$ proposals of box regressor as inputs, we denote the local-scale feature map by PCN as $F^C \in \mathbb{R}^{B\times K\times H'\times W'\times C}$, and set $H' = 56$, $W' = 56$. We perform average pooling on the feature map F^C across the spatial dimension and then reshape it into $A^C \in \mathbb{R}^{BK\times HWC}$, the Case-wise

Gram Matrix $G^C \in \mathbb{R}^{BK \times BK}$ and the final relation matrix R^C are computed as:

$$G^C = A^C \cdot (A^C)^T, \tag{5}$$

$$R^C = \left[\frac{G_1^C}{\left\| G_1^C \right\|_2}, \cdots, \frac{G_{BK}^C}{\left\| G_{BK}^C \right\|_2} \right]^T. \tag{6}$$

The RCC requires the correlation matrix to be stable under ROI features and local-scale features to preserve the semantic relation between patches. We then define the proposed RCC loss as:

$$L_{RCC} = \sum \frac{1}{BK} \left\| R^C(X) - R^R(X) \right\|_2^2, \tag{7}$$

where X is the proposals from the sampled mini-batch, $R^C(X)$ and $R^R(X)$ are the correlation matrices computed on X under different network. By minimizing L_{RCC} during the training process, the network could be enhanced to capture the intrinsic relation between patches, thus helping to extract additional semantic information from cells.

2.4 Optimization

To better optimize the Retinanet detector in a reinforced way, we take the following training strategy, which consists of three major stages. In the first stage, we collect images with doctors' labels for training and initialized the detection net. In the second stage, we train PCN with cross-entropy loss until convergence. In the last stage, we freeze the PCN and optimize the detector. The detector is optimized using the total objective function, which is written as follows:

$$L_{total} = L_{cls} + L_{reg} + \alpha L_{Rank} + \beta L_{RRC}, \tag{8}$$

where L_{cls} and L_{reg} are the ordinary detection loss for each detection head in RetinaNet. L_{cls} is a Cross-Entropy loss for classification and L_{reg} is a Smooth-L_1 loss for bounding box regression. L_{Rank} is the classification ranking loss,L_{RRC} is the RCC loss. α and β are hyper-parameters that denote the different weights of loss. During inference, only the optimized detector is used to output the final detection results without any additional modules.

3 Experimental Results

3.1 Dataset and Experimental Setup

Dataset. For cervical cell detection, our dataset includes 3761 images of 1024×1024 pixels cropped from WSIs. Our private dataset was collected and quality-controlled according to a standard protocol involving three pathologists: A, B, and C. Pathologist A had 33 years of experience in reading cervical cytology

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

Method	AP	AP.5	AP.75	AR
Sparse R-CNN [21]	41.8	72.1	42.4	66.8
Deformable-DETR [25]	40.3	72.6	38.6	64.3
YoloV8 [4]	43.6	74.6	44.0	58.3
Faster R-CNN [17]	43.6	77.0	43.0	58.8
Cascade RRAM and GRAM [11]	44.6	77.5	47.7	60.0
RetinaNet [12]	45.7	81.3	46.2	58.8
Proposed method	**51.1**	**86.6**	**54.3**	**62.5**

images, while pathologists B and C had 10 years of experience each. Initially, the images were randomly assigned to pathologist B or C for initial labeling. Later, the assigned pathologist's annotations were reviewed and verified by the other pathologist. Any discrepancies found were checked and re-labeled by pathologist A. These images were divided into the training set and the testing set according to the ratio of 9:1. We also collect a new dataset of 5000 positive and negative 224×224 cell patches to train the PCN.

