Journal of Medical Systems (2019) 43: 250
https://doi.org/10.1007/s10916-019-1390-6

IMAGE & SIGNAL PROCESSING

Cervical Cancer Detection and Diagnosis Based on Saliency Single Shot MultiBox Detector in Ultrasonic Elastography

Shuli Wei [1] · Peifeng Dai [1] · Zhengping Wang [1]

Received: 8 April 2019 / Accepted: 14 June 2019 / Published online: 27 June 2019

Abstract
Cervical cancer is increasingly threatening the health of women, then early screening and prevention of cervical cancer is very necessary. A traditional saliency cervical cancer detection method in Ultrasound image, assuming that there is only one salient object, is not conductive to practical application. Their effects are dependent on saliency threshold. Object detection model provides a kind of new solutions for multiple salient objects. Shot multiBox detector can accurately detect multi-objects with different scales simultaneously except for small cervical cancer regions. To overcome this drawback, this paper presents a new multi-saliency objects detection model, appending deconvolution module embedded within attention residual module. Experiments show that our proposed diagnosis algorithm achieves higher detection accuracy than comparison algorithms. Also, it improves detection performance for mult-saliency cervical cancer objects with small scales, which greatly improve the diagnosis accuracy of cervical cancer.

Keywords Cervical Cancer · Saliency object detection · Deconvolutional · Attention residual · Shot multiBox detector · Saliency threshold · Deep learning

Introduction

Cervical cancer (CC) is a common malignancy in female reproductive system for worldwide. The incidence rate in China has been younger in recent years, and the incidence rate in developing countries holds the first place. The occurrence and development of cervical cancer is a continuous process from quantitative change to qualitative change [1]. Therefore, early detection of cervical lesions is the key to prevention and treatment of cervical cancer. Ultrasonic examination is easy, safe, non-invasive and reproducible, and is one of the main imaging methods for cervical cancer [2].

Cervical cancer is increasingly threatening the health of women, then early screening and prevention of cervical cancer is very necessary. The artificial error of cervical cell image interpretation can be efficiently decreased by automatic computer-aided diagnosis, as well as labor cost. Then cervical cancer screening technology can be extended rapidly, which has good social value and economic value [3, 4]. Cervical elastography is a new type of ultrasonic diagnostic technique for imaging tissue stiffness, which maps the elastic properties and stiffness of soft tissue. The main idea is that whether the tissue is hard or soft will give diagnostic information about the presence or status of disease [5–7]. The principle is that the tissue hardness of the lesion is different from that of the surrounding normal tissue. The strain is different under pressure. The tissue changed from soft to hard, the elastic coefficient increased gradually, and the filling color changed from red to blue. At present, the commonly used ultrasound elastography technology is compressional ultrasound elastography, and there are intermittent ultrasound elastography and vibrating ultrasound elastography [8–10]. The methods for evaluating elastography images mainly include scoring method, ratio method, shear wave velocity, etc. Liu Yanying et al. used elastography and transvaginal gray-scale ultrasound to diagnose cervical cancer [11]. The results showed that the sensitivity, specificity and accuracy of elastography in the diagnosis of cervical cancer were 58.33%, 44.44%, 55. 13%, and that of transvaginal gray-scale ultrasound is 66.67%, 55.56%, and 64.10%, respectively. The combination

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Zhengping Wang
wangzhengping80@126.com

[1] Department of Ultrasonography, Weifang People's Hospital, Weifang 261041, Shandong, China

diagnosis was 91.67%, 83.33%, and 89.74%. There was no statistically significant difference between elastography and transvaginal gray-scale ultrasound in identifying cervical cancer. But the combination of the two can improve the diagnostic accuracy. Taking pathological results as a reference, Benedet et al. [12] performed a vaginal ultrasound elastography on 113 women (13 patients with cervical cancer) using a scoring method. It was found that there were significant differences between the normal cervical and cervical cancer elastic images, but there was no statistical significance for the difference; Zhang et al. [13] used the strain rate ratio method to measure the cervical mass, and found that the strain rate of the malignant mass was much higher than that of the benign mass (8.19±5.66 vs 2.81±2.24). When the elastic strain rate is greater than 4.53, it can be diagnosed as malignant. Although elastography can provide different imaging features in cervical cancerous areas, its diagnosis depends mainly on professional evaluation.

The rapid development of medical imaging technology provides a new means for the identification of cervical cancer. Doctors can observe the characteristics of the lesion from the image, and then analyze and diagnose it. However, the current medical imaging technology for cervical cancer examination relies heavily on the experience and technology of the specialist, and often has disadvantages such as strong subjectivity, low reproducibility, high labor intensity, and low efficiency. Therefore, the automatic auxiliary detection technology for cervical cancer has extremely important significance in clinical application. Intelligent medical image processing needs to accurately locate the spatial location, size and other status of the lesion and its corresponding relationship with surrounding tissues, assist the medical staff to qualitatively and even accurately quantify the diseased tissues and organs, and then the health status and treatment plan of the cervix uteri have a more accurate judgment. However, most of the existing medical image-based processing algorithms are not universal, and the treatment effects for different human organs are different, and even the difference is great. Therefore, establishing a robust, objective, repeatable, efficient and high-accuracy medical diagnosis method for cervical cancer has important clinical significance for the prevention and treatment of cervical diseases.

