Medical Image Analysis 73 (2021) 102197

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

A novel attention-guided convolutional network for the detection of abnormal cervical cells in cervical cancer screening

Lei Cao[a], Jinying Yang[b], Zhiwei Rong[a], Lulu Li[a], Bairong Xia[c], Chong You[d], Ge Lou[e], Lei Jiang[f], Chun Du[g], Hongxue Meng[g], Wenjie Wang[a], Meng Wang[a], Kang Li[a,*], Yan Hou[h]

[a] Department of Biostatistics, School of Public Health, Harbin Medical University, Harbin 150081, China
[b] Department of Pathology, Heilongjiang Maternal and Child Health Care Hospital, Harbin 150001, China
[c] The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Anhui province cancer hospital, Hefei 230031, Anhui, China
[d] Beijing International Center for Mathematical Research, Peking University, Beijing 100191, China
[e] Department of Gynecology Oncology, Harbin Medical University Cancer Hospital, Harbin 150081, P.R. China
[f] Department of Pathology, the Second Affiliated Hospital of Harbin Medical University, Harbin 150001, China
[g] Department of Pathology, Precision Medical Center, Harbin Medical University Cancer Hospital, Harbin 150081, China
[h] Department of Biostatistics, School of Public Health, Peking University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 11 November 2020
Revised 10 June 2021
Accepted 23 July 2021
Available online 10 August 2021

MSC:
41A05
41A10
65D05
65D17

Keywords:
Detection
Cervical cancer
Cytological image
Attention mechanism
Convolutional neural network
Deep learning
Clinical knowledge

ABSTRACT

Early detection of abnormal cervical cells in cervical cancer screening increases the chances of timely treatment. But manual detection requires experienced pathologists and is time-consuming and error prone. Previously, some methods have been proposed for automated abnormal cervical cell detection, whose performance yet remained debatable. Here, we develop an attention feature pyramid network (AttFPN) for automatic abnormal cervical cell detection in cervical cytology images to assist pathologists to make a more accurate diagnosis. Our proposed method consists of two main components. First, an attention module mimicking the way pathologists reading a cervical cytology image. It learns what features to emphasize or suppress by refining extracted features effectively. Second, a multi-scale region-based feature fusion network guided by clinical knowledge to fuse the refined features for detecting abnormal cervical cells at different scales. The region proposals in the multi-scale network are designed according to the clinical knowledge about size and shape distribution of real abnormal cervical cells. Our method, trained and validated with 7030 annotated cervical cytology images, performs better than the state of art deep learning-based methods. The overall sensitivity, specificity, accuracy, and AUC of an independent testing dataset with 3970 cervical cytology images is 95.83%, 94.81%, 95.08% and 0.991, respectively, which is comparable to that of an experienced pathologist with 10 years of experience. Besides, we further validated our method on an external dataset with 110 cases and 35,013 images from a different organization, the case-level sensitivity, specificity, accuracy, and AUC is 91.30%, 90.62%, 90.91% and 0.934, respectively. Average diagnostic time of our method is 0.04s per image, which is much quicker than the average time of pathologists (14.83s per image). Thus, our AttFPN is effective and efficient in cervical cancer screening, and improvement of clinical workflows for the benefit of potential patients. Our code is available at https://github.com/cl2227619761/TCT_Detection.

1. Introduction

Cervical cancer is the fourth most common malignancy among women globally, with an estimated 570,000 new cases in 2018, and the burden is projected to continue to increase to nearly 700,000 cases and 400,000 deaths in 2030 (Bray et al., 2018). Fortunately, it is also a readily curable disease, if diagnosed and treated early (Boulet et al., 2008; Harlan et al., 1991; Parkin and Bray, 2006; Woodman et al., 2007). The American College of Obstetricians and Gynecologists (ACOG) recommends that women aged 21–69 should undergo a cervical cancer screening every three years to find abnormal cells in the cervix before they become cancer (of Obstetricians et al., 2010). In recent decades, the widespread application of cervical cancer screening has led to early detection and timely treatment. Actually, according to the U.S. Preventive Services Task

* Corresponding author.
E-mail addresses: likang@ems.hrbmu.edu.cn (K. Li), houyan@bjmu.edu.cn (Y. Hou).

https://doi.org/10.1016/j.media.2021.102197
1361-8415/

Force, the effective screening can reduce the incidence of cervical cancer by at least 60% (Landy et al., 2016). However, there are many obstacles to cervical cancer screening, especially in resource-poor regions (Carpenter and Davey, 1999; Papillo et al., 1998; Pradhan et al., 2016). For example, scarcity of highly trained pathologists hinders effective screening for large female populations. In China, for example, there are almost 130,000 new cases of cervical cancer annually, accounting for 28% of all cases diagnosed worldwide every year (Di et al., 2015; Jiang et al., 2017; Malagón, 2019), but there are only about 10,000 licensed pathologists(Huang et al., 2018). Moreover, the cervical cytology images reading is an extremely labor intensive, complex and tedious task (Bhan et al., 2018). Furthermore, reports from the manual reading of cervical cytology images demonstrate large variation across pathologists with various working experience. For instance, the estimates for false-negative rates vary from 2% to 55% according to (Branca and Longatto-Filho, 2015) and (van der Graaf and Vooijs, 1987). Additionally, and more importantly, since abnormal cervical cells usually account for only a small portion of all the sample cells in cytology image for pathologists to investigate, this leads to unnecessary waste of medical resources.

Several studies have investigated the possibility of automatic cervical cancer detection. Bhan et al. (Bhan et al., 2018), for instance, proposed an automatic detection method of nucleus of cervical cell, using geometrical features like perimeter and area etc. to detect whether the cell was cancerous or normal. Li et al. (Li et al., 2019) presented a region detection and classification method based on multi-semantic label, combined with morphological information analysis to assist pathologists in the early diagnosis of cervical cancer. While Zhang et al. (Zhang and Liu, 2004) detected suspected cervical cells using SVM method based on extracting various types of image features including pixel intensity and wavelet features. Despite the effectiveness of these studies, the automatic detection of abnormal cervical cells in screening test is still a challenging task. This is because those machine learning methods leverage traditional handcrafted features designed for a specific task, which are usually not robust enough. Later, deep learning methods such as convolutional neural network (CNN) are applied to medical images (Esteva et al., 2017; Sahlsten et al., 2019; Wang et al., 2017), which can learn hierarchical features of the input without any preprocessing steps such as segmentation or handcrafted feature extraction. However, most CNN models are constructed to do classification tasks, which are not suitable to be used for cell, organ or lesion region localization in medical images. For example, during cervical cancer screening, apart from the cervical cytology image classification result (normal vs. abnormal), abnormal cervical cells location in the image is also an important clue. To address this issue, Tsumori et al (TSUMORI et al., 2019), Li et al (Li et al., 2019), Tobias et al (Tobias et al., 2020), Zhang et al (Zhang et al., 2016), Mahmood et al (Mahmood et al., 2020), Chen et al (Chen et al., 2019), Lin et al (Lin et al., 2019), and Lu et al (Lu et al., 2018) introduced Faster R-CNN (Ren et al., 2015) to medical image detection. The focal areas of these studies include cell detection in microscopy images, fetal head detection in ultrasound images, teeth detection in dental periapical films, and metastatic lymph nodes detection in MR images. The Faster R-CNN (Ren et al., 2015) is a deep learning object detection method and it mainly contains three parts, CNN backbone module to extract useful features from training data, region proposal network (RPN) module to generate high-quality region proposals (ROIs), and Fast R-CNN module to do classification and detection task using those ROI features as inputs.

Despite the effectiveness of Faster R-CNN in various biomedical tasks (Li et al., 2019; TSUMORI et al., 2019; Tobias et al., 2020; Zhang et al., 2016; Mahmood et al., 2020; Chen et al., 2019; Lin et al., 2019; Lu et al., 2018; Liu et al., 2019), abnormal cervical cells detection in cervical cancer screening is still a very challenging task and needs urgent attention. To the best of our knowledge, there are only two studies constructing Faster R-CNN for the cervical cells detection. Tsumori et al (TSUMORI et al., 2019) used Faster R-CNN to detect and classify observed cells into different grade of malignancy. While Li et al (Li et al., 2019) adopted Faster R-CNN to detect and classify cervical exfoliated cells in the early diagnosis of cervical cancer. Both of their methods showed effectiveness in the early diagnosis of cervical cancer. However, some challenges still hinder the efficiency of these methods. For example, cells of various sizes often appear in microscopy images, and original Faster R-CNN cannot deal with the scale variability of different cells, which may lead to missed diagnosis. Also, clinical-knowledge in cervical cancer screening, such as the size and shape of real abnormal cervical cells, despite being important clue in diagnosis, was ignored in their models.

