



# Automatic Detection of Cervical Cells Using Dense-Cascade R-CNN

Lin Yi<sup>1</sup>, Yajie Lei<sup>2</sup>, Zhichen Fan<sup>2</sup>, Yingting Zhou<sup>2</sup>, Dan Chen<sup>2</sup>, and Ran Liu<sup>2(✉)</sup>

<sup>1</sup> Chongqing University Cancer Hospital, Chongqing 400030, China

<sup>2</sup> College of Computer Science, Chongqing University, Chongqing 400044, China

ran.liu\_cqu@qq.com

**Abstract.** It is necessary to realize the automatic detection of cervical cells in Pap Smears. We present an automatic cervical cell detection approach based on the so-called Dense-Cascade R-CNN (Dense-Cascade Region-based Convolutional Neural Networks). The approach consists of three modules: data augmentation, training set balancing (TSB), and the Dense-Cascade R-CNN. The data augmentation module carries out operations such as rotation, scale, flip, etc. on input images to increase the samples. The TSB module is used to balance the number of cervical cell samples of various classes in the training set after data augmentation. As for the Dense-Cascade R-CNN module, the residual neural network (ResNet) with 101 layers in a Cascade R-CNN is replaced by a dense connected convolutional neural network (DenseNet) with 121 layers so as to improve the detection performance of cervical cells. We evaluated the proposed method on the Herlev dataset. The results show that our approach can improve both mean average precision (mAP) and mean average recall (mAR) for Cascade R-CNN. Our cervical cell automatic detection approach can be used as an auxiliary diagnostic tool for cervical cancer screening.

**Keywords:** Cervical cancer · Cervical cells · Automatic detection · Pap smear · Deep learning

## 1 Introduction

With the rapid development of social economy, the natural and social factors that endanger women's health are increasing, and the incidence of gynecological cancer is also increasing year by year. Cervical cancer is the most common gynecological malignant tumor. In China, cervical cancer is the second most common gynecological malignant tumor. As cervical cancer is the only gynecological malignant tumor that can be successfully cured by early diagnosis and treatment, screening for precancerous lesions of the cervix is particularly important. In the 1940s, the Pap Smear method was established by Papanicolaou who used cervical exfoliated cells directly for cervical cancer screening. It was simple for operation and low in price. The widespread of this technique has reduced

---

L. Yi and Y. Lei—These authors contributed equally to this work and should be considered co-first authors.

the incidence and mortality of cervical cancer significantly. Now Pap Smear method is a widely used method for early screening of cervical cancer, which is of great significance for the early diagnosis and treatment of cervical cancer [1, 2]. However, this method relies on manual microscopy, the doctor's workload is large, the screening cost is high, and the accuracy of the screening results depends on the doctor's experience. In addition, the cumbersome operation in the high-powered environment is likely to cause the doctor to fatigue and thus affect the critical result. Therefore, it is necessary to achieve automatic detection of cervical cell in the Pap Smear.

In recent years, the automatic detection of cervical cell images by image processing technology has become the norm [1–3]. Neghina *et al.* introduced a method of cervical cell segmentation and classification based on a polar transformation. One aspect of this method is that a number of parameters can be observed conveniently and evaluated as fuzzy memberships to the non-cell class in the segmented polar representation, out of which the final decision can be determined [4]. Chankong *et al.* used the fuzzy C-means (FCM) clustering technique to segment a single-cell image into the nucleus, cytoplasm, and background; hence, the 2-class problem can be achieved [4].

