

BREAST CANCER DETECTION BASED ON MERGING FOUR MODES MRI USING CONVOLUTIONAL NEURAL NETWORKS

Wenhuan Lu^{*} Zhe Wang^{*} Yuqing He[†] Hong Yu[‡] Naixue Xiong[§] Jianguo Wei^{*}

^{*} College of Intelligence and Computing, Tianjin University, Tianjin, China.

[†] School of Electrical and Information Engineering, Tianjin University, Tianjin, China.

[‡] Center Obstetrics and Gynecology Hospital, Tianjin, China.

[§] Department of Computer Science, Colorado Technical University, Colorado, USA

ABSTRACT

The objective of the study is to develop a framework for automatic breast cancer detection with merging four imaging modes. Attempts were made for tumor classification and segmentation; using a multi-parametric Magnetic Resonance Imaging (MRI) method on breast tumors. MRI data of the breast were obtained from 67 subjects with a 1.5T-MRI scanner. Four imaging modes: were T1 weighted, T2 weighted, Diffusion Weighted and eTHRIVE sequences, and dynamic-contrast-enhanced(DCE)-MRI parameters are acquired. The proposed four-mode linkage backbone in tumor classification, which overcomes the limitations of single-modality image detection and simulates actual diagnosis processes by clinicians, achieves the accuracy of 0.942. The proposed automatic segmentation approach is performed by a refined U-Net architecture, and the result improved segmentation performance significantly. The combination of four-mode linkage classification backbone and improved segmentation network for breast cancer detection forms a computer-aided detection (CAD) system that corresponds to the actual clinical diagnosis work.

Index Terms— four-mode linkage, classification, convolutional neural network, segmentation, MRI

1. INTRODUCTION

Breast cancer is medically known as a common cancer with a high mortality rate, relative to other types of cancer. Early detection and morphological analysis of tumors are critical factors in the diagnosis and treatment of breast cancer, which helps increase the survival rate to a certain extent. Techniques for clinical imaging of the mammary glands include X-ray mammography (mammography), color Doppler ultrasonography, and Magnetic Resonance Imaging (MRI). MRI becomes the most important method for the detection of breast cancer. In this paper, automatic detection with classification and segmentation of breast cancer is based on multi-parametric MR images.

Despite recent advances in computer vision, breast cancer diagnosis continues to rely heavily on visual inspections conducted by experienced pathologists and radiologists [1]. The detection performance may vary due to subjective observations by radiologists and clinicians. Through Computer Aided Diagnosis (CAD) systems, examiners usually study images from a single sequence imaging mode at a time. This is different from clinical practice in which a radiologist observes several types parameters of several imaging modes of MRI breast images simultaneously. We propose another detection backbone based on four sets of images with different imaging modes (T1W, T2W, DWI, and SYN). In the radiology report, clinicians also need to accurately describe the location, shape, and size of the tumor, which will become important foundation for subsequent treatments with chemotherapy or surgical resection. Extracting these features of tumors and finding lesion margins need accurate tumor segmentation. It is often a challenging task due to poor signal-to-noise-ratio and faint edges caused by partial volume effects [2]. Therefore, tumor delineation is affected by variations and uncertainties in the automated lesion segmentation process. These variations of the intensity, texture, and morphology of the tumor lead to inconsistency in diagnostic outcome. Thus, we proposed the improved segmentation model.

2. EXPERIMENTAL METHODOLOGY

In this paper, a new detection backbone with tumor classification and segmentation tasks is proposed. The block diagram of the detection system is shown in Fig. 1. We put forward the four-mode linkage backbone based on the use of Convolutional Neural Networks (CNNs) to simulate the clinical diagnosis, and then explore the segmentation. The upsampling process in the segmentation network is refined for obtaining clear edges of tumors. Meanwhile, using classified network to initialize the segmentation work can better integrate these two related tasks in detection of breast cancer. Once the lesion area is detected and correctly classified, it is automatically segmented and formed into a complete set of diagnosis

system to facilitate the clinician’s further diagnosis of the patient’s breast tumor.