The current study proposed a novel attention feature pyramid network (AttFPN) for automatic detection and classification of abnormal cervical cells in Thinprep Pap Test images. Inspired by feature pyramid network (FPN) (Lin et al., 2017) and attention mechanism (Corbetta and Shulman, 2002; Itti et al., 1998; Rensink, 2000), we developed an attention module and integrated it with multi-scale feature fusion to refine feature representations more effectively. In addition, we designed the sizes and aspect ratios of anchor boxes in RPN according to the real abnormal cervical cells size distribution, instead of the predefined settings commonly used in other related deep learning methods. This process could bring the prior clinical-knowledge into our AttFPN method. To evaluate the performance of our presented model, we conducted a reader study by human expert pathologists to create the ground-truth annotation from the collected images. Experimental results showed that our model outperformed the other related deep learning models as well as pathologists with 10-year experience. In summary, there were five main contributions in our paper:

- We proposed a novel deep learning method named attention feature pyramid network (AttFPN) for abnormal cervical cell detection, which outperformed the other related deep learning methods and was comparable to experienced pathologists with 10-year experience on an independent testing dataset.
- We developed an attention module, and integrated it with multi-scale feature fusion to refine representations more effectively, which can also be used as a plug-and-play module in other similar medical detection tasks.
- We introduced clinical-knowledge into our AttFPN in an easy-to-use manner. The region proposals in detection network were constrained by the prior knowledge about size and shape distributions of real abnormal cervical cells.
- We created a large dataset consisting of 70,552 cervical cytology images collected during cervical cancer screening. The dataset along with its annotations will be made available to other researchers if requested.
- We built a web application based on our AttFPN model. Cervical cytology images could be uploaded to do online prediction. This may give insight to bridge the health resources gap in some resource-poor regions. The temporary url is https://www.linkermed.com/TCT.

2. Materials and methods

2.1. Data preparation and annotation

Two Thinprep Pap Test slides datasets with a total of 325 cases were used in our study. One of them was supplied by the Department of Pathology, Heilongjiang Maternal and Child Health Hos-

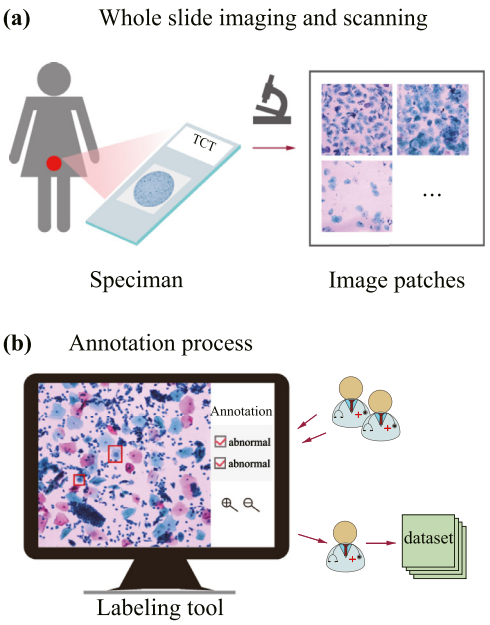

Fig. 1. An illustration of data preparation and annotation. (a) Image patches for training and testing were generated by digitalizing glass slides. (b) A labeling tool was used to manually draw bounding boxes around abnormal cells.

Table 1
Anchors size and aspect ratios at each feature map location.

Feature scale	N_2	N_3	N_4	N_5	N_6
anchor size	32	64	128	256	512
aspect ratios	(1/2,1,2)	(1/2,1,2)	(1/2,1,2)	(1/2,1,2)	(1/2,1,2)

Table 2
Comparison of the detection performance of the proposed method with related methods over average precision (AP) (%) on testing dataset.

Backbone	Faster R-CNN	Faster R-CNN+FPN	AttFPN
ResNet-50	69.98	70.36	**71.75**
ResNet-101	67.72	68.47	**70.80**
ResNet-152	66.44	67.59	**71.02**
DenseNet-121	69.85	71.09	**75.54**
DenseNet-161	67.99	71.60	**73.35**
DenseNet-169	68.91	72.72	**75.09**

pital (HMCHH dataset), including 215 cases (129 abnormal cases and 86 normal cases), acquired during October 2018 to May 2019, which was used for model training. The other was collected from Harbin Medical University Cancer Hospital (HMUCH dataset), including 110 cases (46 abnormal cases and 64 normal cases), which was used for external validation. Each slide was digitalized and divided into 333 non-overlapping patches ($2048 \times 2048 pixel$) at 20x objective magnification, using the Olympus BX53 optical microscope, as shown in Fig. 1a. The image patches with low amount of information were removed; i.e., patches covered by background or blurred patches. This process generated a total of 70,552 patches, among which 22,848 patches (8,037 from the HMCHH dataset, 14,811 from the HMUCH dataset) were from abnormal patients and 47,704 patches (27,502 from the HMCHH dataset, 20,202 from the HMUCH dataset) were from normal patients. The detailed dataset description was shown in supplementary Fig. 1, Table 1 and Table 2.

To get ground-truth annotation, we performed a reader study on those 8037 patches from the HMCHH dataset. Three pathologists were involved in preparing the annotation for this study, with the aim of making exhaustive annotations of abnormal cervical cells in each patch (Fig. 1b). The abnormality or normality of a cervical cell was defined according to the ACOGs guidance (Randel, 2013). We referred to these readers as A, B, and C; reader A had about 33 years of experience in reading cervical cytology images, while reader B and C had about 10 years of experience. The readers draw bounding boxes around abnormal cells in the image patches using the Colabeler tool http://www.colabeler.com/, as shown in Fig. 1b.

A three-step annotation process was performed as follows, the initial labeling step, the verification step, and the final check step. An image was firstly randomly assigned to reader B or C. Once the labeling was finished, the image and annotation were then passed on to another reader for review. Finally, the annotations were checked by reader A.

2.2. The attfpn model

In this work, the cervical cancer detection task was defined as a three-stage task (Fig. 2), where a multi-scale pyramid network integrated with attention module was first designed to automatically detect abnormal cervical cells in an image patch, and then we obtained the image-level classification results based on the detected abnormal cervical cells with corresponding probability; finally, we summarized the classification outputs of all image patches of the target case, and determined the case-level result. Details of our approach were illustrated in Fig. 2, Fig. 3 and the following subsections.

2.2.1. Multi-scale detection network with attention module

The original Faster R-CNN requires two stages: (1) the region proposal network (RPN) is firstly used to generate proposals, and (2) the Fast R-CNN module is then designed to predict the final ROI (Ren et al., 2015). However, since Faster R-CNN only uses mono-scale feature representations, it may fail to detect objects with varying sizes, a very practical challenge that should be tackled in our task of abnormal cervical cells detection. Specifically, clinical prior knowledge shows that cervical cells have a variety of sizes. Fig. 4 is the statistical description of the annotated abnormal cervical cells included in our study, from which we can observe that some cells could be very small (e.g., size of 32×32 pixel, Fig. 4b) while some others could be large (e.g., size of 300×300 pixel, Fig. 4c). In order to robustly detect those cells in different sizes, an intuitive way is to make use of the multi-scale features learned by successive convolutional layers.

Thus, we combined the idea of feature pyramid network (FPN) (Lin et al., 2017) with Faster R-CNN to learn multi-scale region proposals for abnormal cervical cell detection. Besides, in order to refine feature representations more effectively, we further integrated an attention module with our multi-scale network. The architecture of our proposed multi-scale detection network with attention module is presented in Fig. 3. Briefly, relatively large cells and small cells would be detected in our network using feature maps produced by the high-level layers and low-level layers, respectively; meanwhile, the salient feature maps would be emphasized using our attention module. In the rest of this subsection, we would introduce in detail the architecture of our model, and a clinical-knowledge-based strategy for defining detection anchors.

1) Network architecture: as illustrated in Fig. 3, we adopted the general structure of DenseNet-169 (Huang et al., 2017) as the final backbone of our detection network. The DenseNet contains four dense blocks (i.e., stage1 to stage4) that use all the preceding feature maps in its output (supplementary Table 3). Each of the first three dense blocks is followed by a transition layer to down-sample the feature maps by a scale factor of 2. So, similar to that in FPN, the backbone here forms a bottom-up pathway,