The above research mainly uses the traditional image classification method. The problems of these methods are as follows: (1) Manual selection of features is required, so technicians need to master professional knowledge of cytology and pathology; (2) Different classifiers for different types of cervical cells need to be designed. These classifiers are not universal and are difficult to apply to different datasets. As the number of cell types increases, the tasks of the classifier design will increase dramatically; (3) These methods are only good at classification, and the cervical cells cannot be accurately marked in the image. In recent years, deep learning is widely used in the automatic detection of medical images [5]. Some scholars have used a deep convolutional neural network (DCNN) to perform feature extraction and classification on gynecological examination smears such as thinprep cytologic test (TCT), and cross-validation is used to verify its validity [4]. The results show that the performance of the deep learning method significantly outperforms that of the traditional method. However, deep learning is still less used in the automatic detection of cervical cells in Pap Smears, which contains both classification and segmentation tasks. The purpose of this paper is to explore the application of deep learning in the automatic detection of cervical cells. Since the detection performance of DCNN depends heavily on the size of the sample, the balance of the samples of different classes in the input image, and the structure of the DCNN, this paper focuses on the investigation of data augmentation, training set balancing (TSB), and the structure of region-based CNN (a kind of DCNN) to find ways to improve the automatic detection performance of cervical cells. The main contributions of this paper are as follows: (1) A data augmentation method that combining 7 carefully-selected operations is presented to expand the sample size of the data set; (2) A training set balancing (TSB) algorithm is used to balance the number of samples in the data set; (3) A improved network, named Dense-Cascade R-CNN, is given based on Cascade R-CNN: the 101-layer ResNets in the Cascade R-CNN are replaced with 121 layers of DenseNets, and a segmentation branch is added to each detection branch of the Cascade R-CNN to implement segmentation task. All these tricks help to improve the performance of our network significantly in contrast to other networks.

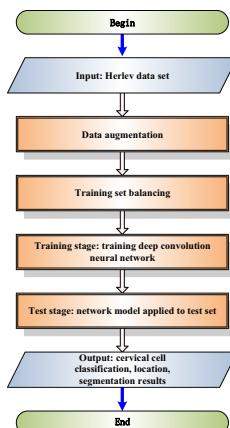
## 2 Materials and Methods

### 2.1 Dataset

We use the Herlev dataset for experiments. The Herlev dataset was produced in collaboration with the Technical University of Denmark and the Herlev University Hospital [4]. The images in this dataset are saved as BMP format. Each image is a single intact cell image with an average resolution of  $150 \times 140$  that has been divided into normal class and abnormal class by cytologists. In this paper, 917 samples (675 abnormal cells and 242 normal cells) in the Herlev dataset are used for experiments [6]. All cells can be further divided into 7 classes (C1–C7, see Table 1). The entire dataset is divided into a training set and a test set. Furthermore, a part of the samples from the training set (about 10%) is taken as the verification set. Table 1 lists the number of training and test samples for the original Herlev dataset.

**Table 1.** Number of training images and test images

Type	Class	Training set	Test set
Normal cell	Superficial squamous epithelial (C1)	58	16
	Intermediate squamous epithelial (C2)	54	16
	Columnar epithelial (C3)	82	16
Abnormal cell	Mild squamous non-keratinizing dysplasia (C4)	166	16
	Moderate squamous non-keratinizing dysplasia (C5)	130	16
	Severe squamous non-keratinizing dysplasia (C6)	181	16
	Squamous cell carcinoma in situ intermediate (C7)	134	16
<b>Total</b>		805	112



**Fig. 1.** Flowchart of the proposed method for cervical cell detection

## 2.2 Detection Method

### Data Augmentation

Since the number of samples in the Herlev dataset is too small, we use data augmentation methods to enrich the dataset by creating new similar samples before training our proposed DCNN. This operation improves the accuracy of DCNNs and reduces overfitting [6]. Random translation [7], rotation [8], scale [8], ZCA whitening [9], feature standardization [10], horizontal flip [7], and vertical flip [7] are used to increase the number of samples. After data augmentation, the number of samples is increased from 805 to 7000.

### Training Set Balancing

As the number of samples in different classes of the Herlev dataset is unbalanced, we balance the number of samples in each class by training set balancing (TSB). In this paper, we first generate a large number of new samples by various methods in the data augmentation module, and then select the required number of samples randomly from generated samples together with the original ones to form the training set, so that the number of cell images in each class is equal.