Traditional shallow learning methods such as machine learning and pattern recognition are sensitive to noise and cannot effectively process and analyze large-scale image data. Compared with traditional shallow learning, deep learning represented by convolutional neural networks emphasizes the depth of model structure and highlights the importance of feature learning. By layer-by-layer feature transformation, the feature representation of the sample in the original space is transformed into a new feature space, making classification or prediction easier. Compared with the manually design features, the use of big data to learn features is more representative of data-rich intrinsic information. Shallow learning lacks the generalization ability for complex problems due to the limitations of samples and computational units. Deep learning can learn a deep nonlinear network structure, realize complex function approximation, characterize the distributed representation of input data, and demonstrate the powerful ability to learn the essential characteristics of data from the sample set. In view of the difference in gray scale of ultrasound images of cervical cancer, this paper proposes the detection and recognition of cervical cancer region using deep features, which improves the automatic detection ability and diagnostic accuracy.

Materials and Methods

Staging of Cervical Cancer

Staging of cervical cancer is very important for its treatment. Its specific stage is: stage 0: carcinoma in situ. Stage I: The focus of cancer is confined to the cervix. Stage II: The lesion had exceeded the cervix, but did not reach the pelvic wall or the lower 1/3 of the vagina. Stage IIA showed that the lesion occupied 2/3 of the vagina without obvious parametrial infiltration. Stage IIB showed obvious parametrial infiltration but did not reach the pelvic wall. Stage III: The cancer has spread to the pelvic wall. There is no gap between the pelvic wall and the tumour during the digital rectal examination. The cancer affects 1/3 of the lower vagina. All cases of hydronephrosis or renal insufficiency were included. Stage IIIA involved the lower 1/3 of the vagina, but did not spread to the pelvic wall; Stage IIIB spread to the pelvic wall, or hydronephrosis or renal dysfunction; Stage IV: cancer spread beyond the true pelvic, or clinical infiltration of bladder and (or) rectal mucosa.

In order to achieve accurate diagnosis of cervical cancer, the existing method is to use ultrasound analysis. Figure 1 is the result of color Doppler ultrasound and Elastography. It can be seen that the results obtained by these two methods are quite different. The statistical results show that the results of Elastography can better reflect the real situation. But the image needs to be interpreted by professionals. Therefore, this paper proposes to use deep learning for automatic recognition.

Saliency Detection

Saliency detection is not only helpful to the completion of visual tasks such as image thumbnail generation, image editing and summary, but also helpful to the realization of image understanding system. It has been one of the research hotspots in the field of computer vision. Early studies were mostly detection model based on saliency information, that is, based on saliency image, to find the optimal detection box containing salient objects. Liu et al. defined the optimal detection box as the smallest rectangular box with sufficient

(a) (b) (c)

Fig. 1 Imaging comparison for cervical cancer. (a) Color Doppler ultrasound image of cervical cancer; (b) Color Doppler ultrasound image of normal cervix; (c) Elastography of cervical cancer

importance and important objects [14]. And solve the rectangular box on saliency image using greedy algorithm. Valenti et al. combined edges, colors and shapes to obtain saliency images [15]. The information obtained was used as the weights of the output of the segmentation algorithm. After the saliency images were obtained, an efficient sub-window search method was used to find the optimal rectangular window of saliency object. Luo et al. regarded saliency object detection as the location problem of the region with the greatest saliency density [16]. First, the saliency object is represented by the greatest saliency density on the saliency image, and then the branch and bound search is used to speed up the detection and obtain the optimal solution. The performance of this saliency-based detection model depends greatly on the selection of saliency threshold. And it is usually assumed that there is only a single saliency object in the image, which is not consistent with the actual application. However, it is difficult to separate multiple saliency objects directly from the saliency image once the hypothesis is divorced.

Some research works attempt to extract characteristics manually from the original image and further extract saliency objects. Feng et al. proposed the notion of saliency window and proposed a segmentation-based characterization, and used the greedy algorithm to obtain the optimal saliency object detection box [17]. Yildirim et al. quantify the color to get the histogram of CIELAB color space, and calculate the spatial center and variance of quantized color by bilateral filtering to get the saliency object detection box [18]. The work detection effect based on manual characteristic extraction is better than that based on saliency information, but there are still some shortcomings such as insufficient characteristic extraction and poor universality.

To overcome the shortcomings of manual characteristic extraction, a detection model based on deep learning has emerged. Its basic idea is to regard saliency objects as special cases of objects and apply supervised deep learning to achieve object detection algorithm. After the SOD model proposed by Zhang et al. transfers the object detection model multibox, the experimental results show that the detection accuracy is improved by more than 16% compared with the model based on manual characteristic extraction [19].

The SSD model can detect multiple objects with different scales accurately and quickly at the same time. This paper attempts to apply its transfer to saliency object detection. In order to overcome the shortcoming of inaccuracy of SSD model in detecting small size objects, deconvolution module and attention residual module are introduced to increase context information, and then a DAR-SSD model is constructed. The experimental results show that transferring SSD model to saliency object detection is feasible, and the introduction of deconvolution and attention residual module can effectively improve the detection accuracy.

