

A Method for Localization and Classification of Breast Ultrasound Tumors

Wanying Mo^{1((\Box)}, Yuntao Zhu², and Chaoyun Wang²

¹ Computer Science and Technology College, Jilin University, Changchun 130012, China mowanying@163.com

² School of Automation, Harbin Engineering University, Harbin 150000, China

Abstract. Ultrasound instruments are suitable for large-scale examination of breast tumors, especially for women from Asian whose glands are dense. However, ultrasound images have the low contrast and resolution, blurred boundary and artifacts, which bring great difficulties to the interpretation of the junior doctor. However, traditional methods of breast ultrasound tumor recognition often use manually extracted features to gradually realize ROI region location and tumor classification with low accuracy, poor robustness and weak universality. Deep learning is limited to the location of tumor ROI region or the classification of a given tumor ROI region. In this paper, YOLOV3 algorithm is used for breast ultrasound tumor recognition, which could realize ROI localization and tumor classification at the same time. In addition, K-Means is optimized by K-Means++ and K-Mediods algorithm to generate anchor boxes of YOLOV3, and based on the Darknet-53 network structure of YOLOV3, ResNet and DenseNet are combined to design ResNet-DenseNet_Darknet-53. The proposed method is tested on the breast ultrasound tumor data set. Experiments show that the improved YOLOV3 algorithm shows better detection results on multiple evaluation indicators.

Keywords: Breast tumor recognition · YOLOV3 · Anchor boxes · ResNet-DenseNet_Darknet-53

1 Introduction

According to the American Cancer Statistics of 2016 published by the American Cancer Society, breast cancer is the most common malignant tumor in women and is expected to monopolize 29% of all new women's cases [1]. Therefore, screening and early diagnosis of breast cancer are very necessary. At this stage, the main methods to diagnose breast cancer are MRI, PET, ultrasound imaging and X-ray, Compared with MRI, PET and X-ray, ultrasound imaging technology is not only inexpensive and non-radioactive, but also can show most of the lesion areas in each section, especially for more compact breast tissues of young Asian women [2, 3].

Although ultrasound imaging has become the main tool for early diagnosis, it still faces great challenges:

(1) Ultrasound images have a lot of noise;

© Springer Nature Switzerland AG 2020 Y. Tan et al. (Eds.): ICSI 2020, LNCS 12145, pp. 564–574, 2020. https://doi.org/10.1007/978-3-030-53956-6_52

- (2) the boundaries in the ultrasound image is blurred;
- (3) Because the acoustic impedance of different human tissues and organs has large differences, it is easy to cause artifacts in the ultrasound image [4].

In order to solve this problem, researchers have proposed different methods.

2 Related Work

In recent years, with the rise of deep learning, some researchers have proposed a method based on deep learning for tumor recognition. But the methods of breast cancer recognition based on deep learning are basically limited to use image segmentation to locate ROI region or classify a given ROI region as benign or malignant, but they can not simultaneously locate ROI region and classify tumors, which brings great inconvenience to the diagnosis of medical staff. Recently, Yap et al. and Chiao et al. proposed to use improved FCN-AlexNet and improved Mask R-CNN semantics segmentation model to realize end-to-end ultrasound tumor recognition [5, 6], which locate and classify tumors at the same time, not only accurately locate the boundary of ultrasound tumors, but also realize the recognition of benign and malignant tumors. However, this method of image semantics segmentation requires manual labeling of a large number of tumor segmentation data sets, and the process of making the segmentation data sets is very cumbersome, which greatly increases the burden of professional doctor labeling.

In this paper, we propose a target detection algorithm based on deep learning for breast cancer recognition based on YOLOV3[7], which is the best target detection algorithm among all the other methods and could realize ROI localization and tumor classification at the same time. In addition, K-Means [8] is optimized by K-Means++ [9] and K-Mediods [10] algorithm to generate anchor boxes of YOLOV3 to solve the problem of instability of the initial center point and the problem of outlier sensitivity, respectively, and then based on the Darknet-53 network structure of YOLOV3, ResNet and DenseNet [11] are combined to design ResNet-DenseNet_Darknet-53 to solve the problem that features of ultrasound image of breast tumor are more difficult to extract than other images.