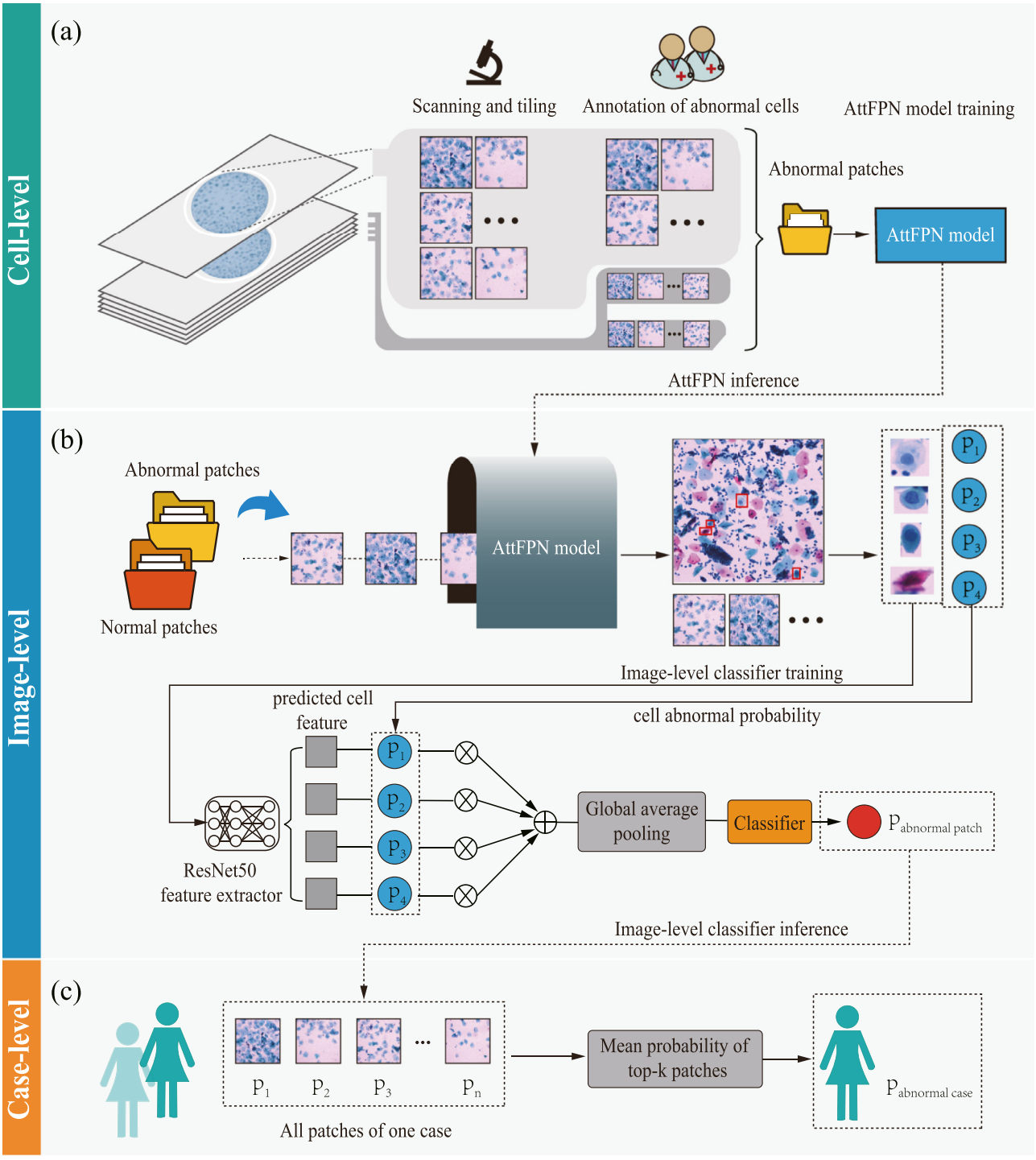

Fig. 2. The three-stage process of our framework. (a) Cell-level object detection. The abnormal patches annotated with abnormal cells are used for the training of AttFPN model, which focuses on detecting abnormal cells in a given image patch. (b) Image-level classification. By using a classification network, the normal/abnormal probability for each image patch is predicted based on detected abnormal cells and their corresponding abnormal probability. (c) Case-level classification. Summarizing the predicted results of all patches from one case to derive the final case-level classification report for the case.

which computes multi-scale feature representations with progressively increased abstraction but decreased spatial resolution. The specific structure of dense blocks was kept the same as that in the original paper.

To detect cells with various sizes, and to refine features more effectively, we designed a top-down pathway integrated with attention modules. Firstly, we selected the output of the 1×1 conv layer of each transition layers as the multi-scale features, namely Conv_2, Conv_3, Conv_4. For an easier expression, we abbreviated them to C_2, C_3, C_4. Since there is no transition layer after the stage 4 (supplementary Table 3), we directly selected the output of stage4, denoted as C_5. And P_2, P_3, P_4, P_5 were generated from C_2, C_3, C_4, C_5 through four 1×1 conv layers, respectively (Fig. 3a). This operation reduced the number of feature map channels to C=256. Secondly, according to the way pathologists reading images, not all cervical cells are of equal importance. Thus, different empha-

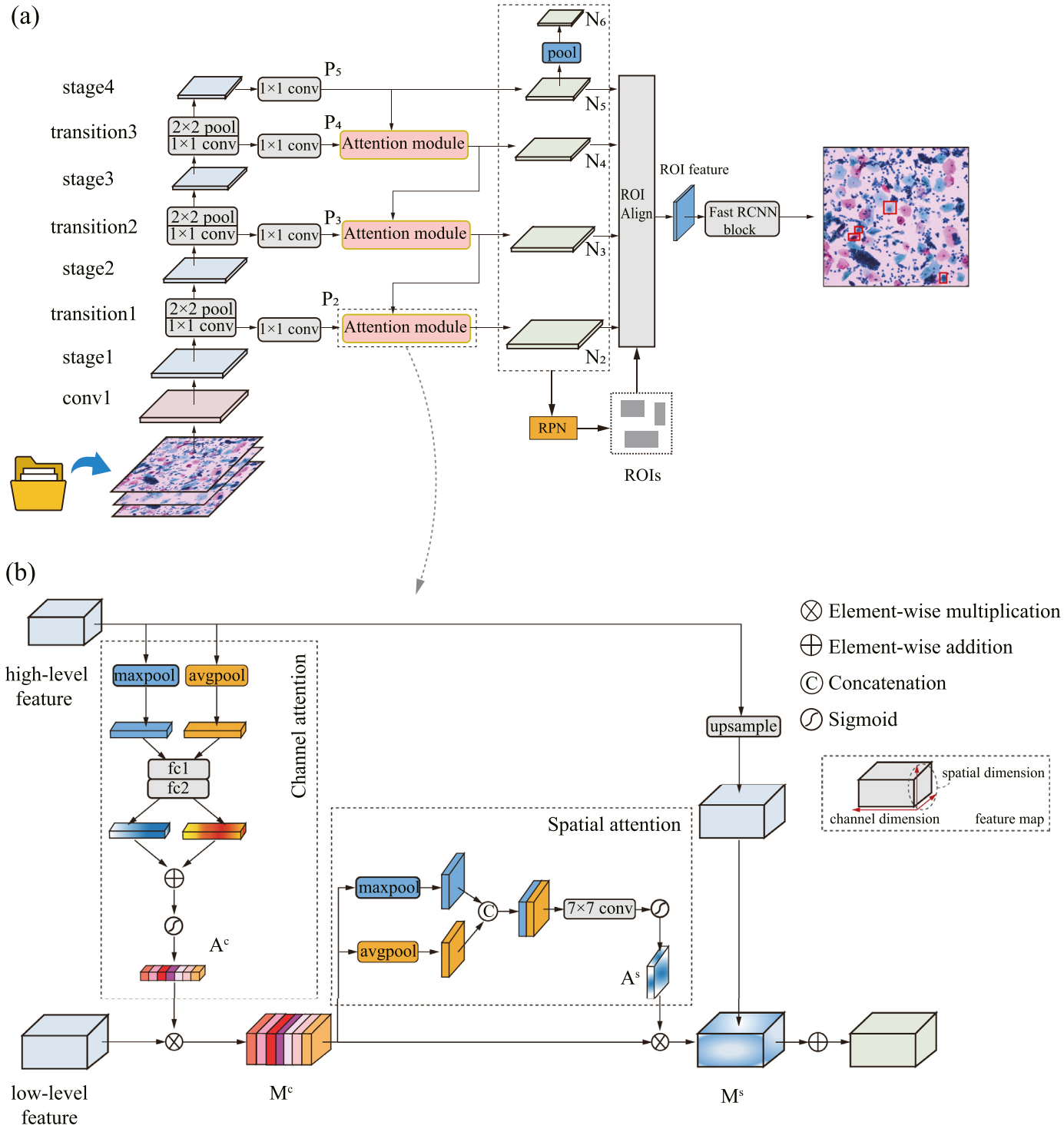

Fig. 3. Overview of the attention feature pyramid network (AttFPN) presented in our study. (a) The AttFPN that combines attention mechanism with multi-scale feature fusion, the backbone network is DenseNet-169. In the detection part, input images are fed into the model, and the model learns on the inputs. After the training procedure is done, given an input image, abnormal cervical cells can be detected through the trained model. Their probability of being abnormal can be obtained. Then in the classification part, the maximum probability of those detected cells is considered as the images probability of being abnormal. (b) The detailed structure of our Attention module. The low-level features are refined along the channel dimension through channel attention, then the refined features are fed into the spatial attention to be refined again along the spatial dimension. Finally, the refined low-level features are fused with high-level features.

sis should be put on different cells. Inspired by this clinical fact, from P_2 to P_5, we injected an attention module between any two adjacent stages (Fig. 3a) to give different weights on them. In each attention module, as shown in Fig. 3b, we first produced a channel attention map $A_i^c \in R^{C\times1\times1}$ $(i = 2, 3, 4)$ by exploiting the inter-channel relationship of high-level features to weight low-level features to select precise details, and the refined low-level feature after this operation was defined as $M_i^c \in R^{C\times H\times W}$ $(i = 2, 3, 4)$, and then we produced a spatial attention map $A_i^s \in R^{1\times H\times W}$ $(i = 2, 3, 4)$ by using the inter-spatial relationship of M_i^c, after this operation, the final refined low-level feature $M_i^s \in R^{C\times H\times W}$ $(i = 2, 3, 4)$ was generated. The overall top-down pathway integrated with attention