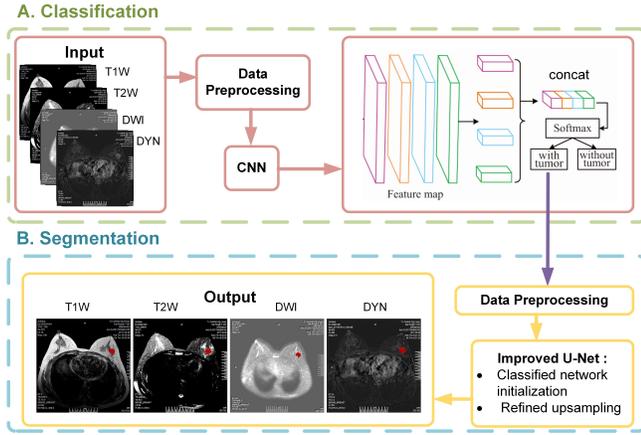


Fig. 1. The breast cancer detection system based on four-mode linkage classification backbone and improved segmentation network system.

The dataset used in our experiment is an MRI database built at the radiology department of the Tianjin Center Obstetrics and Gynecology Hospital. A total of 67 MRI breast examinations from high-risk patients are included, containing 8,132 images. After a selection process, 6000 images were used for classification, and 1800 images are used for segmentation. The data is divided into training set, verification set and testing set by 8:1:1. All sequences are perfectly aligned, it is because of the imaging protocol. The screening protocol used is as follows. Bilateral MRI is performed simultaneously using a 1.5T scanner (PHILIPS MR Systems Achieva, version 3.2). On account of the T2W_SPARI, T1W_TSE, DWI_SSh, and DYN_eTHRIVE+C images are linked for observing the tumor in clinical diagnosis, these four sequence imaging modes are mainly to be considered in this experiment.

2.1. Classification in breast cancer detection

In order to verify that the proposed four-mode linkage classification backbone is effective, three comparative analyses were performed using four major convolutional neural networks. The major techniques that successfully employ CNNs to medical image classification and applied in our classification task are training the “CNN from scratch” [3, 4, 5, 6] and transfer learning [7]. Also, multiple patches are selected in the image. The method here uses random-size cropping [8]. In addition to the random crops, only random horizontal flipping is adopted as a standard data augmentation scheme. Because the left and right breasts are almost symmetric, this data augmentation method can better increase the number of valid samples.

In the proposed backbone, the images of the four sequences are passed through the convolutional neural network in order. Then, four types of feature maps are acquired as the output of the CNN models. As shown in Fig. 2, each type of image after passing through the CNN can output a feature map. The dimension of all the four feature maps got from CNN are the same, depending on the numbers of convolutional layers in CNNs. The thickness is denoted by t here. Next, we concatenate the features F_1 , F_2 , F_3 , and F_4 extracted from each feature map, and convert them into a higher-dimensional feature. The algorithm of the feature fusion before concatenation can be explained in detail as shown in Algorithm 1.

Algorithm 1 Procedure of the feature fusion algorithm.

Input: N features extracted from the CNN, $\{f_n\}_{n=1}^N$.

- 1: **hyper param** number of relations N_r
- 2: **hyper param** d_k : key feature dimension
- 3: **learnt weights**: $\{W_K^r, W_Q^r, W_V^r\}_{r=1}^{N_r}$
- 4: **for every** (n, r) **do**
- 5: compute $\{\omega_A^{mn,r}\}_{m=1}^N$ using Eq. 3
- 6: compute $\{\omega^{mn,r}\}_{m=1}^N$ using Eq. 2
- 7: compute $f_n^{new_{N_r}}$ using Eq. 1
- 8: **end for**