### Dense-Cascade R-CNN

Currently, there are many object detection networks based on DCNN, such as R-CNN [4], Fast R-CNN [4], Faster R-CNN [4], Mask R-CNN [11], and Cascade R-CNN [4]. They are networks based on regional recommendations. Note that the Mask R-CNN can simultaneously perform classification, location, and pixel-level segmentation [4], hence it is always chosen as a benchmark for performance evaluation. In our experiments, we also chose it as a benchmark for comparison purpose.

Different from Mask R-CNN or Faster R-CNN, Cascade R-CNN is composed of a sequence of detectors trained with increasing IoU thresholds, and the detectors are trained sequentially, using the output of a detector as the training set for the next. Actually, Cascade R-CNN is a multi-stage extension of the Faster R-CNN. In the two-stage architecture of the Faster R-CNN, The first stage is a proposal sub-network, which is applied to produce preliminary detection hypotheses. In the second stage, these hypotheses are processed, a final classification score and a bounding box are assigned per hypothesis. Cascade R-CNN contains four stages, i.e. an RPN network and three detectors. In these three detectors, the input of each detector is the result of the regression of the boundary box of the previous detector, and the IoU thresholds of the three detectors are different and increasing.

Cascade R-CNN using ResNet backbone has achieved excellent results in the segmentation of many large datasets [11]. It outperforms most state-of-the-art detectors. However, ResNet is not an optimal choice for Cascade R-CNN for feature extraction for small datasets. Some research argues that for a relatively small dataset, the network called DenseNet (Densely Connected Convolutional Neural Network) always shows better performance [12], because it can solve the problem of overfitting better [12].

The basic idea of DenseNet is similar to ResNet, but it establishes a dense connection between all the front and back layers. DenseNet is mainly divided into two modules:

Dense block and Transition layer. The Dense block consists of two convolution layers with a convolution kernel size of  $1 \times 1$  and  $3 \times 3$ , respectively. The Transition layer consists of a convolution layer of  $1 \times 1$  convolution kernel and an average pooled layer. Similar to ResNet, DenseNet is initially a  $7 \times 7$  convolution layer and a  $3 \times 3$  maximum pooling layer, followed by the Dense block and Transition layer, and finally the  $7 \times 7$  global pooling layer and the fully connected layer. The feature map size of the corresponding layer of DenseNet can match the feature map size of the 5 stages output in ResNet. Because of this architecture, we can replace the ResNet in Cascade R-CNN with DenseNet. Another feature of DenseNet is that it can reuse feature through the connection of features on Channel [12]. These features allow DenseNet to achieve better performance than ResNet in the case of fewer parameters and computational costs. Studies have shown that the feature extraction performance of DenseNet is superior to ResNet [12].

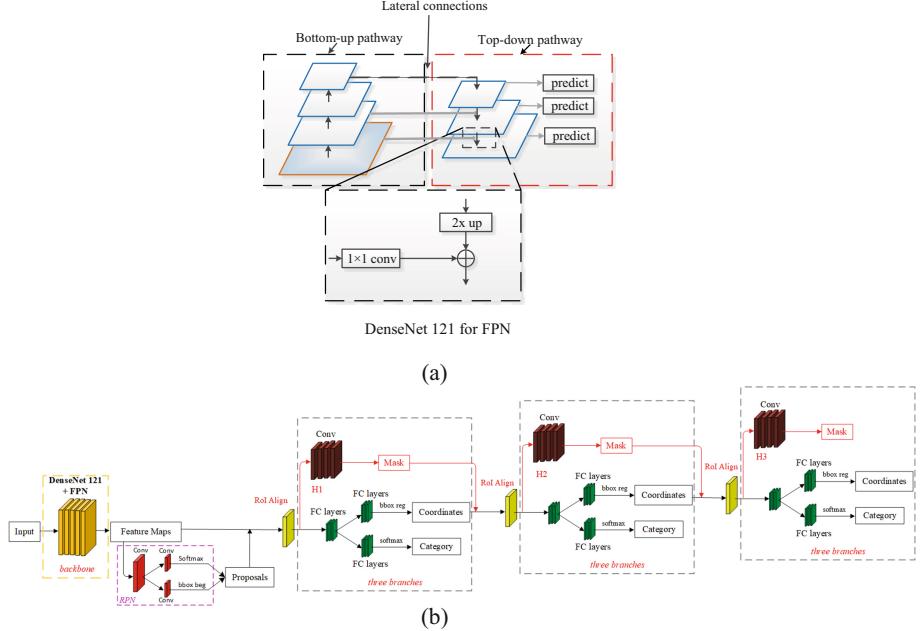
Inspired by this conclusion, we attempt to replace the 101-layer ResNet in Cascade R-CNN with the 121-layer DenseNet. Moreover, we add a segmentation branch to each detection branch of Cascade R-CNN to implement segmentation tasks. Figure 2(b) shows the architecture of our proposed DCNN, which uses the 121-layer DenseNet instead. We call this network Dense-Cascade R-CNN. We also manage to load the weights of DenseNet and Cascade R-CNN pre-trained models to initialize the proposed network. This operation speeds up the convergence rate of the gradient. We also replace ResNet in Mask R-CNN for comparison purpose. We call it Dense R-CNN.