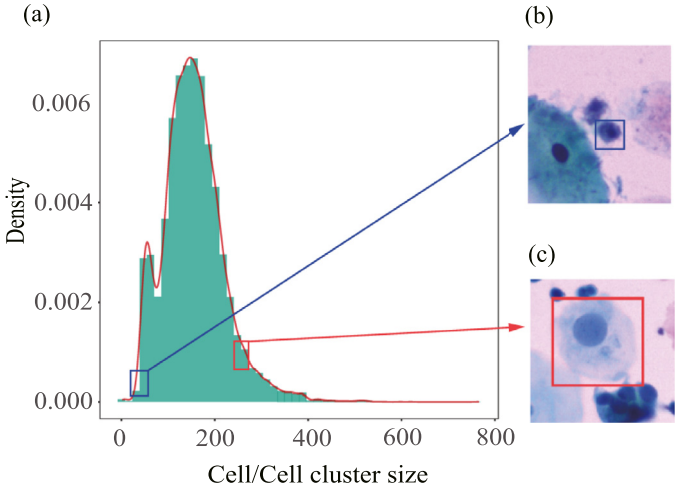

Fig. 4. Statistical description of abnormal cervical cells in our dataset. (a) The scale histogram (in pixels) of the abnormal cervical cells. (b) An abnormal cervical cell with relatively small size. (c) An abnormal cervical cell with relatively large size. The blue and red boxes in (a) indicate the cell size locations in the area distribution for (b) and (c). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Comparison of the proposed method performance (%) with related methods on our annotated dataset. All results are reported from the testing dataset.

Method	Backbone	AUROC	AUPRC
Faster R-CNN	ResNet-50	97.59	93.15
	ResNet-101	96.53	91.23
	ResNet-152	98.29	94.91
	DenseNet-121	98.14	95.35
	DenseNet-161	98.42	96.25
	DenseNet-169	97.73	94.18
Faster R-CNN+FPN	ResNet-50	98.66	96.27
	ResNet-101	98.57	96.72
	ResNet-152	98.21	95.36
	DenseNet-121	98.76	96.62
	DenseNet-161	98.58	96.38
	DenseNet-169	98.14	95.31
AttFPN	ResNet-50	**98.86**	**96.95**
	ResNet-101	**98.94**	**97.03**
	ResNet-152	**98.72**	**96.69**
	DenseNet-121	**99.29**	**98.25**
	DenseNet-161	**98.73**	**96.86**
	DenseNet-169	**99.09**	**97.60**

module could be summarized as:

$$N_5 = P_5 \tag{1}$$

$$N_6 = maxpool(N_5) \tag{2}$$

$$M_i^c = P_i \otimes A_i^c \tag{3}$$

$$M_i^s = M_i^c \otimes A_i^s \tag{4}$$

$$N_i = M_i^s \oplus f_{up}(N_{i+1}), i = 2, 3, 4 \tag{5}$$

where $\otimes$, $\oplus$ denotes element-wise multiplication and element-wise addition, respectively. N_2, N_3, N_4, N_5, N_6 refer to fused features. Both kernel size and stride of maxpool is 2. f_{up} is an upsampling process with a scale factor of 2 that acts on the fused feature. Note that the attention maps, A_i^c and A_i^s, are produced by attention module. The attention module consists of a channel attention part and a spatial attention part, as shown in Fig. 3b. The details of each part are described as follows.

In channel attention part, we aggregated spatial information of high-level feature by using both max-pooling and average-pooling, generating two different features: $F_{max,i}^c \in R^{C\times1\times1}$ and $F_{avg,i}^c \in R^{C\times1\times1}$, $(i = 2, 3, 4)$. Both of them were then forwarded to a shared two-layer fully connected network to generate the channel attention feature map A_i^c, encoding which channel to emphasize or suppress. The shared two-layer fully connected network could both reduce parameters compared with two parallel separated two-layer fully connected layers and add more linearity to $F_{max,i}^c$ and $F_{avg,i}^c$. In short, the channel attention was computed as:

$$F_{max,i}^c = maxpool(N_{i+1}) \tag{6}$$

$$F_{avg,i}^c = avgpool(N_{i+1}) \tag{7}$$

$$A_i^c = \sigma(W_2\delta(W_1F_{max,i}^c) + W_2\delta(W_1F_{avg,i}^c)) \tag{8}$$

where the kernel size and stride of both maxpool and avgpool equal to the size of N_{i+1}. σ and δ refer to the sigmoid function and ReLU function, respectively, $W_1 \in R^{C/r\times C}$ and $W_2 \in R^{C\times C/r}$ are weights of the two-layer fully connected network, and r is the reduction ratio used to do dimension reduction in the two-layer fully connected network. In Eq. 8, we used addition instead of concatenation in the channel attention module in order to control the dimension of the feature maps as same as that of the corresponding low-level feature.

In spatial attention part, firstly, we used max-pooling and average-pooling along the channel axis of M_i^c to generate two different features: $F_{max,i}^s \in R^{1\times H\times W}$ and $F_{avg,i}^s \in R^{1\times H\times W}$. The informative regions would be highlighted by applying the pooling operation. Then, we concatenated them and applied a 7×7 conv layer to produce the final spatial attention feature map A_i^s, encoding where to emphasize or suppress for M_i^c. In short, the spatial attention was computed as:

$$F_{max,i}^s = maxpool(M_i^c) \tag{9}$$

$$F_{avg,i}^s = avgpool(M_i^c) \tag{10}$$

$$A_i^s = \sigma(f_{7\times7}(concat(F_{max,i}^s, F_{avg,i}^s))), i = 2, 3, 4 \tag{11}$$

where σ refers to the sigmoid function, $f_{7\times7}$ denotes a convolution operation with a kernel size of 7, and *concat* represents a concatenate operation along the channel axis. In Eq. 11, since the feature representation of concatenation can cover that of addition in training process, we used concatenation instead of addition to combine $F_{max,i}^s$ and $F_{avg,i}^s$. In order to deciding spatially important regions, large kernel and effective receptive field (Peng et al., 2017), i.e 7×7 convolution with sigmoid function, was used.

After the multi-scale network integrated with attention module, we got the fused feature maps N_2, N_3, N_4, N_5, N_6. In this way, our proposed network yielded region proposals through RPN at five different scales (i.e., N_2, N_3, N_4, N_5, N_6), for which the feature map sizes were 1/4, 1/8, 1/16, 1/32, and 1/64 of the original input image size, respectively. Together with the corresponding feature maps, those region proposals were then fed into the Fast R-CNN block to detect abnormal cervical cells with different sizes, as shown in Fig. 3a.

2) Detection anchors restricted by clinical-knowledge: similar to the original Faster R-CNN, our proposed network also predicted region proposals using anchors with designed sizes and aspect ratios

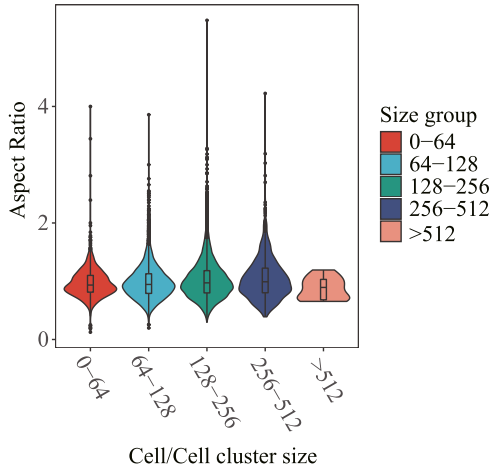

Fig. 5. The distribution of aspect ratios with respect to the size of the abnormal cervical cells. Here, we use the square of area (in pixels) to measure abnormal cervical cell size. Each violin shows the distribution of aspect ratios for a specific range of cell or cell cluster sizes.

(i.e., height divided by width). Generally, anchors are proposal candidates with fixed sizes defined on feature maps. These anchors are mapped back to the input image to get proposals after shifting and resizing. Thus, the way defining sizes and aspect ratios of these anchors directly influences the quality of the predicted region proposals. To ensure that the anchors can cover the variability of real abnormal cervical cells, we defined anchors differently at different feature maps N_2, N_3, N_4, N_5, N_6 based on clinical knowledge.

Specifically, according to clinical knowledge, most cervical exfoliated cells are roughly circle or oval shaped and rarely expand in only one direction. Thus, their aspect ratios actually lie in a certain range. Inspired by this fact, we conducted a statistical description of the cervical cells used in our study with the results shown in Fig. 5. We could observe that the abnormal cervical cells aspect ratios mainly fall between 0.4 and 2.0, which indicated that the aspect ratio for anchors and proposed regions should be selected within this range. Therefore, to ensure that the size and aspect ratio of region proposals to be consistent with the real cervical cell distribution, we generated anchor boxes with 1 optional size and 3 optional aspect ratios at each feature map location. The size (i.e., height in pixels) and aspect ratios were defined at different scales, which could cover the different abnormal cervical cells sizes in our current images, as shown in Table 1. The size is represented by the number of pixels in the original image. Thus, we could effectively improve the quality of region proposals for detection based on the clinical knowledge on the aspect ratios and sizes of abnormal cervical cells.