Implementation Details. The backbone of the suspicious cell detection network is RetinaNet with ResNet-50 [7]. The backbone of the pre-trained patch classification network is SE-ResNeXt-50. All parameters are optimized by Adam [9] with an initial learning rate of 4×10^{-5}. We set α to 0.25 and β to 1 during training. The model is implemented by PyTorch on 2 Nvidia Tesla P100 GPUs. We conduct a quantitative evaluation using two metrics: the COCO-style [13] average precision (AP) and average recall (AR). We calculate the average AP over multiple IoU thresholds from 0.5 to 0.95 with a step size of 0.05, and individually evaluated AP at the IoU thresholds of 0.5 and 0.75 (denoted as AP.5 and AP.75), respectively.

3.2 Evaluation of Cervical Abnormal Cell Detection

Comparison with SOTA Methods. We compare the performance of our proposed method against known methods for cervical lesion detection as well as representative methods for object detection. Table 1 presents the results, from which several observations can be drawn. (1) Among the models for object detection, Retinanet is generally superior to the other models. (2) Based on Retinanet, our method improves the detection performance significantly, especially AP.5 shows great performance improvement. This confirms the necessity and effectiveness of introducing the classification ranking and ROI-correlation consistency schemes for cervical lesion detection.

Ablation Study. We also perform an ablation study to further evaluate the contributions of each part in our method. Table 2 reports the detailed ablation results, from which several observations can be drawn. (1) Compared with the baseline model, Retinanet, our classification ranking loss achieves considerably

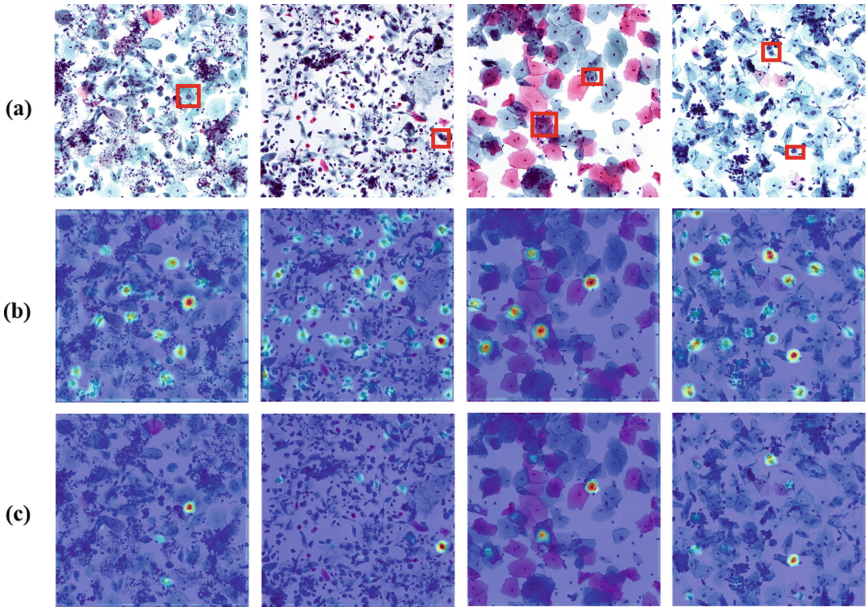

Fig. 2. Feature map visualization of RetinaNet and our method. (a) shows input images with ground-truth annotations. (b) shows feature maps from RetinaNet. (c) shows feature maps of our proposed method.

Table 2. Performance of ablation study for our local-scale consistency refinement.

Method	AP	AP.5	AP.75	AR
Baseline	45.7	81.3	46.2	58.8
+Ranking Loss	47.8	83.2	49.0	59.7
+RCC Loss	47.4	82.7	46.1	59.2
+Ranking Loss and RCC Loss	**51.1**	**86.6**	**54.3**	**62.5**

better performance, especially in AP.75, with an improvement of 2.8. (2) The RCC loss is also effective for learning better feature representations and distinguishing them well, and the AP is improved by 1.7. (3) With both ranking loss and RCC loss, our method has the best performance, which surpasses the baseline model by a large margin, validating the effectiveness of our method.