### 3 Results and Analysis

In this paper, we use the samples listed in Table 1 for the experiment. We first evaluate the performance of the data augmentation module and TSB module, showing how they can contribute to detection; then, we evaluate the whole approach and present the cervical cell detection results.

#### 3.1 Evaluation of Data Augmentation

In order to evaluate the effect of the data augmentation module in this paper, we design the ablation experiment shown in Table 2, in which the first row is the number of each experiment, and the first column lists the methods available in the data augmentation module. The “ $\checkmark$ ” in the table indicates that this method is used in the experiment of this column. We evaluate the effect of each method on cervical cell performance by combining the methods in the module. Table 3 shows the detection performance scores on the test set when using different data augmentation methods. In order to make a fair comparison, E1–E7 does not use the TSB method, but uses each data augmentation method separately to expand the original training set in Table 1 by one time to get  $805 \times 3$  training samples (original included), and then randomly selects 24 samples from each class (hence 192 samples) to form the verification set. E8 merges the training sets of E1–E7 together to obtain  $805 \times 8$  training samples, and then randomly selects 96 samples from each class to form the verification set. The test set is always the same, and the default hyperparameters of the network are applied.



**Fig. 2.** Architecture of the proposed Dense-Cascade R-CNN. (a) FPN is constructed by DenseNet 121. (b) Overall architecture.