Problem Description

Given a large number of pairs of cervical cell pathological data set $S=\{X_1, X_2, \cdots, X_N\}$ and their saliency object bounding box truth set $Y=\{Y_1, Y_2, \cdots, Y_N\}$, N is the size of the data set. Supposing image X_i contains l saliency objects, and $Y_i=\{[\left(x^i_{j1}, y^i_{j1}\right), (x^i_{j2}, y^i_{j2})]\Big|^l_1\}$ is the true value of saliency object bounding box of image X_i, where $\left(x^i_{j1}, y^i_{j1}\right)$ is the upper left coordinate of saliency object j bounding box and $\left(x^i_{j2}, y^i_{j2}\right)$ is the lower right coordinate [20, 21].

Given the convolution neural network D, input X_i and Y_i can be trained in pairs to obtain the characteristic function $F_{final}(\cdot)$, and then the saliency value object can be predicted for any input image. When process the saliency object detection on image X_i of S, the set of candidate saliency object detection box set $Y'_i=\{[\left(x'^i_{j1}, y'^i_{j1}\right), (x'^i_{j2}, y'^i_{j2}), conf^i_j]\Big|^l_1\}$ is output, where $\left(x'^i_{j1}, y'^i_{j1}\right)$ is the upper left coordinate of the saliency object box, $\left(x'^i_{j2}, y'^i_{j2}\right)$ is the lower right coordinate and $conf^i_j$ is the confidence of the truth saliency object.

SOD Model

The SOD model uses CNN to predict the confidence of saliency objects for 100 prior windows, in which the prior windows are clustered by the real bounding box of the training set. The higher the confidence is, the greater the probability that the prior windows are saliency object detection boxes. Then, a compact group of saliency object detection results are generated by using the optimization formula based on Maximum A Posteriori (MAP).

The model uses VGG network infrastructure, and the initial generated 100 candidate detection boxes can only provide rough positioning of saliency objects. Inspired by attention mechanism, the first M prior windows with high confidence are selected from high to low. According to each selected prior window, the corresponding positions of the original image are double sized sampled and then re-entered into CNN. So the confidence of the M group prior windows is obtained. The top 10 prior windows with the highest confidence are screened for each group, and merged with the first 20 prior windows with the highest initial confidence to form the candidate set of final detection box.

In SOD model, prior windows are directly used as candidate bounding boxes. The obtained detection box is not usually accurate enough, and it only generates the initial candidate detection box on a single scale. When there are multiple saliency objects of different scales in the image, the detection accuracy is not good.

Saliency Detection Model DAR-SSD

Liu et al. proposed SSD model in 2016 based on feedforward convolution network to generate bounding box sets and corresponding class confidence, using Non-Maximum Suppression (NMS) to produce the final results. The model uses multi-scale characteristic images and predicts the default box offset, which makes the detection results more accurate. And it can also predict objects of different scales at the same time. Compared with SOD, it is more suitable for saliency object detection.

Basic Ideas of Saliency Detection Based on SSD Model

Figure 2 shows the network structure of the improved SSD model. The basic network structure SSD Base consists of nine convolution modules, including 2, 2, 3, 3, 3, 4, 2, 2, 2 and 2 convolutions, of which the last convolution simultaneous outputs of 4, 6, 7, 8, 9 convolution modules are used for prediction module. The prediction uses a set of 3×3 convolution kernels to predict the subordinate saliency object for each location of the characteristic pattern and the confidence of two background classes, and the offset is relative to the default box.

The default box is defined as formula (1):

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1), \quad k \in [1, m] \tag{1}$$

where $s_{\min}$ and $s_{\max}$ are set as 0.2 and 0.9, and represent the scale of the lowest layer and the highest layer, respectively; the mesosphere scale is evenly spaced; m is the number of convolution layers used for prediction, with a value of 5. Default box can set different aspect ratio. When aspect ratio is 1, add a default box with scale $s_k' = \sqrt{s_k s_{k-1}}$. The width and height of each default box are $w_k^{\alpha} = s_k \sqrt{\alpha_k}$ and $h_k^{\alpha} = s_k / \sqrt{\alpha_k}$, the center is $((i+0.5)/|f_k|, (j+0.5)/|f_k|)(i, j \in [0, |f_k|])$, and $|f_k|$ is the size of the kth characteristic pattern. Combining all scales and aspect ratios, the default box can generate a prediction covering all sizes and shapes of input.

Let N be the number of default boxes and the objective loss function be the weight sum of confidence loss and location loss:

$$\mathrm{L}(x, c, l, g) = \frac{1}{N}\left(L_{conf}(x, c) - \alpha L_{\mathrm{Loc}}(x, l, g)\right) \tag{2}$$

Location loss is the smoothing loss $L1$ of prediction box l and real bounding box g. It is used for regression center (cx, cy). It is the offset of default box d whose width is 2 and height is h. its definition is shown as follows (3):

$$L(x, c, l, g) = \sum_i \sum_m x_{ij}^k f_{L1}\left(l_i^m - g_j^m\right) \tag{3}$$

where c denotes class; f_{L1} denotes smoothing function and is defined as $\left\|l_i^m - g_j^m\right\|_1^1$, $g_j^{cx} = \left(g_j^{cx} - d_i^{cx}\right)/d_i^{\omega}$,$g_j^{cy} = \left(g_j^{cy} - d_i^{cy}\right)/d_i^h$ and $g_i^{\omega} = \lg\left(\frac{g_j^{\omega}}{d_i^{\omega}}\right) g_i^h = \lg\left(\frac{g_j^h}{d_i^h}\right)$.