Output: The fused new feature $\{f'_n\}_{n=1}^N$ using Eq. 4

After four images pass through the CNN, the features are f_1, f_2, f_3 , and f_4 . Then, in order to merge the features of each image into the features of the other three images, for the n -th feature f_n , we calculate the new feature as

$$f_n^{new} = \sum_{m=1}^4 \omega^{mn} \times (W_V \times f_m) \quad (1)$$

Where f_m is the feature of the m -th image, W_V indicates that the feature is linearly transformed to a low dimension, and ω^{mn} represents the weight of the linear sum, the weight can be calculated as

$$\omega^{mn} = \frac{\exp(\omega_A^{mn})}{\sum_k \exp(\omega_A^{kn})} \quad (2)$$

$$\omega_A^{mn} = \frac{\text{dot}(W_K f_m, W_Q f_m)}{\sqrt{d_k}} \quad (3)$$

W_K, W_Q are matrices and play the same roles as W_V , they all represent linearly transform the features. Appearance weight ω_A^{mn} is computed as dot product. d_k is the dimension of the feature after being transformed. After doing the calculation, for the n -th feature f_n , there are N_r features like $f_n^{new_1}, f_n^{new_2}, \dots, f_n^{new_{N_r}}$. Then, the feature of the n -th image can be transformed as

$$f'_n = f_n + \text{Concat} [f_n^{new_1}, f_n^{new_2}, \dots, f_n^{new_{N_r}}] \quad (4)$$

Four new features $f'_1, f'_2, f'_3,$ and f'_4 are obtained after the transformation, we can simply concatenate them [9]. Generally, we set $N_r = 8$, if the dimension of the initial feature $f_1, f_2, f_3,$ and f_4 is 1024, the dimension of the feature $f_n^{new_1}, f_n^{new_2}, \dots, f_n^{new_{N_r}}$ can be $1024 \div 8 = 128$. Finally, the 4t dimension feature that can be expressed as $F=[F_1, F_2, F_3, F_4]$ is put into the softmax process for the classification work.

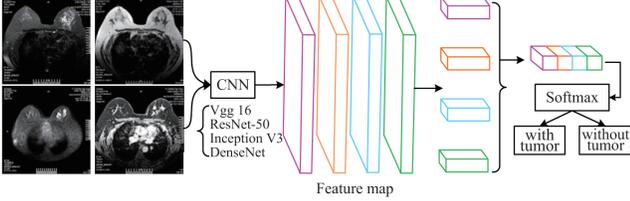


Fig. 2. A group images put into a convolutional neural network.

In the softmax regression, we solve the multi-classification problem (as opposed to the binary problem solved by logistic regression). The y class labels can take k different values (instead of two values). Therefore, with the training sets $\{(x^1, y^1), (x^2, y^2), \dots, (x^m, y^m)\}$, there is $y^i \in \{1, 2, \dots, k\}$. For example, there are k different categories in a classification task. For the given test input x , we want to use the hypothesis function to estimate the probability value $p(y = j | x)$ for each category j . That is, we want to estimate the probability of the occurrence of each classification result of x . Therefore, our hypothesis function will output a vector of k dimensions (the sum of the vector elements is 1) to represent k estimated probability values. Specifically, our hypothetical function $h_\theta(x)$ has the following form:

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} \theta_1^T x^{(i)} \\ \theta_2^T x^{(i)} \\ \vdots \\ \theta_k^T x^{(i)} \end{bmatrix} \quad (5)$$

Where $\theta^1, \theta^2, \dots, \theta^k \in \mathbb{R}^{n+1}$ is the model parameter. Note that item $\frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}}$ normalizes the probability distribution so that the sum of all the probabilities is 1.

In the training, four types of convolutional neural networks, Vgg 16 [10], ResNet-50 [11], Inception V3 [12] and DenseNet [13], are selected as different network architectures for the verification procedure. These network architectures have achieved excellent results on ImageNet, and have shown to be effective for extracting image features.

All the networks are trained using the stochastic gradient descent (SGD) method. We train using the batch size of 16 for 40 epochs, respectively. The learning rate is set to 0.01 initially, and is lowered by 10 times at the epoch 20 and 30. In addition to this, a weight decay of $1e-5$ and a Nesterov

momentum [14] of 0.9 without dampening are used. The loss function used in the training part is defined as

$$L = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n I\{y^{(i)} = j\} \ln p_j^{(i)} \quad (6)$$

where m is the number of samples, n for the number of classifications, I represents the indicator function, $y^{(i)} = j$ here represents the ground truth for the i -th sample belongs to the j -th class, and $p_j^{(i)}$ is the probability that the i -th sample belongs to the j -th class.