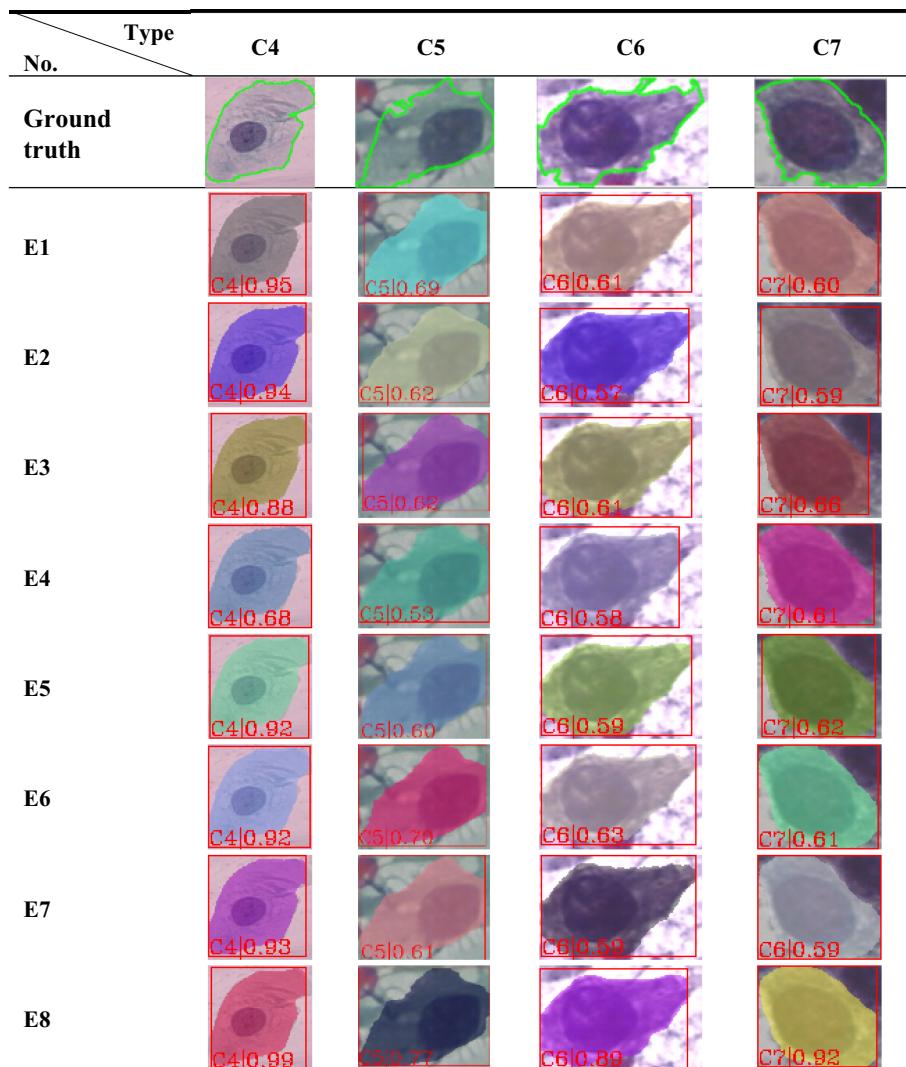
**Table 2.** Results of ablation experiments on the data augmentation module. Methods used in the module were combined to evaluate the impact of each method on the detection performance. The “√” in the table indicates that the method was used in the experiment listed in the same column

No. Method	E1	E2	E3	E4	E5	E6	E7	E8
Random translation	√							√
Rotation		√						√
Scale			√					√
ZCA whitening				√				√
Feature standardization					√			√
Horizontal flip						√		√
Vertical flip							√	√

In Table 2, E1–E7 are experiments using a single data augmentation method to evaluate the impact of the method on the detection performance of cervical cells. The purpose of E8 is designed to assess the impact of the combination of all data augmentation methods on the detection performance of cervical cells. It can be seen from Table 3 that when a single data augmentation method is applied, the best performance of mean

**Table 3.** Detection results on the test set using different data augmentation methods

No. Metric	E1	E2	E3	E4	E5	E6	E7	E8
mAP	0.539	0.483	0.505	0.464	0.478	0.539	0.525	0.787
mAR	0.620	0.577	0.588	0.547	0.560	0.611	0.589	0.833

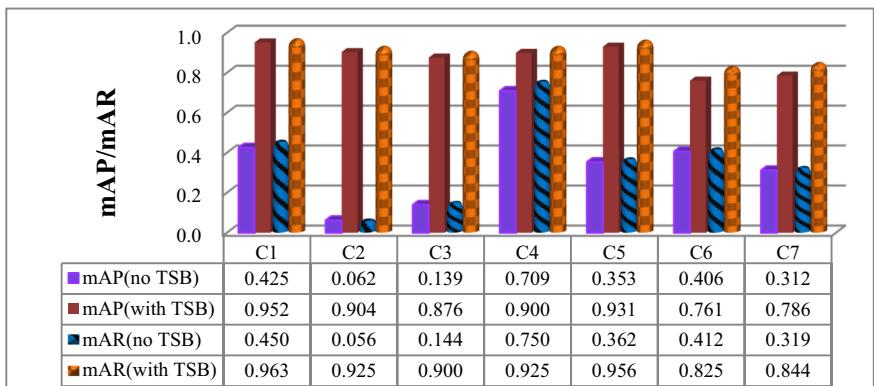
**Fig. 3.** Cell segmentation results of various data augmentation methods on Herlev dataset, taking abnormal cells as an example