Confidence loss is defined as multiple categories of *Softmax* loss, as shown in formula (4):

$$L_{conf}(x, c) = -\sum_{i \in Pos}^{N} x_{ij}^p \lg\left(\hat{c}_i^p\right) - \sum_{i \in Neg}^{N} \lg\left(\hat{c}_i^0\right) \tag{4}$$

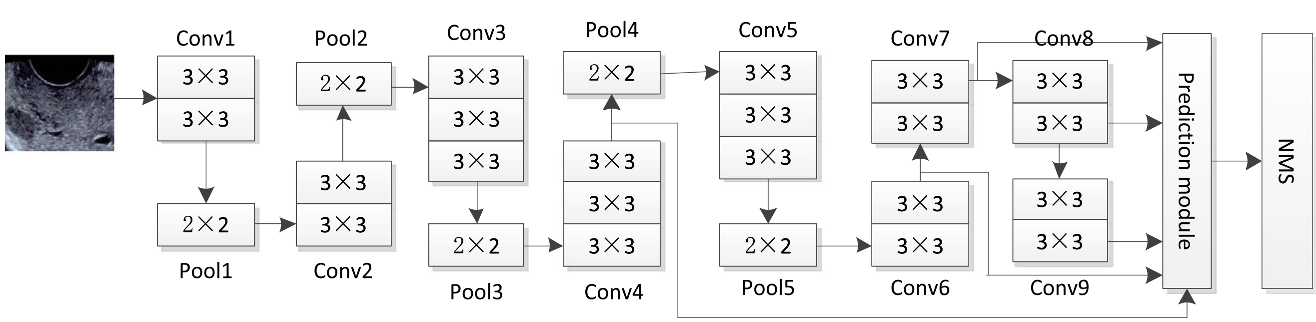

Fig. 2 Network structure of the improved SSD model

where $\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$; if the default box i matches the true bounding box j of category p, then x_{ij}^p is 1, otherwise it is 0.

Deconvolution Block

SSD is sensitive to object scale and poor in small-scale object detection. According to the structure of SSD Base, when the object size is small, its real detection box corresponds roughly to the characteristics of the fourth convolution layer. It should be pointed out that the characteristics of this layer are relatively low-level and contain insufficient semantic characteristics. If we add high-level semantic characteristics to the underlying characteristics, it will help to improve the detection rate of small-scale objects.

Literature [22] proposes that the combination of high-level and low-level characteristic patterns can realize the fusion of high-level semantics and low-level appearance information, and achieve good results in image semantics segmentation. Therefore, the deconvolution module is introduced to deconvolute the high-level characteristic pattern, and to fuse the low-level and high-level semantic characteristics to enhance the context information. Specifically, the SSD Base structure directly uses the characteristic layer which is reduced by two times in length and width at different scales to realize multi-scale object detection. Thus, the convolution, whose convolution kernel size is 4×4 and the step size is 2, is firstly used to deconvolute the high-level characteristic pattern, and to expand its size twice. Further, as shown in the yellow part of fig. 3, the crop layer is added to adapt to any size of input, so that align the size of the deconvoluted characteristic pattern with that of the previous layer; the two characteristic patterns are multiplied pixel by pixel to get the fusion characteristic pattern of high-level semantics and low-level.

Attention Residual Module

Literature [23] points out that add task-oriented sub-networks after size-dependent object detectors can effectively improve the detection accuracy. On the other hand, literature [24] uses spatial pyramid model to realize object detection on VGG network. The visualization results of its fifth convolution layer show that each characteristic pattern has its specific strongest activation region compared with the original image.

Considering that saliency object detection usually uses visual attention mechanism to analyze the saliency of candidate objects, this paper proposes introducing attention residual module between SSD Base and prediction module to enhance the selection of interesting regions. If the strongest activation region of a characteristic pattern is a non-saliency region, reducing its saliency weight is obviously beneficial to saliency object detection. Specifically, as shown in the green part of fig. 3, attention values between 0 and 1 are obtained by convolution with convolution kernel size of 1×1 and siomoid functions. Attention patterns of each convolution characteristic pattern are learned and weighted with the original characteristic pattern.

In addition, by embedding the attention residual module into the deconvolution module, the Deconvolution Attention Residual module (DAR) can be obtained. Its complete structure is shown in fig. 3.

Overall Structure

Aiming at the problem of poor performance in small-scale object detection and lack of saliency information for multi-saliency object detection using SSD model, this paper proposes a DAR-SSD model. Comparing with SSD Base, the final results of the 4th, 6th, 7th, 8th, 9th convolution module in feature convolution layer are output to a DAR module. And the candidate saliency object detection box and confidence of the final output are screened through NMS, as shown in fig. 4.