In addition, to further show the effectiveness of our method, we visualize the feature maps of Retinanet and the proposed method in Fig. 1. Those feature maps are from the Conv3 stages of the class-subnet backbone. Specifically, we sum and average the features in the channel dimension, and upsample them to the original image size. As shown in Fig. 2, our method can really learn better feature representations for abnormal cells, with the help of our proposed classification ranking refinement and ROI-correlation consistency learning. By model

learning, our method can gradually enhance the features of abnormal cell regions while repressing noise or other suspicious but non-lesion regions.

4 Conclusion

In this paper, we integrate a distillation strategy that uses the knowledge learned from the pre-trained PCN to guide the training of the detection model to minimize the effects of noisy labels and explore the feature interaction between cells. Our method constructs RetinaNet with the PCN module which provides the refined scores and local-scale features of extracted patches. Specifically, we propose the ranking loss by utilizing refined scores to optimize the RetinaNet proposal classifier by reducing the impact of noisy labels. In addition, the ROI features generated by the detector and local-scale features from the PCN are used for correlation consistency learning, which explores the extracted cells' relationship. Our work can achieve better performance without adding new modules during inference. Experiments demonstrate the effectiveness and robustness of our method on the task of cervical abnormal cell detection.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (No. 62001292).

References

1. Bengtsson, E., Malm, P.: Screening for cervical cancer using automated analysis of pap-smears. In: Computational and Mathematical Methods in Medicine 2014 (2014)
2. Cao, L., et al.: A novel attention-guided convolutional network for the detection of abnormal cervical cells in cervical cancer screening. Med. Image Anal. **73**, 102197 (2021)
3. Chen, T., et al.: A task decomposing and cell comparing method for cervical lesion cell detection. IEEE Trans. Med. Imaging **41**(9), 2432–2442 (2022)
4. Contributors, M.: Mmyolo: Openmmlab yolo series toolbox and benchmark (2022)
5. Davey, E., et al.: Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review. Lancet **367**(9505), 122–132 (2006)
6. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
7. 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)
8. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
9. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
10. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)

11. Liang, Y., et al.: Exploring contextual relationships for cervical abnormal cell detection. arXiv preprint arXiv:2207.04693 (2022)
12. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
13. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
14. Liu, Q., Yu, L., Luo, L., Dou, Q., Heng, P.A.: Semi-supervised medical image classification with relation-driven self-ensembling model. IEEE Trans. Med. Imaging **39**(11), 3429–3440 (2020)
15. Nayar, R., Wilbur, D.C. (eds.): The Bethesda System for Reporting Cervical Cytology. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-11074-5
16. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
17. 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 28 (2015)
18. Saslow, D., et al.: American cancer society, American society for colposcopy and cervical pathology, and American society for clinical pathology screening guidelines for the prevention and early detection of cervical cancer. Am. J. Clin. Pathol. **137**(4), 516–542 (2012)
19. Schiffman, M., Castle, P.E., Jeronimo, J., Rodriguez, A.C., Wacholder, S.: Human papillomavirus and cervical cancer. Lancet **370**(9590), 890–907 (2007)
20. Stoler, M.H., Schiffman, M., et al.: Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ascus-lsil triage study. JAMA **285**(11), 1500–1505 (2001)
21. Sun, P., et al.: SparseR-CNN: end-to-end object detection with learnable proposals. arXiv preprint arXiv:2011.12450 (2020)
22. Yi, L., Lei, Y., Fan, Z., Zhou, Y., Chen, D., Liu, R.: Automatic detection of cervical cells using dense-cascade R-CNN. In: Peng, Y., et al. (eds.) PRCV 2020. LNCS, vol. 12306, pp. 602–613. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60639-8_50
23. Zhou, M., et al.: Hierarchical pathology screening for cervical abnormality. Comput. Med. Imaging Graph. **89**, 101892 (2021)
24. Zhu, X., et al.: Hybrid ai-assistive diagnostic model permits rapid tbs classification of cervical liquid-based thin-layer cell smears. Nat. Commun. **12**(1), 3541 (2021)
25. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159 (2020)