2.2.2. Image patch classification

It was noteworthy that the detection results indicated the locations and probability of abnormal cervical cells in the input image patch. Those location and probability information could provide valuable diagnosis advice for pathologists in screening process. In addition to the detection results, the classification result of the input image patch is also an important reference for pathologists. Actually, whether the image patch is normal or not depends on the detection results. Thus, as shown in Fig. 2b, inspired by (Zhou et al., 2021), we further trained a convolutional neural network to further determine if the image patch is recognized as normal or abnormal based on the detection results in cell-level detection stage.

First, we extracted 4 detected bounding boxes with high abnormal probabilities from the given image patch. The upright locations of the bounding boxes were obtained from the cell-level detection. If there were not enough bounding boxes, then the detected bounding boxes were randomly augmented. Second, the extracted bounding boxes were then forwarded to ResNet50 based feature extractor to extract the feature map of each input bounding box patch. The abnormal probabilities provided by the cell-level detection stage were passed in as the auxiliary context input at the same time. A cell with high abnormal probability was more likely to be abnormal. Thus the features extracted from different bounding box patches should be given different weights. The feature maps were multiplied by the corresponding abnormal probabilities to obtain the weighted feature maps. Finally, the weighted feature maps were fused to complete image-level classification. Specifically, element-wise addition was performed on the weighted feature maps, followed by global average pooling, and classification result was finally obtained through a fully connected layer.

2.2.3. Case-level classification

After the image-level classification, each image patch from a given case had its corresponding abnormal probabilities. Generally, an abnormal case tend to have more abnormal image patches, and the number of abnormal image patches in a normal case should be near to 0. Thus, in case-level classification, we empirically collected the top-10 image patches with high abnormal probabilities per case, and used their mean value as the case-level abnormal probability value.

3. Experiments

To evaluate the abnormal cervical cell detection, image-level classification and case-level classification performance of our proposed model, we conducted experiments using the collected images that we prepared for this task.

3.1. Evaluation criteria

In our experiments, the average precision (AP) and the free-response receiver operating characteristic (FROC) analysis (Chakraborty and Winter, 1990) were used to evaluate the detection performance. To evaluate the classification performance, AUROC (area under the ROC curve), AUPRC (area under the precision-recall curve), sensitivity, specificity and accuracy were used. These metrics were defined as follows:

$$AP = \sum_{k=1}^{m} P(k)\Delta r(k) \tag{12}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{13}$$

$$Specificity = \frac{TN}{TN + FP} \tag{14}$$

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{15}$$

where k is the number of abnormal cervical cells that have been detected, $P(k)$ is the precision at the cut-off k in the list, and $\Delta r(k)$ is the change in the recall from items $k-1$ to k, TP, TN, FP, FN represented the number of true positive, true negative, false positive, false negative image patches or cases, respectively.

The FROC analysis was used for detection studies. In an ROC curve, sensitivity is plotted on the y-axis and 1-specificity is plotted along the x-axis, while the FROC curve replaced the 1-specificity by the number of false positives per image.

3.2. Experiment set-up

Training and deployment of our model was performed using a workstation equipped with two NVIDIA Quadro GV100 32GB GPU cards. We implemented the model using the PyTorch (Paszke et al., 2017) (version 1.5) framework. To measure the effectiveness of our model, we randomly split the cases into 80%, 10%, 10% to ensure image patches from a case go to the same dataset, and the corresponding images were used for training, validation, and testing, respectively (Supplementary Table 1 and Table 2). The model was trained using the training subset and validated using the validation subset. The model with best validation performance was saved and evaluated using the testing subset. We trained a total of 128,560 iterations using the stochastic gradient descent (Bottou, 2012) with a batch size of 2 and momentum of 0.9. The weight decay was kept as 0.0005 to avoid overfitting. The initial learning rate (lr_0) was 0.005 and was decayed according to the equation:

$$lr(x) = \begin{cases} lr_0 \times \frac{1+0.999x}{1000} & x \leq 1000 \\ lr_0 & 1000 < x \leq 32140 \\ lr_0 \times \gamma & 32140 < x \leq 96420 \\ lr_0 \times \gamma^2 & x > 96420 \end{cases} \tag{16}$$

where $lr(x)$ denotes the learning rate of current iteration x, γ refers to the learning rate decay factor, which was set as 0.1. Here, learning rate warmup (Luo et al., 2019) was applied during the first 1000 iterations to bootstrap the network for better performance. For the $W_1 \in R^{C/r \times C}$ and $W_2 \in R^{C \times C/r}$ in Eq. 8, the r was set to 16.

Data augmentation was used to overcome the overfitting. In our experiment, we performed random horizontal flip, vertical flip, rotation, brightness change, and gaussian blur. These affine transformations helped the model to have a better understanding of the input image since it viewed the images in many transformed views. The input images were rotated randomly by 90, 180, 270 degrees. The range between [0.5, 1.5] was used for brightness change, and a sigma value between [0, 5] was used for the gaussian blur.

4. Results

4.1. Cell-level detection results

We compared the detection performance of our method with original Faster R-CNN method (Ren et al., 2015) and Faster R-CNN+FPN method (Lin et al., 2017) over various network architectures including ResNet (He et al., 2016) and DenseNet (Huang et al., 2017). The corresponding results are summarized in Table 2. Faster R-CNN and Faster R-CNN based methods were commonly used in previous cervical cytology detection and classification studies (Li et al., 2019; TSUMORI et al., 2019; Zhou et al., 2021; Liang et al., 2021; Tan et al., 2021; Sompawong et al., 2019). Our proposed method clearly outperformed the original Faster R-CNN in terms of average precision (AP) values over all backbone networks. Specifically, compared with Faster R-CNN, our method improved the AP by an average of 4.44% over all backbone networks. For example, the AP of our proposed method using DenseNet-121 as backbone network was 75.54%, which was better than Faster R-CNN method (69.85%). Secondly, our method, integrated with attention module, yielded better performance compared with Faster R-CNN+FPN method, e.g., the AP increased by an average of 2.62% over all backbone networks. For instance, the AP achieved by our method was 75.54%, which was better than Faster R-CNN+FPN (71.09%). These results suggest that with our attention module, the abnormal cervical cell detection accuracy can be effectively improved, since the presented attention module can refine feature learning more effectively.

Fig. 6 presents three FROC (Methods) curves obtained from our novel AttFPN, original Faster R-CNN and Faster R-CNN+FPN model based on testing dataset, respectively. Our results demonstrate that our model which combines multi-scale feature fusion and the presented attention module achieved the best performance. It reflects that the proposed attention module learned what and where to emphasize or suppress, and encouraged the model to refine features effectively. Meanwhile, the multi-scale feature fusion made full use of multi-scale refined features and detected abnormal cervical cells with various sizes. For example, as shown in Fig. 6a the proposed model achieved TPR=94% with FPR=3 per image, which was higher than Faster R-CNN (88%) and Faster R-CNN+FPN (87%).

Apart from the quantitative comparison presented in Table 2 and Fig. 6, we also qualitatively compared the detection results, with some examples shown in Fig. 7. The results strongly support our methods comparability with the pathologists performance in correctly detecting the abnormal cervical cells. Moreover, while Faster R-CNN or Faster R-CNN+FPN method failed to detect some small cells, our method could detect abnormal cervical cells with various sizes more accurately.

Table 4
Sensitivity, specificity, accuracy, and diagnostic time comparisons between AttFPN and the four pathologists. For the proposed model, the optimum cut-off point was selected according to Youden index analysis.

	Sensitivity	Specificity	Accuracy	Time(s/image)
AttFPN	0.980	0.980	0.980	0.04
Pathologist 1	0.400	0.980	0.690	17.36
Pathologist 2	0.413	0.960	0.687	22.73
Pathologist 3	0.753	0.900	0.827	11.55
Pathologist 4	0.960	0.913	0.937	7.67
Average pathologist	0.632	0.938	0.785	14.83

4.2. Image patch-level classification results

We evaluated the image patch-level classification performance of the proposed method, and the results of the same for different methods are summarized in Table 3. With a 99.09% AUROC score and a 97.60% AUPRC score using DenseNet-169 as backbone network, our method has outperformed Faster R-CNN and Faster R-CNN+FPN. In order to test the statistical significance of AUROC improvement, Delong test (DeLong et al., 1988) was conducted. The improvements caused by our method are statistically significant ($P < 0.05$).

4.3. Comparison with human experts

To further evaluate the performance of the AttFPN, we conducted a reader study to classify cervical cytology images into normal/abnormal class. 300 image patches were randomly chosen as an independent testing dataset used in our study. As in Supplementary Table 2, a total of 150 abnormal images were randomly chosen from 1007 testing abnormal images, and 150 normal images were randomly chosen from 2963 testing normal images. Four pathologists with different experience were recruited to read these images independently. The first two pathologists (1 and 2) were with 4-year work experience, the third pathologist (3) was with 7-year work experience, and the fourth pathologist (4) was with more than 10-year work experience.