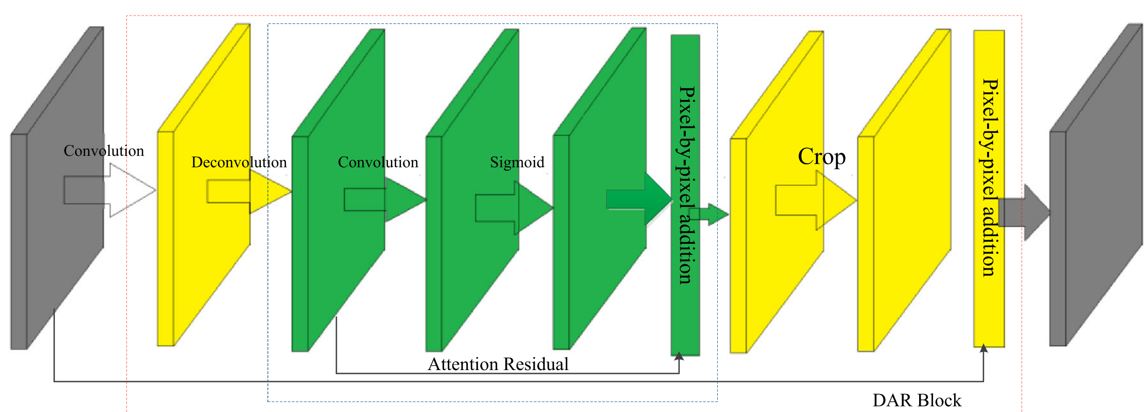

Fig. 3 DAR Block

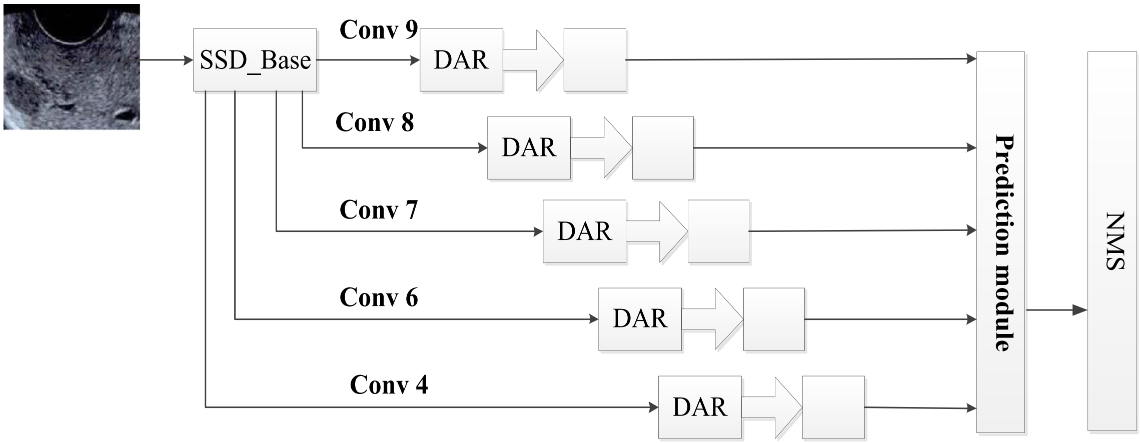

Fig. 4 Overall Structure

Experimental Results and Analysis

Data Set

To verify the effectiveness of our proposed algorithm, a self-built cervical ultrasound dataset is adopted as a training set and there are 5244 images in all, where 1293 normal images with no saliency cervical cancer region; 2573 images with single significant cervical cancer region; The number of images with 2 and 3 significant cervical cancer regions are 777 and 601, respectively.

The testing set is a dataset of the cervical images of the patients we collected at work. Cervical cancer has been diagnosed, with a total of 1224 images containing 338 normal cervical images. In order to facilitate cross-validation, we divide the test data set into four sub-sets, which are named as CCD_A, CCD_B, CCD_C, and CCD_D. In this paper, all test data are manually labeled according to the saliency feature, and the manual cannot be judged, which is not included in the test set; Table 1 summarizes the test sets based on the number and scale of saliency cervical cancer regions, where the small scale is defined as the saliency cervical cancer object whose real bounding box size is smaller than 1/4 of the image.

Experimental Setup

In this paper, subjective comparison analysis of SOD, SSD and DAR-SSD is carried out for different scales and multiple saliency cervical cancer regions, and objective curve analysis is performed with PR curve and Average Precision (AP) as quantitative indexes.

Consistent with the literature [16], when DAR-SSD and SSD are trained, the input image size is 224×224, the initial learning rate is 0.001, the momentum is 0.9, the batch is 32, the weight attenuation is 0.0005, and the training iteration is terminated to 3000 times. At the time of testing, 80 candidate bounding boxes are taken for each of the three models, and the final saliency object bounding box is obtained by NMS and MAP respectively. The setting of MAP value is consistent with the literature [16], and the threshold of NMS was 0.5. In subjective experiments, The confidence threshold is 0.5.

Subjective Comparison Analysis

In order to analyze the effectiveness of the proposed ultrasound cervical cancer diagnosis algorithm, the experimental results are qualitatively compared. Due to space limitations, we only descriptively analyze the experimental results. It can be seen from the experimental results that both SOD and DAR-SSD accurately detected cancer region with small scale, and SSD could not be detected. For some obvious lesion areas, the three methods were successfully detected and recognized without exception, but SOD obtained the lowest confidence. SSD and DAR-SSD accurately detect 3 saliency cervical cancer regions with different scales in the cervix, reflecting the advantages of multi-scale detection, while

Table 1 Dataset statistics

	CCD_A	CCD_B	CCD_C	CCD_D
Number of images with CC	886	954	4890	3930
Number of images with single-object	611	807	4124	1345
Number of images with multi-object	285	138	801	1253
Number of images with small-scale	446	325	3065	3102
Number of images with large-scale	424	635	1825	920