2.2. Segmentation

In the segmentation part, for obtaining the most accurate segmentation results, the upsampling process in the traditional U-Net[15] architecture is refined, and the U-Net is initialized using the classified network. For preprocessing, the data are normalized using the means and standard deviations of the dataset. The data augmentation scheme (random horizontal flipping) is adopted for the datasets in classification tasks. Also, the data balancing strategy is employed as described in [13] for mini-batch sampling.

The refined upsampling process specifically use the sub-pixel method. The sub-pixel convolution is a standard convolution followed by additional rearrangement of feature values, termed the phase shift [14]. The process of upsampling is to put the feature map into a layer convolution in which the length and width are in 1×1 , and the number of channels is $4C$. The feature map becomes $H \times W \times 4C$. Then, it is reshaped into $2H \times 2W \times C$. As shown in Fig. 3, such upsampling can achieve better image resolution through learning parameters.

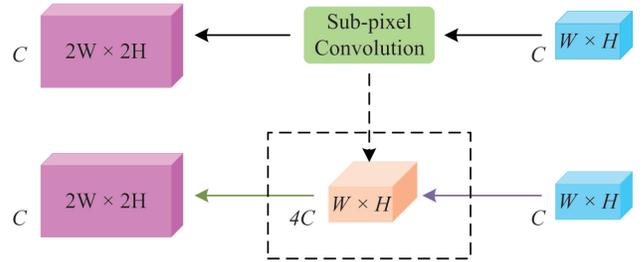


Fig. 3. The refined upsampling process.

The network is trained using the stochastic gradient descent (SGD) method. In the training, the batch size 8 for 60 epochs and the classified network to initialize the downsampling part of U-Net are used to increase the accuracy of the segmentation. The initial learning rate is set to 0.001, and is divided by 10 at 50% and 75% of the total number of the training epochs. The remaining settings are the same as those for classifications. The weight decay of $1e-5$ and the Nesterov momentum of 0.9 without dampening are used. The

loss function used in the segmentation training part is defined as

$$L = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^h \sum_{k=1}^w \sum_{q=1}^n I \{y_{jk}^{(i)} = q\} \ln p_{jkq}^{(i)} \quad (7)$$

where m is the number of samples, n for the number of classifications, I represents the indicator function, $y_{jk}^{(i)} = q$ represents the pixel of the j -th row and the k -th column in the i -th sample belongs to the q -th class, $p_{jkq}^{(i)}$ is the probability that in the i -th sample, the pixel of the j -th row and the k -th column belongs to the q -th class, h represents the height of the image, and w represents the width of the image.

3. RESULTS

Table 1 and 2 reveals that the comparisons across the performance obviously favors the four-mode linkage backbone. All the accuracy data show that the backbone of the four-mode linkage proposed in this study outperforms the existing method of the single-modality image diagnosis in all the four convolutional neural networks. Our best results on DenseNet are more encouraging: the four-mode linkage backbone reaches the accuracy of 94.2%. At the same time, the accuracy data can strongly demonstrate applying the transfer learning strategy, and use the transfer learning strategy and data augmentation simultaneously, the results have a certain degree of improvement. Anyway, the four-mode linkage backbone apparently perform better than the single-modality image backbone. Facts have proved that our new backbone is effective in tumor classification.

Table 1. The accuracy data in single-modality image methods in all four convolutional neural networks in classification.

CNN	Single	Single+trans	Single+trans+aug
Vgg 16	0.876	0.890	0.905
ResNet-50	0.890	0.891	0.907
Inception V3	0.894	0.896	0.909
DenseNet	0.906	0.907	0.911

¹ CNN: Convolutional Neural Network

² Single: the single-modality image backbone

³ Single+trans: the backbone with transfer learning

⁴ Single+trans+aug: the backbone with transfer learning and data augmentation

Whether using the classified network to initialize the segmentation network, and whether using the refine upsampling process in the segmentation task based on U-Net architecture are mainly compared. The results of the comparison are shown in Table 3. It can be seen that combining classified network initialization and the refined upsampling process, the segmentation can be more accurate. The best dice coefficient value reaches 0.865.