The results in Fig. 8 demonstrate that the proposed model achieved an AUROC of 0.99, an accuracy of 98.0%, a sensitivity of 98.0%, and a specificity of 98.0% (Table 4), which is comparable to 10-year experienced pathologist (accuracy: 93.7%, sensitivity: 96.0%, specificity: 91.3%). Furthermore, diagnostic discrepancy among pathologists is significant. Third, the average diagnostic time of pathologists was 14.83s per image, while it only took 0.04s per image for our AttFPN method. The improvements indi-

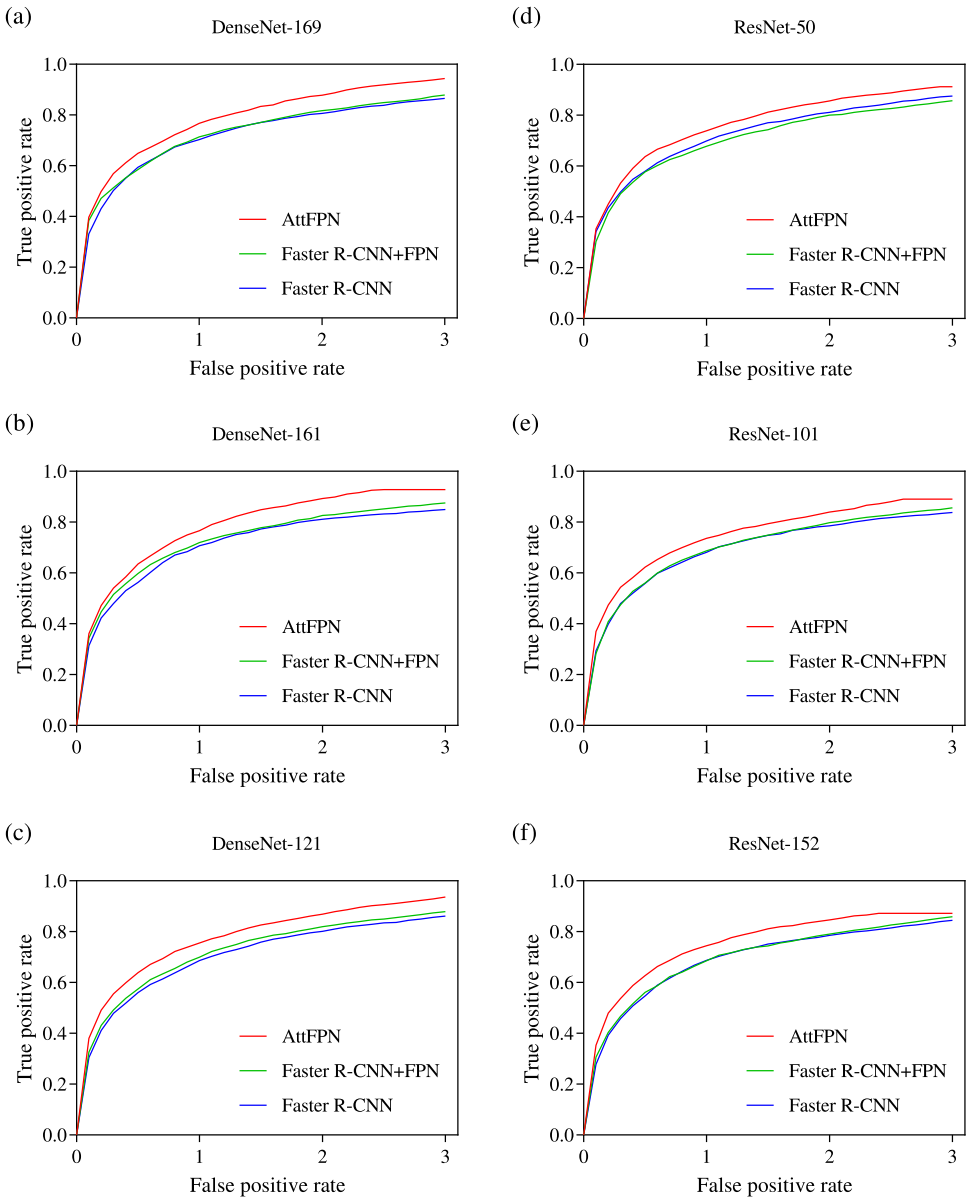

Fig. 6. Evaluating the proposed models performance for abnormal cervical cell detection in terms of FROC curves with the false positive rate (FPN) on the X-axis and the true positive rate (TPR) on the Y-axis. DenseNet and ResNet are used as backbone networks, respectively.

Table 5
Case-level performance comparison between AttFPN and the other methods. DenseNet-169 was used as the backbone network. The optimum cut-off point was selected according to Youden index analysis.

	AUROC	AUPRC	Sensitivity	Specificity	Accuracy
AttFPN	**0.934**	**0.920**	**0.913**	**0.906**	**0.909**
Faster R-CNN	0.781	0.806	0.717	0.797	0.764
Faster R-CNN+FPN	0.915	0.908	0.826	0.891	0.864

cated that our proposed model could be a potential assistant to help pathologists make a more rapid and accurate diagnosis.

4.4. Case-level classification results

We used an external dataset from a different organization (the HMCHH dataset, Supplementary Fig. 1b) to evaluate the case-level classification performance of our AttFPN model. The results are summarized in Table 5 and Fig. 9. Our AttFPN method clearly outperformed the other two related methods. In the ROC curve (Fig. 9), the area enclosed by our AttFPN is the largest, Therefore, for the case-level classification, our method can classify abnormal cases from normal cases more accurately.

4.5. Feature learning

To verify the ability of our proposed attention module to refine intermediate features effectively, we visualized the multi-scale fused features generated after feature fusion in both our method and the Faster R-CNN+FPN, as shown in Fig. 10. Compared with the feature maps generated by Faster R-CNN+FPN, the abnormal cervical cells were emphasized and the normal cells were suppressed in the feature maps generated by our method. It suggests that the proposed attention module encourages the model to learn more discriminative features.

5. Discussion

In this study, we developed a novel attention feature pyramid network that can integrate pathological experience and extracted

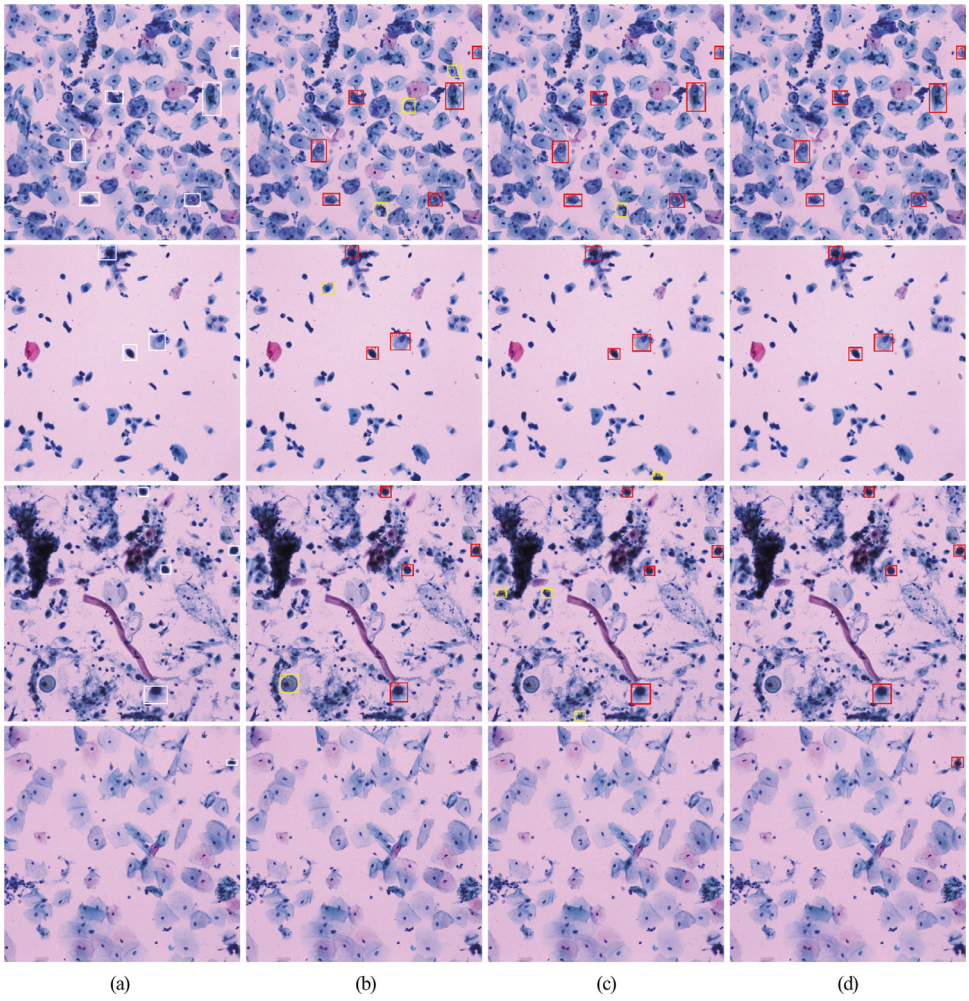

Fig. 7. Illustration of the predicted detection results using image patches. DenseNet-169 is used as the backbone network. (a) The ground-truth annotations. The detection results and their corresponding abnormal scores predicted by Faster R-CNN, Faster R-CNN+FPN, and our AttFPN method are shown in (b), (c) and (d). White boxes represent the ground-truth annotations, yellow boxes illustrate wrongly detected cells, and red boxes illustrate successfully detected cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

features from the images in a training model and evaluated the performance of this model on an independent external dataset. To the best of our knowledge, this is the first study to detect the abnormal cervical cells by ThinPrep Pap Test images features extracted from cytological images based on novel deep learning techniques. Our study demonstrated that our AttFPN could improve the accuracy in cervical cancer screening: it almost unambiguously classified normal versus abnormal cervical cytology images, which is comparable to that of an experienced pathologist with 10 years of experience. Furthermore, our AttFPN model is almost 380 times faster than an average pathologist. Our results imply that AttFPN method has been proven to be extremely rapid and effective in detection of abnormal cervical cells with Pap test images.