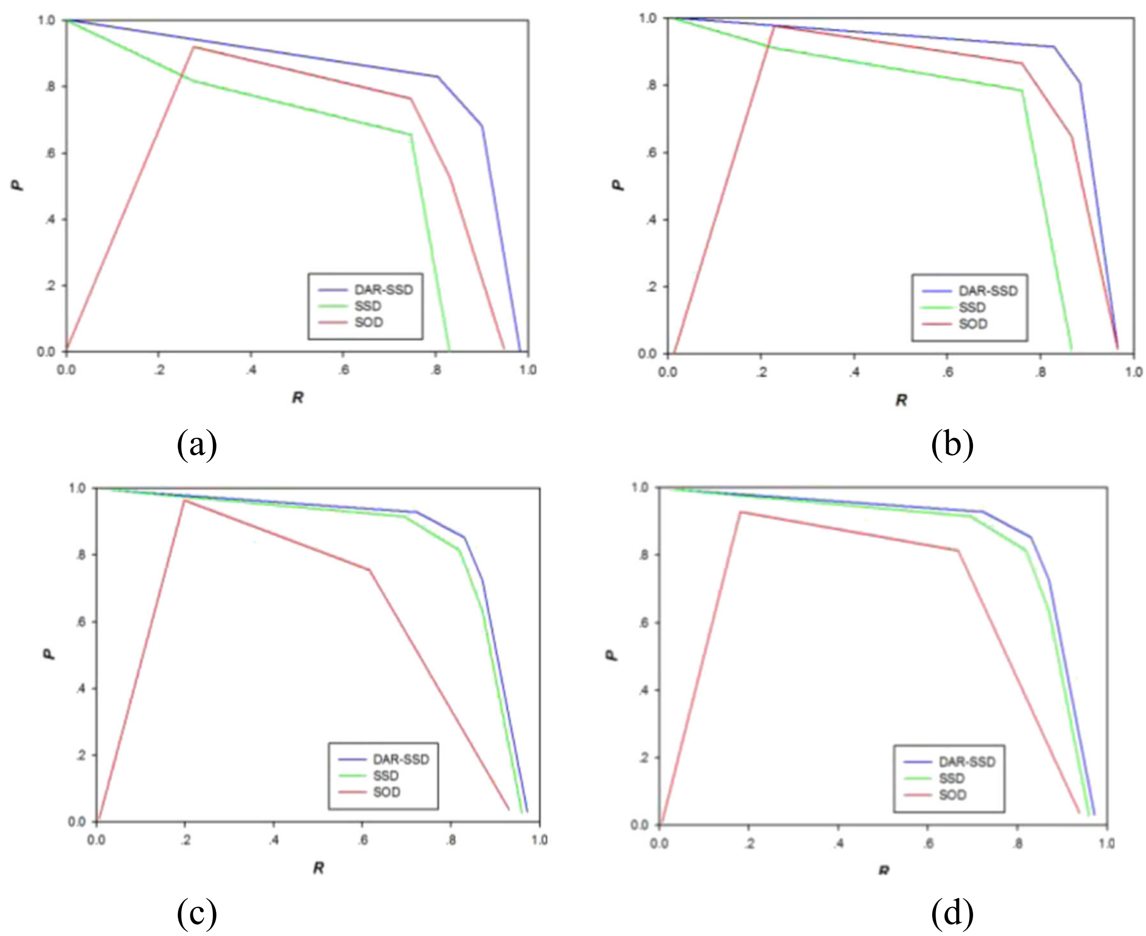

Fig. 5 Comparison of PR curve under NAP methoda)CCD_A; (b)CCD_B; (c)CCD_C; (d)CCD_D

SOD lost the object of the smallest cervical cancer, and the deviation of the bounding box is large compared with the benchmark position.

In summary, DAR-SSD inherits the advantages of the original SSD model to simultaneously detect multiple different scale objects. On the other hand, the introduction of DAR module not only improves the shortcomings of the original SSD model to detect small-scale objects, but also improve the position reliability and accuracy through the attention mechanism, where it performs well under different conditions, and the bounding box is the most accurate.

Analysis of Quantitative Experimental Results

Figure 5 and Fig. 6 show the PR (Precision-Recall) curves obtained when the MAP and NMS are used in the three models. The corresponding AP values are shown in Table 2 and Table 3. The AP values of the three models are consistent. The performance of DAR-SSD and SSD on PR and AP is significantly better than SOD, which is nearly 10% higher than that of NMS. It is shown that the original object detection is still good performance after SSD transfer application, and its ability to simultaneously detect multi-scale objects is a better performance than SOD.

As shown in Table 2 and Table 3, compared with NMS, the performance of SOD is significantly improved when using MaP, and the SSD is improved by an average of about 1%, but the increase of DAR-SSD is not obvious. Since the SSD detects each position on the feature map and outputs it, the SOD re-detects the first five detection region of the confidence level, and the obtained detection frame is far more dense than the SSD, so the performance of the MAP to the SOD is significantly stronger than the other two. In addition, when the noise is high, the MAP detection performance is better than that of the NMS. When the bounding region is purer and the noise is less, the relative advantage of the MAP is weakened. The DAR-SSD detection bounding box is the most accurate, and the SSD is second. Therefore, the performance improvement of MAP under DAR-SSD is the weakest, followed by SSD.