average precision (mAP) is E1 (53.9%), the worst is E4 (46.4%), the best performance of mean average recall (mAR) is E1 (62.0%), and the worst is E4 (54.7%). This shows that for cervical cell detection, different data augmentation methods have different impacts on the detection results; combined with various data augmentation methods, mAP and mAR have been significantly improved compared to the best results using single method: mAP increased by 24.8% point, and mAR increased by 21.3% point. Figure 3 shows the segmentation results of abnormal cells with different data augmentation methods (E1–E8).

### 3.2 Evaluation of Training Set Balancing

In order to verify whether the TSB is beneficial to the cervical cell automatic detection performance of the network, we conduct a comparative experiment of the network cell segmentation performance without and with TSB. Default hyperparameters of the network are applied. In the experiment without TSB, the original Herlev dataset shown in Table 1 is used, with a total of 805 training set samples. For the experiment with TSB, the new training set (the number of cells in each class is 1000) and data augmentation method E8 in Table 2 are used. The results show that, when there is no TSB, the mAP of the network is 29.6%, and the mAR is 39.1%, which are much smaller than the values when TSB is applied (mAP: 79.2%, mAR: 89.0%).

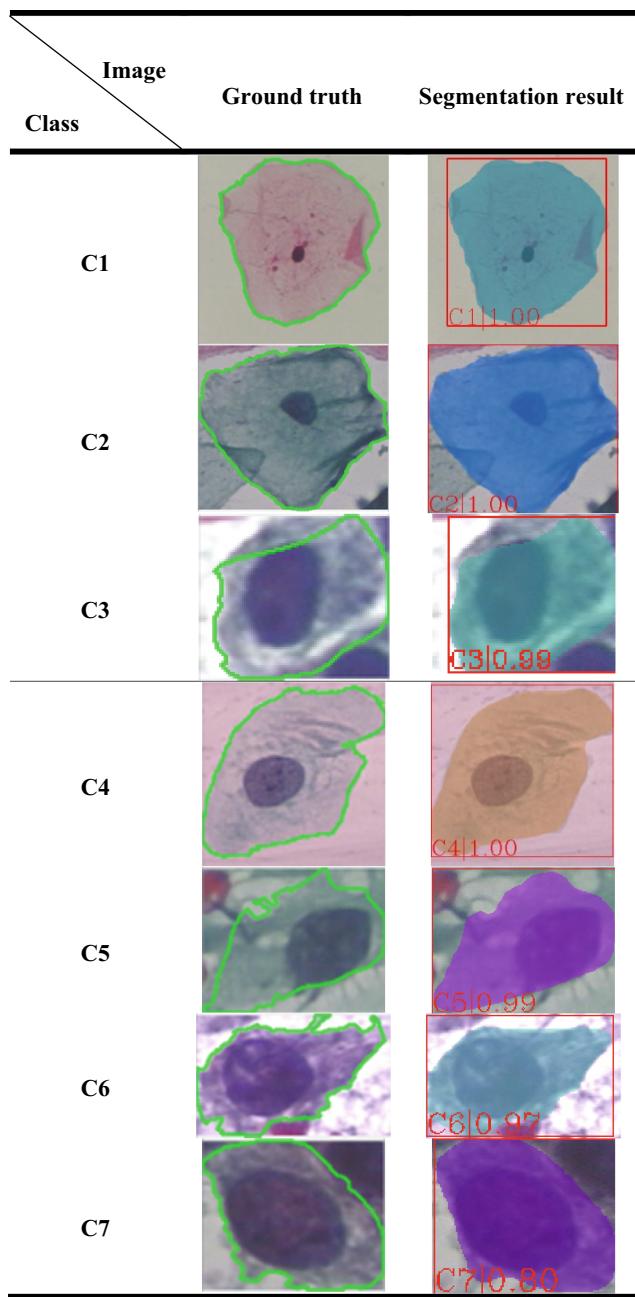
Figure 4 compares the detection performance of the network without and with TSB for each class in the test set. It can be seen from Fig. 4 that TSB improves the performance of the network in all classes, and it has the greatest influence on class C2 (intermediate squamous epithelial).