So far, the vast amount of valuable information contained in ThinPrep Pap Testing images has posed a big challenge to pathologists regards of experiences and time. The huge dimension of the cervical cytology slide makes it extremely difficult to manipulate, and the workflows requiring manual scanning all the fields of vision is time-consuming and labor-intensive. Although several studies have been conducted to intensify the work efficacy and accuracy of abnormal cell detection in cervical cancer screening, but they have either used handcrafted features or conducted transfer learning from natural image datasets. However, the handcrafted features are not robust enough, and there are fundamental differences in data sizes, features and task specifications between natural images classification and the target medical tasks.

Our approach has several advantages in improving the screening performance. First of all, no additional human effort is needed in our informatics workflow other than the manual annotation for abnormal cervical cells for the training data. Moreover, instead of directly using transfer learning, the prior clinical knowledge about size and shape distributions of real abnormal cervical cells is introduced into our method. As we know, the cell changes or abnormalities in Pap test images are expected to be biologically informative, this method utilized the novel attention module which could refine feature representations more effectively. Most importantly, we make full use of multi-level, channel-wise and spatial features within neural network, where we recalibrate the channel-wise feature responses using channel attention module with emphasizing on salient intermediate features and exploit the inner-spatial relationship of the intermediate features using spatial attention module. These features capture both the local anatomical structure (for example, shape of the abnormal cells) and more global patterns (for example, texture) of the abnormal cells. As a benchmark for the utility of our objective features, machine-learning models with selected features successfully identified images with abnormal cells and classified patients into positive and negative, showing that our

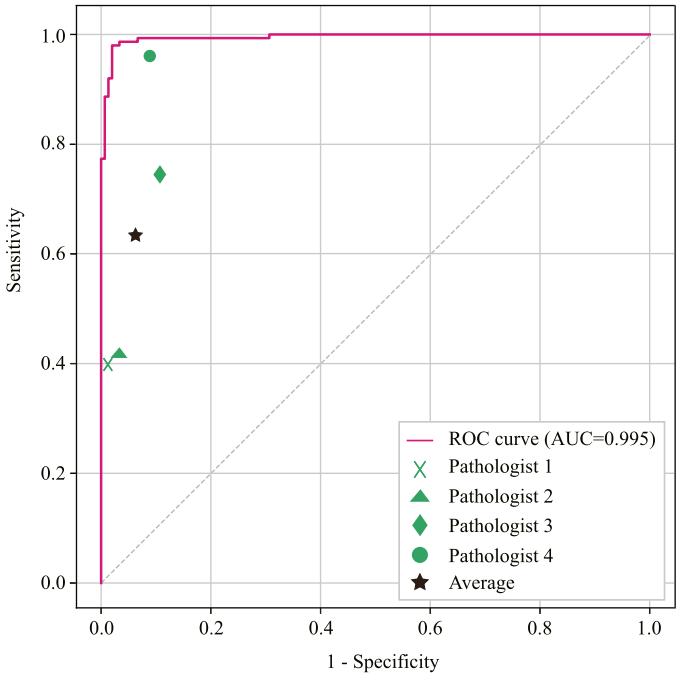

Fig. 8. Performance comparisons between AttFPN and pathologists. 150 images with abnormal cells and 150 images without abnormal cells were involved and four pathologists were instructed to classify each image as negative or positive. The model prediction ROC curve and 4 pathologists performance.

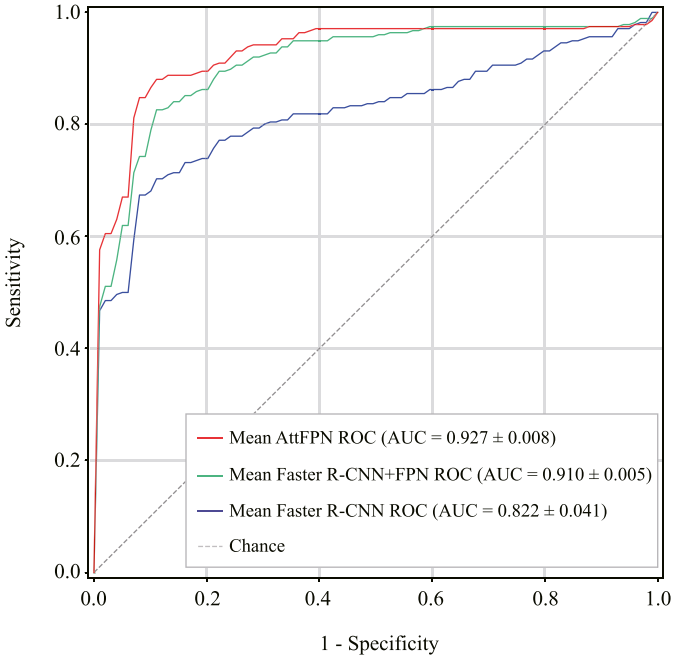

Fig. 9. The case-level average ROC of our AttFPN method and the other two related methods on the external dataset from a different organization. For each method, we took the mean AUC value over different backbone networks and plotted the average ROC curve.

image features could recapture the important information based on the annotation by trained pathologists.

Our findings demonstrate that the extracted morphological features in Pap test images can identify patients with abnormal cells in cervical cancer screening and provide the potential location for the lesions. While these image features are generally difficult to spot by manual inspection, computerized methods can efficiently

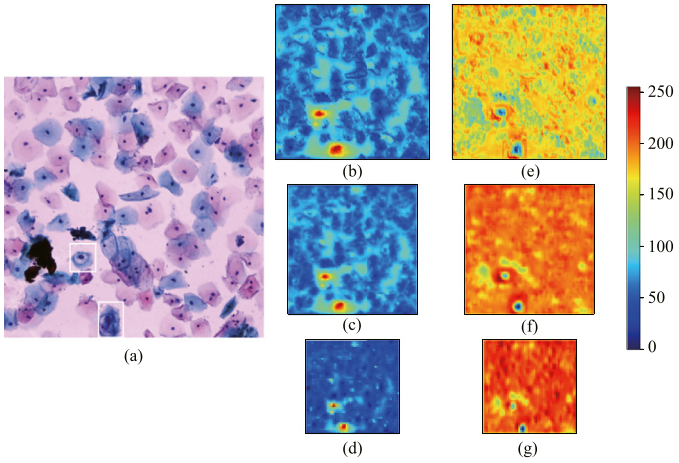

Fig. 10. Comparison of feature maps between our proposed method and Faster R-CNN+FPN with DenseNet-169 as backbone network. Input image with ground-truth annotations is shown in (a). N_2, N_3, N_4 layers features of our method are shown in (b), (c), and (d), respectively. The counterparts of Faster R-CNN+FPN method are shown in (e), (f), and (g), respectively.

and effectively identify them. Since Pap test images are routinely prepared and reviewed in current clinical practice, our classifiers can be efficiently applied to routine practice. We validated our model for abnormal prediction by an independent screening data set, signifying the generalizability of our approach. This method is well-suited for analyzing large amounts of data and large number of features in our analysis. Accurate diagnosis in cervical cancer screening generated by our models can guide clinical decision making and enhance precision medicine.

Our method has multiple clinical benefits. First of all, manual cytology image reading in cervical cytology screening is a costly, tedious and time-consuming task. In China, pathologists in most general hospitals need to read almost 100 cervical cytology slides every day (Di et al., 2015). Rapid and accurate abnormal cervical cells detection can help liberating pathologists from the heavy image reading task and allow the pathologists to concentrate on higher-level decisions, such as making treatment decisions for individual patients. Secondly, the diagnosis accuracy varies across pathologists, with false-negative rates reported as high as 55% (Branca and Longatto-Filho, 2015; van der Graaf and Vooijs, 1987). Our proposed method will potentially help reduce misdiagnosis rate, especially for inexperienced pathologists. Thirdly, cervical cytology image reading relies on the skill and experience of pathologists, and, in developing countries, the number of such pathologists is always insufficient. Our method will potentially benefit those countries, especially for resource-poor regions.