3 Method

3.1 Traditional YOLOV3 Algorithm

The YOLOV3 algorithm combines the tasks of classification and location into a step to directly predict the position and category of objects. It includes a new feature extraction network Darknet-53 and three scales of YOLO layer, which are used for feature extraction and multi-scale prediction, respectively. Its network structure is shown in Fig. 1 [11].

The Darknet-53 layer consists of one convolution block DBL and five residual blocks resn (n = 1, 2, 4, 8). DBL is a collection of convolution (conv), batch normalization (BN), and activation function (Leaky relu), which is the smallest component in YOLOV3. After feature extraction from Darknet-53 network, output the feature map with size of 13 * 13 * 1024. After upper sampling and shallow feature map splicing (see concat



Fig. 1. YOLOV3 network architecture

in Fig. 1) [12], the feature map with three scales is output for predicting the results of YOLO layer detection. That is to say, each box is responsible for the regression of three anchor boxes.

3.2 Optimization of Anchor Box

The anchor box is used as a priori box to predict the target boundary and is designed according to different data sets. The existing YOLOV3 uses K-Means algorithm to cluster the annotated data of COCO dataset. First, it reads the data set file of PASCLA VOC format to get the category and border of the target. Then, it normalizes the border data to get the width *w* and height *h* of the target, which is used as the data to be clustered. The normalization process is shown in the formulas (1-1) and (1-2)where x_{\min} , y_{\min} and x_{\max} , y_{\max} represent the upper left corner coordinates and the lower right corner coordinates of the border, respectively. *width* and *height* represent the image size, respectively.

$$w = \frac{x_{\max} - x_{\min}}{width}$$
(1-1)

$$h = \frac{y_{\text{max}} - y_{\text{min}}}{height} \tag{1-2}$$

After that, nine clustering centers are initialized, each of which describes the length and width of the rectangular box. Computing IOU (the intersection and parallelism ratio) of rectangular boxes described by clustering center and data to be classified, using distance D as clustering basis, updating clustering center by means of in-cluster mean, and getting the final anchor box by optimizing the minimum clustering loss function J. The IOU expression is (1-3), the distance calculation formula D is (1-4), and the loss function J is (1-5).

$$IOU = \frac{\text{area (box \cap centroid)}}{\text{area (box \cup centroid)}}$$
(1-3)

$$D = 1 - IOU_{(box, centriod)}$$
(1-4)

$$J = \sum_{k=1}^{k} \sum_{i=1}^{n_k} D_i$$
 (1-5)

where, 'box' represents the real target box, 'centroid' the candidate box, J the clustering loss function, K the number of categories, and n_k the number of sample points in a cluster.

In order to make the network better to learn the location and feature size of the target, this paper uses K-Means++ algorithm and K-Mediods algorithm to optimize the original K-Means algorithm to obtain a merger anchor box with higher intersection. The specific steps of the improved algorithm based on K-Means are as follows (where K = 9):

Step 1: Obtain m samples $(Q_m = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\})$ which are clustered with normalized pretreatment of breast ultrasound tumor data sets;

Step 2: Select a sample $c_i = (x_i, y_i)(i \in [1, m])$ randomly from the data set as the clustering center;

Step 3: Calculate the distance d_j between each sample point c_j and c_i and the probability p_j according to the formulas in Formulas (1-6) and (1-7), respectively, where $j \in [1, m]$ and $j \neq i$;

$$d_j = 1 - IOU_{(c_j, c_i)}$$
(1-6)

$$p_j = d_j^2 \tag{1-7}$$

Step 4: Select the sample points with the largest probability p_j as the new clustering center.

Step 5: Repeat step 2 and step 3 until K cluster centers are selected. The corresponding K clusters are generated at the same time. The K-Means++ algorithm process is over.