Table 2. The accuracy data in four-mode linkage methods in all four convolutional neural networks in classification.

CNN	Linkage	Linkage+trans	Linkage+trans+aug
Vgg 16	0.902	0.915	0.931
ResNet-50	0.924	0.926	0.936
Inception V3	0.927	0.928	0.937
DenseNet	0.931	0.937	0.942

¹ CNN: Convolutional Neural Network

² Linkage: the four-mode linkage backbone

³ Linkage+trans: the backbone with transfer learning

⁴ Linkage+trans+aug: the backbone with transfer learning and data augmentation

Table 3. The dice coefficient values of tumor segmentation in traditional U-Net and improved U-Nets.

Method	U-Net	U-Net+init	U-Net+init+rf
Vgg 16	0.832	0.847	0.860
ResNet-50	0.829	0.833	0.847
Inception V3	0.815	0.821	0.839
DenseNet	0.836	0.849	0.865

¹ U-Net+init: U-Net with the initialization of the classified network

² U-Net+init+rf: U-Net with the initialization of the classified network and refining of upsampling process

4. CONCLUSION

In this paper, a novel breast cancer detection framework based on merging four modes of MR images is developed to enhance the diagnosis process by clinicians. We propose a four-mode linkage backbone for the classification task in tumor detection. This backbone considers different imaging modalities for detecting tumors in multiple signal-weighted imaging sequences through combining MRI-based features and image information to detect breast tumors. In addition, we use an optimized U-Net architecture to segment the filtered tumors. We refine the upsampling process and initialize the segmentation backbone by the classified network. The proposed four-mode linkage backbone in classification and the improved segmentation procedure can be integrated to form a breast cancer detection framework. This framework can be extended to clinical medicine and has important implications for the diagnosis of breast tumors.

5. REFERENCES

- [1] Weiming Zhi, Henry Wing Fung Yueng, Zhenghao Chen, Seid Miad Zandavi, Zhicheng Lu, and Yuk Ying Chung, "Using transfer learning with convolutional neu-

- ral networks to diagnose breast cancer from histopathological images,” pp. 669–676, 2017.
- [2] Fahira Afzal Maken and Andrew P. Bradley, “Multiple instance learning for breast mri based on generic spatio-temporal features,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015, pp. 902–906.
- [3] B Menze, M Reyes, and Leemput K Van, “The multimodal brain tumor image segmentation benchmark (brats).,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993, 2015.
- [4] Wei Shen, Mu Zhou, Feng Yang, Caiyun Yang, and Jie Tian, “Multi-scale convolutional neural networks for lung nodule classification,” 2015, p. 588.
- [5] Gustavo Carneiro, Jacinto Nascimento, and Andrew P. Bradley, “Unregistered multiview mammogram analysis with pre-trained deep learning models,” *International Conference on Medical Image Computing Computer-assisted Intervention*, 2015.
- [6] Jelmer M. Wolterink, Tim Leiner, Max A. Viergever, and Ivana Igum, *Automatic Coronary Calcium Scoring in Cardiac CT Angiography Using Convolutional Neural Networks*, Springer International Publishing, 2015.
- [7] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I Noguees, J. Yao, D Mollura, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [9] Han Hu, Jiayuan Gu, Zheng Zhang, Jifeng Dai, and Yichen Wei, “Relation networks for object detection,” 2017.
- [10] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *Computer Science*, 2014.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” pp. 770–778, 2015.
- [12] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, “Rethinking the inception architecture for computer vision,” *Computer Science*, pp. 2818–2826, 2015.
- [13] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton, “On the importance of initialization and momentum in deep learning,” in *International Conference on Machine Learning*, 2013, pp. III–1139.
- [14] Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” pp. 1874–1883, 2016.
- [15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.