**Fig. 4.** Comparison of the detection performance of the network without and with TSB

### 3.3 Evaluation of Dense-Cascade R-CNN

In this section, we apply the proposed approach to the test set to evaluate its performance. The experiments are implemented based on PyTorch and MMDetection.



**Fig. 5.** Cell segmentation results of the proposed methods on the Herlev dataset. C1–C3: normal cell; C4–C7: abnormal cell

We fine-tune the hyperparameters of the Dense-Cascade R-CNN to get an excellent result. Figure 5 shows the detection results. As can be seen from Fig. 5, all the cells are classified, located, and segmented accurately.

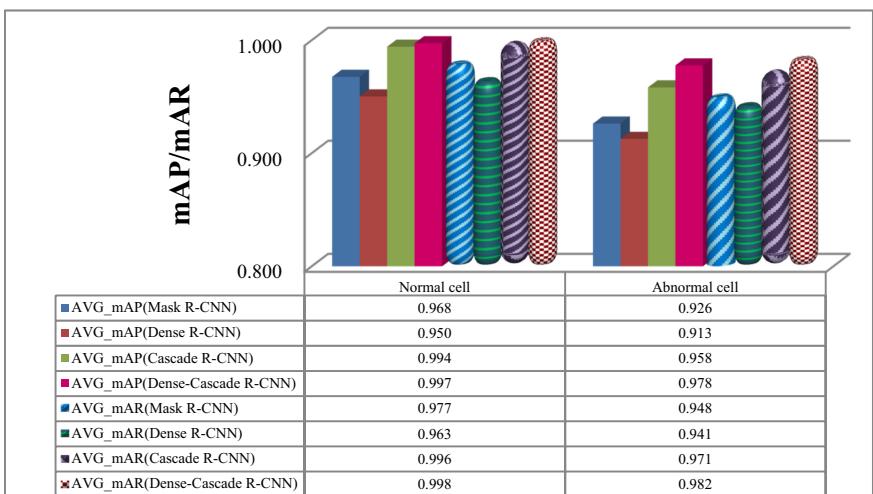
We compare the performance of several state-of-the-art approaches (Mask R-CNN, Dense R-CNN, Cascade R-CNN [13], and Dense Cascade-R-CNN) on the test set. Table 4 shows the comparison results.

**Table 4.** Performances of different approaches for cervical cell detection. The data underlined indicate the best performance metrics among these approaches. (Epoch = 200)

Metric	Mask R-CNN	Dense R-CNN	Cascade R-CNN	Dense-Cascade R-CNN
mAP	0.922	0.899	0.964	<u>0.979</u>
mAR	0.968	0.959	0.973	<u>0.988</u>

As can be seen from Table 4, Dense R-CNN achieves both the lowest mAP and mAR, indicating that DenseNet replacing ResNet is the cause of performance degradation, hence the framework of Mask R-CNN might not suitable for DenseNet. Please note both mAP and mAR are increased after ResNets are replaced by DensNets in Cascade R-CNN: the mAP of Dense-Cascade R-CNN is 1.5% points higher than that of Cascade R-CNN; for the mAR, it also increases 1.5 percentage points. Both mAP and mAR of Dense-Cascade R-CNN are highest among all approaches. This indicates our method does work better for the cervical cell detection task, and it is served as the detector according to clinical needs.

In addition, we also compare the performance of the above approaches in dealing with abnormal cells and normal cells. As shown in Fig. 6, the detection performance



**Fig. 6.** Performance comparison of different methods for segmenting abnormal and normal cells

of our approach outperforms that of other methods for both abnormal and normal cells. This may be because our network can better learn the shape and the contour details of cells.