Step 6: Calculate the distance \hat{d}_j from center c_j of each sample in Q_m to cluster centers n_1, n_2, \ldots, n_k , according to formula (1-6), corresponding samples to minimum distance are assigned to M_i in the cluster to form K clusters;

Step 7: Calculate the distance from each sample point \hat{x}_i ($i \in (1, n_i)$, n_i denotes the number of all sample points) to all other sample points in each M_i according to formula (1-8), and use the sample point with the smallest distance to all other sample points as a new clustering center to update K clustering centers ($n_1, n_2, ..., n_k$)

$$n_i = \underset{\hat{x}_i \in M_i}{\arg\min} \sum_{\hat{y} \in M_i} \left(1 - IOU(\hat{x}_i, \hat{y})\right)$$
(1-8)

Step 8: If the objective function E in formula (1-9) does not change, it will be optimal, otherwise, the algorithm will repeat step 7 and step 8.

$$E = \sum_{i=1}^{K} \sum_{j=1}^{m} (1 - IOU(n_i, c_j))$$
(1-9)

where n_i denotes the number of sample points in cluster *i*.

3.3 Design of ResNet-DenseNet_Darknet-53

In this section, based on the existing Darknet-53 network structure, we use ResNet and the DensetNet to design a method named ResNet_DenseNet-Darknet-53 for recognizing ultrasound images.

(1) ResNet-DenseNet_Darknet-53

In traditional Darknet-53, when the feature map of the previous layer is input to resn (n = 1, 2, 4, 8), a conv3 down-sampling operation is first performed, and then transmitted to a network containing n (n = 1, 2, 4, 8) residuals Res to obtain the final output. The output of each residual network Res is constructed by feature fusion of a conv1 and a conv3 with short connections. But unlike Inception and other networks, this fusion does not use tensor splicing. It directly adds the pixels of each channel, so the number of channels will not be changed, as shown in Fig. 2.



Fig. 2. Structural diagram of Res network

In order to describe the ResNet-DenseNet_Darknet-53 conveniently, this paper takes res4 as an example to explain this process in detail. res4 consists of a conv3 and four Res. And the size of the input feature graph of res 4 is 26 * 26 * 512. First, a conv3 is used to downsample the feature graph to make its size 13 * 13 * 1024. Then, the output result is obtained after four consecutive Res processing of the feature graph. Because the corresponding channel pixels are directly added in Res, the input and output sizes remain to be unchanged after each Res, as shown in Fig. 3.



Fig. 3. Structural diagram of res4 network

After introducing ResNet and DensetNet, the improved res4 network is shown in Fig. 4. The size of input feature graph is 26 * 26 * 512. After conv3 down-sampling, it becomes 13 * 13 * 1024. It should be pointed out that the number of conv1 and conv3 in each Res in the improved res4 is 256, so the output dimension of each Res is 256. Now all Res outputs are connected in pairs, so that each layer of the network output accepts the characteristics of all the layers in front of it as input. After the output of each residual network Res is spliced by features, the output feature maps are 13 * 13 * 1280, 13 * 13 * 1536, 13 * 13 * 1792 and 13 * 13 * 2048, respectively. Considering the output of the last residual network Res, the dimension of the feature map after feature splicing

is twice as large as that of the feature map after conv3 downsampling. Drawing on the idea of pixel addition of the corresponding channel of ResNet, the feature map of the output after the first conv3 downsampling and the feature map after the conv1 feature reduced dimension are added to the corresponding channel pixels (as shown by the green line in the figure). Finally, a feature graph of 13 * 13 * 1024 size is output.



Fig. 4. Structural diagram of improved res4 network

Since res1 contains only one Res, no improvement is needed. Similarly, other resns (n = 2, 8) are improved according to the idea of res4. The difference is that the number of conv1 and conv3 in each Res is different. The number of the first res2, the first res8, the second res8 and the first res4 are 64, 32, 64 and 256, respectively. Thus, we construct a new Darknet network based on the ResNet and the DenseNet.

4 Experiments and Results Analysis

4.1 Data Set

The breast ultrasound data used in the experiment have been screened and eliminated from the 2011 original cases of different ages, regions and ethnicities by professional doctors of Harbin Medical University from 2015 to 2019. There are 13,586 original images with a size of 768*576. The effective images are 3259 pictures containing tumors, including 1199 benign patients and 2060 malignant patients. The data set is directly divided into training set and test set according to 4: 1. As shown in Table 1, there are 2608 training sets, including 960 benign training sets and 1648 malignant training sets. There are a total of 651 test sets, including 239 benign tests and 412 malignant tests.