In order to further highlight the improvement of the DAR module in DAR-SSD, this paper conducted an AP comparative analysis of different cases of small-scale cervical cancer, as shown in Table 4. In the case of single cervical-cancer

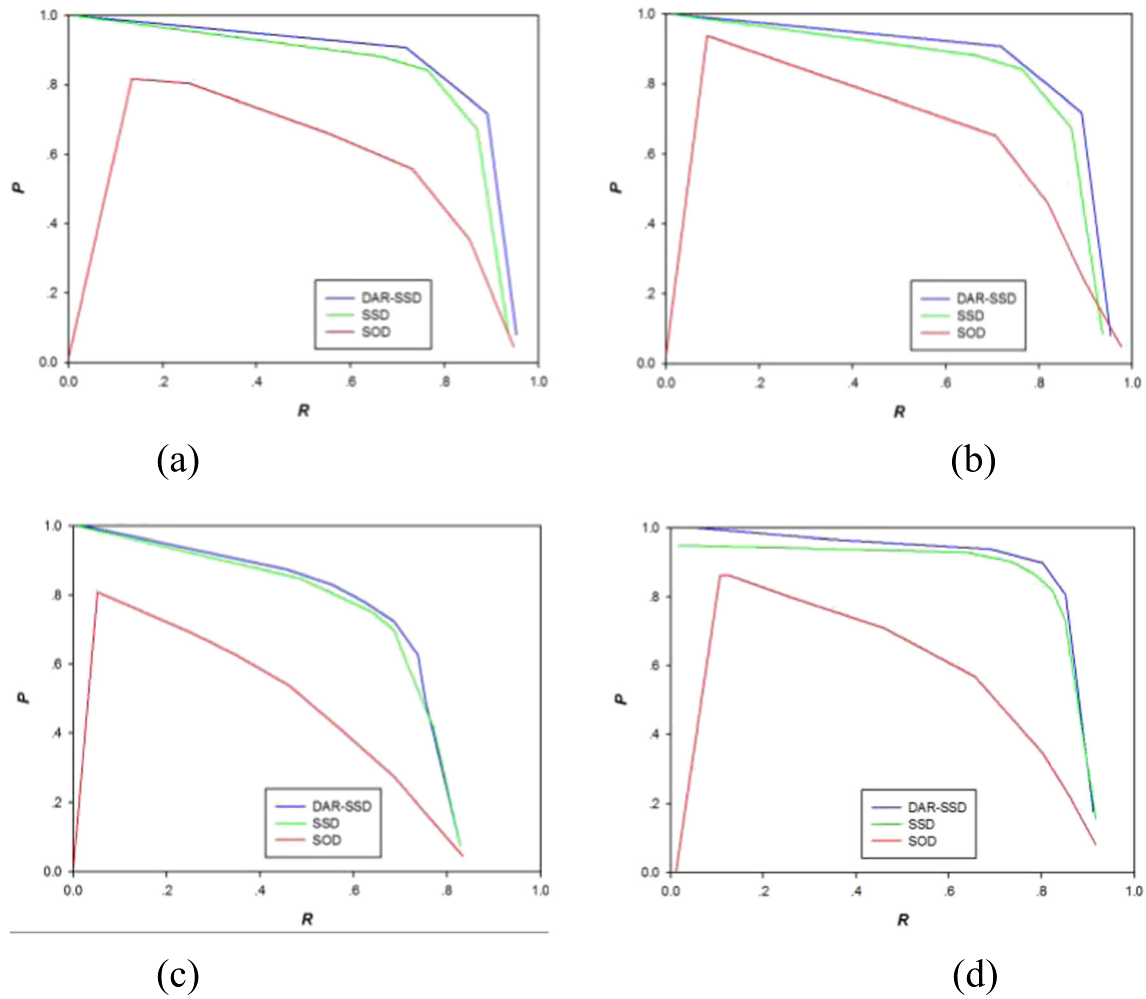

Fig. 6 Comparison of PR curve under NMS method; (a)CCD_A; (b)CCD_B; (c)CCD_C; (d)CCD_D

object, DAR-SSD has similar performance to SSD on CCD_A and CCD_D, and is superior to SSD in CCD_B and CCD_C, and NMS is increased by more than 1.5%. In fact, the first two data sets are simple, and the latter two data sets contain a large number of complex background. In addition, DAR-SSD introduces the attention mechanism through the attention residual module, and can learn more relevant features of the region of interest. It helps to better detect saliency cervical cancer in complex contexts.

In the case of multi-objects, DAR-SSD has a 1.5% improvement in all four data sets; in the case of small-scale saliency objects, DAR-SSD performance improvement is more obvious, at least 2.4% improvement in 4 data sets. The experimental results show that the deconvolution module fuses the high-level semantics with the underlying features, which helps to improve the detection accuracy of small-scale cervical-cancer objects. In the case of large-scale saliency cervical-cancer objects, there is no difference in performance between DAR-SSD and SSD. The experimental results show that the deconvolution module has small-scale object sensitivity and does not reduce the detection performance of large-scale objects for transvaginal ultrasound elastography.