Although our proposed method effectively improved the detection performance of abnormal cervical cells, there are also several limitations. First of all, it is a single center study. Although we have external testing dataset, yet more high-quality images from multiple centers should be collected to further evaluate our method. Secondly, abnormal cervical cells include Atypical squamous cells of undetermined significance (ASC-US), Low-grade squamous intraepithelial lesion (LSIL), and so on. We did not subclass abnormal cervical cells into those different types, which may be explored in future research.

6. Conclusion

This paper presented a novel deep learning method named AttFPN as an automated detection model for abnormal cervical cells in cervical cancer screening. The model was guided by clinical knowledge and attention mechanism, consisting of a multi-

scale feature fusion structure and an attention module. The proposed method outperformed the related stat-of-the-art deep learning methods and was comparable to a pathologist with 10-year experience. Our method can be used as a reliable assistant to aid pathologists to more effectively and accurately determine abnormal cervical cells in cervical cancer screening. Furthermore, it may provide helpful suggestions for inexperienced pathologists, which is important especially in resource-poor regions.

Declaration of Competing Interest

None.

Acknowledgments

This work was supported by National Natural Science Foundation of China (project number 81773551) and National Natural Science Foundation of China (project number 81773550).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2021.102197.

References

Bhan, A., Sharma, D., Mishra, S., 2018. Computer based automatic segmentation of pap smear cells for cervical cancer detection. In: 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, pp. 883–887.
Bottou, L., 2012. Stochastic Gradient Descent Tricks. In: Neural networks: Tricks of the trade. Springer, pp. 421–436.
Boulet, G.A., Horvath, C.A., Berghmans, S., Bogers, J., 2008. Human papillomavirus in cervical cancer screening: important role as biomarker. Cancer Epidemiology and Prevention Biomarkers 17 (4), 810–817.
Branca, M., Longatto-Filho, A., 2015. Recommendations on quality control and quality assurance in cervical cytology. Acta Cytol. 59 (5), 361–369.
Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68 (6), 394–424.
Carpenter, A.B., Davey, D.D., 1999. Thinprep® pap test performance and biopsy follow-up in a university hospital. Cancer Cytopathology: Interdisciplinary International Journal of the American Cancer Society 87 (3), 105–112.
Chakraborty, D.P., Winter, L., 1990. Free-response methodology: alternate analysis and a new observer-performance experiment.. Radiology 174 (3), 873–881.
Chen, H., Zhang, K., Lyu, P., Li, H., Zhang, L., Wu, J., Lee, C.-H., 2019. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. Sci Rep 9 (1), 1–11.
Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3 (3), 201–215.
DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 837–845.
Di, J., Rutherford, S., Chu, C., 2015. Review of the cervical cancer burden and population-based cervical cancer screening in china. Asian Pac J Cancer Prev 16 (17), 7401–7407.
Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542 (7639), 115–118.
van der Graaf, Y., Vooijs, G., 1987. False negative rate in cervical cytology.. J. Clin. Pathol. 40 (4), 438–442.
Harlan, L.C., Bernstein, A.B., Kessler, L.G., 1991. Cervical cancer screening: who is not screened and why? Am J Public Health 81 (7), 885–890.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
Huang, Y., Lei, Y., Wang, Q., Li, D., Ma, L., Guo, L., Tang, M., Liu, G., Yan, Q., Shen, L., et al., 2018. Telepathology consultation for frozen section diagnosis in china. Diagn Pathol 13 (1), 1–6.
Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20 (11), 1254–1259.
Jiang, X., Tang, H., Chen, T., 2017. Epidemiology of gynecologic cancers in china. J Gynecol Oncol 29 (1).
Landy, R., Pesola, F., Castañón, A., Sasieni, P., 2016. Impact of cervical screening on cervical cancer mortality: estimation using stage-specific results from a nested case–control study. Br. J. Cancer 115 (9), 1140–1146.
Li, X., Li, Q., et al., 2019. Detection and classification of cervical exfoliated cells based on faster r-cnn. In: 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT). IEEE, pp. 52–57.
Liang, Y., Tang, Z., Yan, M., Chen, J., Liu, Q., Xiang, Y., 2021. Comparison detector for cervical cell/clumps detection in the limited data scenario. Neurocomputing 437, 195–205.
Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125.
Lin, Z., Li, S., Ni, D., Liao, Y., Wen, H., Du, J., Chen, S., Wang, T., Lei, B., 2019. Multi–task learning for quality assessment of fetal head ultrasound images. Med Image Anal 58, 101548.
Liu, T., Guo, Q., Lian, C., Ren, X., Liang, S., Yu, J., Niu, L., Sun, W., Shen, D., 2019. Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks. Med Image Anal 58, 101555.
Lu, Y., Yu, Q., Gao, Y., Zhou, Y., Liu, G., Dong, Q., Ma, J., Ding, L., Yao, H., Zhang, Z., et al., 2018. Identification of metastatic lymph nodes in mr imaging with faster region-based convolutional neural networks. Cancer Res. 78 (17), 5135–5143.
Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W., 2019. Bag of tricks and a strong baseline for deep person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, p. 0.
Mahmood, T., Arsalan, M., Owais, M., Lee, M.B., Park, K.R., 2020. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster r-cnn and deep cnns. J Clin Med 9 (3), 749.
Malagón, T., 2019. Reasons for optimism about eliminating cervical cancer in china. The Lancet Public Health 4 (9), e434–e435.
of Obstetricians, A.C., Gynecologists, et al., 2010. Cervical cancer in adolescents: screening, evaluation, and management. committee opinion no. 463. Obstetrics & Gynecology 116, 462–472.
Papillo, J.L., Zarka, M.A., St John, T.L., 1998. Evaluation of the thinprep pap test in clinical practice. a seven-month, 16,314-case experience in northern vermont.. Acta Cytol. 42 (1), 203–208.
Parkin, D.M., Bray, F., 2006. The burden of hpv-related cancers. Vaccine 24, S11–S25.
Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch.
Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J., 2017. Large kernel matters–improve semantic segmentation by global convolutional network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4353–4361.
Pradhan, D., Li, Z., Ocque, R., Patadji, S., Zhao, C., 2016. Clinical significance of atypical glandular cells in pap tests: an analysis of more than 3000 cases at a large academic women's center. Cancer Cytopathol 124 (8), 589–595.
Randel, A., 2013. Acog releases guidelines on cervical cancer screening. Am Fam Physician 88 (11), 776–777.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp. 91–99.
Rensink, R.A., 2000. The dynamic representation of scenes. Vis cogn 7 (1–3), 17–42.
Sahlsten, J., Jaskari, J., Kivinen, J., Turunen, L., Jaanio, E., Hietala, K., Kaski, K., 2019. Deep learning fundus image analysis for diabetic retinopathy and macular edema grading. Sci Rep 9 (1), 1–11.
Sompawong, N., Mopan, J., Pooprasert, P., Himakhun, W., Suwannarurk, K., Ngamvirojcharoen, J., Vachiramon, T., Tantibundhit, C., 2019. Automated pap smear cervical cancer screening using deep learning. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 7044–7048.
Tan, X., Li, K., Zhang, J., Wang, W., Wu, B., Wu, J., Li, X., Huang, X., 2021. Automatic model for cervical cancer screening based on convolutional neural network: a retrospective, multicohort, multicenter study. Cancer Cell Int. 21 (1), 1–10.
Tobias, R.R., De Jesus, L.C., Mital, M.E., Lauguico, S., Guillermo, M., Vicerra, R.R., Bandala, A., Dadios, E., 2020. Faster r-cnn model with momentum optimizer for rbc and wbc variants classification. In: 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech). IEEE, pp. 235–239.
TSUMORI, T., KIDO, S., HIRANO, Y., MORI, M., INAI, K., IMAMURA, Y., 2019. Detection of malignant cells in cervical cancer cytology using faster r-cnn. Medical Imaging Technology 37 (3), 155–163. doi:10.11409/mit.37.155.
Wang, J., Ding, H., Bidgoli, F.A., Zhou, B., Iribarren, C., Molloi, S., Baldi, P., 2017. Detecting cardiovascular disease from mammograms with deep learning. IEEE Trans Med Imaging 36 (5), 1172–1181.
Woodman, C.B., Collins, S.I., Young, L.S., 2007. The natural history of cervical hpv infection: unresolved issues. Nat. Rev. Cancer 7 (1), 11–22.
Zhang, J., Hu, H., Chen, S., Huang, Y., Guan, Q., 2016. Cancer cells detection in phase-contrast microscopy images based on faster r-cnn. In: 2016 9th International Symposium on Computational Intelligence and Design (ISCID), 1. IEEE, pp. 363–367.
Zhang, J., Liu, Y., 2004. Cervical cancer detection using svm based feature screening. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 873–880.
Zhou, M., Zhang, L., Du, X., Ouyang, X., Zhang, X., Shen, Q., Luo, D., Fan, X., Wang, Q., 2021. Hierarchical pathology screening for cervical abnormality. Computerized Medical Imaging and Graphics 89, 101892.