4.2 Anchor Box Calculation

The average cross-and-merge ratio (Avg-IOU) is used as the evaluation index (the average of the cross-and-merge ratio between the calculated anchor box and all data). In order to improve the coincidence between the priori box and the data set, and to verify the effectiveness of the algorithm, eight experiments are repeated. As shown in Fig. 5, the maximum Avg-IOU of the original method is 0.8322, and the maximum Avg-IOU of this method is 0.8421, which improves the coincidence between anchor box and data set, and makes tasks easier to learn.

	Benign	Malignant	Total
The training set	960	1648	2608
The testing set	239	412	651
Total	1199	2060	3259

 Table 1. Breast ultrasound data set partition



Fig. 5. Avg-IOU of traditional K-Means and improved K-Means

Anchor box obtained from the largest cluster experiment of Avg-IOU is used as a priori box in this paper. The results of clustering experiments are shown in Fig. 6. Different clusters are marked with different colors and clustering centers are drawn with black forks. After inverse transformation, nine anchor boxes are obtained as follows:(52,51), (79,72), (93,102), (122,94), (125,135), (159,117), (160,164), (190,225), (204,166).

4.3 Results Analysis Based on Index

In this paper, Faster R-CNN [13] algorithm and YOLOV3 algorithm with different configurations are selected for 1000 training sessions. The different network configurations are shown in Table 2. YOLOV3 (1) is the original YOLOV3 algorithm without optimizing anchor box and using ResNet-DenseNet_Darknet-53; YOLOV3 (2) is the YOLOV3 algorithm with introducing anchor box optimization; YOLOV3 (3) is the YOLOV3 algorithm with optimizing anchor box and using ResNet-DenseNet_Darknet-53 at the same time. In order to accelerate the training speed of the network and prevent over-fitting, this paper uses Adma algorithm to do gradient optimization. The initial learning rate is set to 0.001. After every 100 iterations, the learning rate decreases to 1/10 of the original, the impulse is 0.9, the attenuation coefficient is 0.0002, and the batch_size is 100. Then, the performance of the model is evaluated on the test set according to multiple evaluation indicators.



Fig. 6. Clustering experiment

	Faster R-CNN	YOLOV(1)	YOLOV (2)
Optimizing anchor box	×	×	\checkmark
Using ResNet-DenseNet_Darknet-53	×	×	×

By setting different confidence thresholds, AP of benign samples and malignant samples on training set and test set can be calculated, respectively. Compared with Fig. 10 and Fig. 11, the AP value of malignant samples is higher than that of benign samples, whether in training set or test set. This is Because the number of malignant samples in the training set is larger than that of benign samples.

From Fig. 7 and Fig. 8, it can be seen that the positive and malignant AP of YOLOV3 (1) algorithm are 87.85% and 90.25% respectively in the training set and 83.01% and 88.25% respectively in the test set, which are superior to Faster R-CNN algorithm, showing that the performance of YOLOV3 algorithm is better. The benign and malignant AP of YOLOV3 (2) algorithm is 2.03% and 2.76% higher than that of YOLOV3 (1) in training set, and 2.67% and 2% higher than that of YOLOV3 (1) in test set. This is because in YOLOV3 (2) algorithm, better anchor box can be obtained after clustering optimization. Compared with YOLOV3 (2) algorithm, the benign and malignant AP of YOLOV3 (3) algorithm is improved by 0.71% and 2.2% in training set and 1.37% and 3.22% in test set, respectively.

After calculating the AP of benign and malignant samples, the mAP of the model can be obtained. As shown in Fig. 9, in the four algorithms, the mAP on the training



Fig. 7. Benign AP on training set and test set



Fig. 8. malignant AP on training set and test set

set is higher than that on the test set, which is consistent with the reality. The mAP of YOLOV3 (1) algorithm is higher than that of Faster R-CNN in training set and test set. Compared with YOLOV3 (1) algorithm, the mAP of YOLOV3 (2) algorithm increased by 2.4% and 1.42% in training set and test set, indicating the effectiveness of optimizing anchor box. Compared with YOLOV3 (2), the mAP of YOLOV3 (3) algorithm increases by 1.45% and 2.25% in training set and test set respectively.