## 4 Conclusion

Currently, most hospitals depend on the doctors to examine Pap smears for the screening of precancerous lesions in the cervix. This method is time-consuming and labor-intensive, and the accuracy of the screening results depends on the experience of the doctor. In addition, it can easily lead to fatigue of the doctor. It is of great significance to implement the automatic detection of cervical cells in Pap Smear. In this paper, a new approach for the automatic detection of Pap Smear cervical cells is proposed. Three modules, data augmentation, training set balancing, and Dense-Cascade R-CNN, are contained in our approach. One important aspect of the proposed Dense-Cascade R-CNN is that the DenseNet is used to extract features instead of ResNet in the networks so as to improve the detection performance of cervical cells. Experimental results show that compared with other state-of-the-art approaches, our proposed system can improve the automatic detection performance of cervical cells. Therefore, the proposed approach can be used as an auxiliary diagnostic tool for cervical cancer screening.

As the performance of the network relies heavily on hyperparameters, our method needs more experiments to verify the parameter sensitivity. We plan to design and carry out these experiments to investigate the effects of hyperparameters in the next step.

**Acknowledgment.** This work was supported in part by the Fundamental Research Funds for the Central Universities (2018CDXYJSJ0026, 2019CDYGZD004); the Chongqing Foundation and Advanced Research Project (cstc2019jcyj-msxmX0622); the Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN201800111), the Sichuan Science and Technology Program (2019YFSY0026), and the Entrepreneurship and Innovation Program for Chongqing Overseas Returned Scholars (No. cx2017094).

## Appendix

The supplementary materials (including the source code of the proposed approach) for this paper can be downloaded from [https://github.com/threedeam/cell\\_detection](https://github.com/threedeam/cell_detection).

## References

1. Wang, P., Wang, L., Li, Y., Song, Q., Lv, S., Hu, X.: Automatic cell nuclei segmentation and classification of cervical pap smear images. *Biomed. Sig. Process. Control* **48**, 93–103 (2019)
2. Kurnianingsih, et al.: Segmentation and classification of cervical cells using deep learning. *IEEE Access* **7**, 116925–116941 (2019)
3. Sompawong, N., et al.: Automated pap smear cervical cancer screening using deep learning. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7044–7048. Institute of Electrical and Electronics Engineers Inc., Berlin (2019)

4. Lin, T.Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944. Institute of Electrical and Electronics Engineers Inc., Honolulu (2017)
5. Kermany, D.S., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **172**(5), 1122–1131 (2018)
6. Zhang, L., Lu, L., Nogues, I., Summers, R.M., Liu, S., Yao, J.: DeepPap: deep convolutional networks for cervical cell classification. *IEEE J. Biomed. Health Inform.* **21**(6), 1633–1643 (2017)
7. Liu, R., Zhang, Y., Zheng, Y., Liu, Y., Zhao, Y., Yi, L.: Automated detection of vulnerable plaque for intravascular optical coherence tomography images. *Cardiovasc. Eng. Technol.* **10**(4), 590–603 (2019). <https://doi.org/10.1007/s13239-019-00425-2>
8. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *J. Big Data* **6**(1), 1–48 (2019)
9. Chiu, T.Y.: Understanding generalized whitening and coloring transform for universal style transfer. In: 17th IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4451–4459. Institute of Electrical and Electronics Engineers Inc., Seoul (2019)
10. Reiche, B., Moody, A.R., Khademi, A.: Effect of image standardization on FLAIR MRI for brain extraction. *SIViP* **9**(1 supplement), 11–16 (2015)
11. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(2), 386–397 (2018)
12. Gao, H., Zhuang, L., Maaten, L.V.D., Weinberger, K.Q.: Densely connected convolutional networks. In: 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269. Institute of Electrical and Electronics Engineers Inc., Honolulu (2017)
13. Singh, K.R., Chaudhury, S.: A cascade network for the classification of rice grain based on single rice Kernel. *Complex Intell. Syst.* **6**, 321–334 (2020)