Table 2 AP comparison under MAP method

	CCD_A	CCD_B	CCD_C	CCD_D	Means
SOD	0.7336	0.8152	0.5476	0.6938	0.6976
SSD	0.8229	0.8654	0.6698	0.8165	0.7934
DAR-SSD	0.8361	0.8766	0.6774	0.8317	0.8054

Table 3 AP comparison under NMS method

	CCD_A	CCD_B	CCD_C	CCD_D	Means
SOD	0.6104	0.7157	0.4409	0.5865	0.5871
SSD	0.8120	0.8619	0.6585	0.7974	0.7825
DAR-SSD	0.8387	0.865	0.6738	0.8256	0.8025

Table 4 Comparative analysis for small-scale cervical cancer

	CCD_A	CCD_B	CCD_C	CCD_D
SSD + NMS	0.7984	0.7719	0.5734	0.7889
DAR-SSD + NMS	0.8285	0.7951	0.5980	0.8117
SSD + MAP	0.8052	0.7844	0.5762	0.8035
DAR-SSD + MAP	0.8310	0.8065	0.6001	0.8125

Conclusion

Traditional saliency cervical cancer detection methods in Ultrasound image, assuming that there is only one salient object, is not conductive to practical application. Their effects are dependent on saliency threshold. Object detection model provides a kind of new solutions. Shot multiBox detector can accurately detect multi-objects with different scales simultaneously except for small cervical cancer regions. To overcome this drawback, this paper presents a new multi-saliency objects detection model, appending deconvolution module embedded with in attention residual module. Experiments show that our proposed diagnosis algorithm achieves higher detection accuracy than comparison algorithms. Also, it improves detection performance for mult-saliency cervical cancer objects on small scales, compared with original shot multiBox detector, and it has an advantage over complicated background, compared with MDF and DCL, which also are deep model based methods. In future work, we will optimize the algorithm and embed the module into the ultrasound equipment to feedback the diagnosis results in real time and accurately so as to improve the automation level of ultrasound examination.

Compliance with Ethical Standards

Conflict of interest We declare that we have no conflict of interest.

Human and animal studies The paper does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Lewis, G. C., Howson, et al. Cervical cancer detection.[J]. Medical Journal of Australia 1(13):708–710, 2017.
2. Koss, L. G., The Papanicolaou test for cervical cancer detection. A triumph and a tragedy[J]. International Journal of Gynecology & Obstetrics 30(1):79, 2016.
3. Reinhardt, M. J., Ehritt-Braun, C., Vogelgesang, D. et al., Metastatic Lymph Nodes in Patients with Cervical Cancer: Detection with MR Imaging and FDG PET1[J]. Radiology 218(3):776–782, 2001.
4. Zhang, J., Malmberg, F., and Sclaroff, S., Unconstrained Salient Object Detection.
5. Kaijian, X. et al., Optimization of diagnosis and treatment of chronic diseases based on association analysis under the background of regional integration. *J. Med. Syst.* 43:46, 2019.
6. Pengjiang, Q. et al., Knowledge-leveraged transfer fuzzy C -Means for texture image segmentation with self-adaptive cluster prototype matching. *Knowl.-Based Syst.* 130:33–50, 2017.
7. Zhang, J., et al. Salient Object Subitizing. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE Computer Society, 2015.
8. Cheng, M. M., et al. Global Contrast Based Salient Region Detection. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on IEEE, 2011.
9. Yan, Y. et al., Unsupervised Image Saliency Detection with Gestalt-laws Guided Optimization and Visual Attention Based Refinement. Pattern Recognition 32(12):23–41, 2018.
10. Li, G., and Yu, Y., Visual Saliency Detection Based on Multiscale Deep CNN Features. IEEE Transactions on Image Processing 25(11):5012–5024, 2016.
11. Klaes, R., Woerner, S. M., Ridder, R. et al., Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes.[J]. Cancer Research 59(24):6132–6136, 1999.
12. Benedet, J. L., Anderson, G. H., and Matisic, J. P., A comprehensive program for cervical cancer detection and management[J]. American Journal of Obstetrics and Gynecology 166(4):1254–1259, 2016.
13. Zhang, J., Liu, Y., Cervical Cancer Detection Using SVM Based Feature Screening[C]// Medical Image Computing & Computer-assisted Intervention-miccai, International Conference Saint-malo, France, September. DBLP, 2004.
14. Liu, F., and Gleicher, M., " [ACM Press the 18th annual ACM symposium - Seattle, WA, USA (2005.10.23–2005.10.26)] Proceedings of the 18th annual ACM symposium on User interface software and technology, − UIST \"05-Automatic image retracking with fisheye-view warping." (2005):153–175.
15. Valenti, R., Sebe, N., Gevers, T., Image saliency by isocentric curvedness and color[C]// IEEE Conference on Computer Vision. 2009.
16. Luo, Y. et al., Saliency Density Maximization for Efficient Visual Objects Discovery. IEEE Transactions on Circuits and Systems for Video Technology 21(12):1822–1834, 2011.
17. Feng, Jie, et al. Salient object detection by composition. 2015.
18. Yildirim, G. and Süsstrunk, S., FASA: Fast, Accurate, and Size-Aware Salient Object Detection. 2014.
19. Ren, S. et al., Object Detection Networks on Convolutional Feature Maps. IEEE Transactions on Pattern Analysis & Machine Intelligence 39(7):1476–1481, 2015.
20. Redmon, J., et al. You Only Look Once: Unified, Real-Time Object Detection. 2015.
21. Ren, S., et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. 2015.
22. Long, J., Shelhamer, E., and Darrell, T., Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence 2014.
23. Zhou, Q. et al., Multi-scale deep context convolutional neural networks for semantic segmentation. World Wide Web 22(2):555–570, 2019.
24. He, K. et al., Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence 37(9):1904–1916, 2014.