4.4 Results and Analysis Based on Prediction

One malignant samples are randomly selected from database and the above four algorithms are run, respectively. As shown in Fig. 10, the first image is the original image, and the second image is a label image labeled sample by a professional doctor. The blue area in the image is the location of the tumor. The third, fourth, fifth and sixth pictures are the results of running Faster R-CNN, YOLOV3 (1), YOLOV3 (2) and YOLOV3 (3), respectively. The four algorithms in Fig. 10 identify it as malignant with confidence levels of 87.25%, 95.91%, 96.72%, and 97.85%, respectively.



Fig. 9. Benign and malignant mAP on training and test set



Fig. 10. Recognition effect under four different algorithms



Fig. 11. Recognition results under different algorithms

Then, one benign samples are randomly selected from database, and the above four algorithms are run, respectively. As shown in Figs. 11, the order of the six images is

the same as that in Fig. 10. Figures 11 shows that all the four algorithms recognize it as benign, with confidence of 71.79%, 89.39%, 92.51% and 95.45% respectively. Obviously, compared with the first three algorithms, the YOLOV3 (3) algorithm after anchor box optimization and using ResNet-DenseNet_Darknet-53 is obviously better than the first three algorithms.

5 Conclusions

In this paper, YOLOV3 algorithm is proposed to classify benign and malignant tumors and locate ROI regions simultaneously. It improves YOLOV3 algorithm by optimizing anchor and designing ResNet_DenseNet-Darknet-53 for related questions. Experiments show YOLOV3 algorithm after optimizing anchor and using ResNet_DenseNet-Darknet-53 achieves the best results. The proposed method not only realizes the localization of ROI region and the classification of benign and malignant tumors, but also achieves good detection results. It makes the application of artificial intelligence closer to the breast cancer detection in actual operating environment.

References

- Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2016. CA Cancer J. Clin. 66(1), 7–30 (2016)
- Ikedo, Y., Morita, T., Fukuoka, D., et al.: Automated analysis of breast parenchymal patterns in whole breast ultrasound images: preliminary experience. Comput. Assist. Radiol. Sur. 4(3), 299–306 (2009). https://doi.org/10.1007/s11548-009-0295-0
- 3. Huang, Y.L., Jiang, Y.R., Chen, D.R., et al.: Computer-aided diagnosis with morphological features for breast lesion on sonograms. Ultrasound Obstet. Gynecol. **32**(4), 565–572 (2008)
- Ting, X., Lei, L., Kai, L., et al.: Comparison of transferred deep neural networks in ultrasonic breast masses discrimination. Biomed. Res. Int. 2018, 1–9 (2018)
- Yap, M.H., et al.: End-to-end breast ultrasound lesions recognition with a deep learning approach. Proc. SPIE 10578, 1057819 (2018). https://doi.org/10.1117/12.2293498
- Chiao, J.-Y., et al.: Detection and classification the breast tumors using mask R-CNN on sonograms. Medicine 98(19), pe15200 (2019)
- 7. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement (2018)
- 8. Jain, A.K.: Data Clustering: 50 years beyond k-means (2008)
- 9. Arthur, D.: K-Means++: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. ACM (2007)
- Zhou, Z., Si, G., Chen, J., et al.: A novel method of transformer fault diagnosis based on K-Mediods and decision tree algorithm. In: IEEE 2017 1st International Conference on Electrical Materials and Power Equipment (ICEMPE), Xian, China 2017 1st International Conference on Electrical Materials and Power Equipment (ICEMPE), pp. 369–373 (2017)
- Gao, H., Zhuang, L., Maaten, L.V.D., et al.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision & Pattern Recognition (2017)
- Zhao, Y., Zhu, H., Shen, Q., et al.: Practical adversarial attack against object detector. 24 (2018)
- 13. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: NIPS (2015)