IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Multi-Tasking U-shaped Network for benign and malignant classification of breast masses

HAICHAO CAO[1], SHILIANG PU[1], WENMING TAN[1], JUNYAN TONG[1], and DI ZHANG[1]

[1]Hikvision Digital Technology Company Limited, Hangzhou 310051, China

Corresponding author: SHILIANG PU (pushiliang@hikvision.com).

Project funded by China Postdoctoral Science Foundation under Grant No. 2020TQ0086.

ABSTRACT The benign and malignant (BM) classification of breast masses based on mammography is a key step in the diagnosis of early breast cancer and an effective way to improve the survival rate of patients. Nevertheless, due to the differences in size, shape and texture of breast masses and the visual similarity between masses of the same category, it is difficult to obtain a robust classification model using conventional deep learning methods. To address this problem, we proposed a Multi-Tasking U-shaped Network (MT-UNet), which contains three key ideas: 1) the U-shaped classification architecture constructed can well adapt to the heterogeneity of breast masses; 2) the combination of the proposed truncated normalization method and adaptive histogram equalization method can enhance the contrast of image; 3) training with label smoothing method can alleviate the problem of convergence difficulty caused by insufficient training data. The performance of the proposed scheme is evaluated on the public dataset of DDSM and INbreast. On the DDSM dataset, the Area Under Curve (AUC) and accuracy (ACC) reached 0.9963 and 0.9817, respectively. On the INbreast dataset, the AUC and ACC reached 0.9767 and 0.9391, respectively. Experimental results show that the proposed method can obtain a competitive performance.

INDEX TERMS classification of benign and malignant, image enhance, label smoothing, transfer learning, computer aid diagnosis

I. INTRODUCTION

Breast cancer is one of the most common cancers in the global female population [1], [2]. According to statistics, there were more than 2 million new breast cancer cases worldwide in 2018, accounting for 11.6% of new cancer cases, and more than 600,000 people died of breast cancer in the same year, accounting for 6.6% of cancer deaths [3]. However, if breast cancer can be diagnosed at an early stage, it is possible to improve the survival rate of patients and save the medical cost in a great way [4]. Mammography has been regarded as the preferred method of breast cancer, due to its clear imaging and sensitivity to early breast masses [5]. Also, the use of CNN-based methods to classify the benign and malignant (BM) of breast masses in mammography has important clinical significance and practical application value [6]–[8]. At present, when traditional manual radiograph readings are used to classify the BM of breast masses, the professional knowledge and clinical experience of different doctor is different, resulting in poor reproducibility of diagnostic results. In addition, owing to prolonged reading, it is easy for the doctor to feel tired or even irritable, which will greatly affect the accuracy of the diagnosis results [9]. Therefore, the development of an automatic breast cancer diagnosis system is essential to improve the efficiency and accuracy of doctors' diagnosis.

Studies have shown that computer-aid diagnostic systems can provide fast and reproducible analysis results, while greatly reducing the work pressure of radiologists, and improving the diagnosis accuracy of breast cancer [10], [11]. However, due to the heterogeneity of breast masses in mammography, it is difficult to obtain a robust BM classification model for breast masses. The various difference of breast masses in size, shape and texture can be seen through comparison of Fig. 1 (m_4) and Fig. 1 (m_5), Fig. 1 (m_2) and Fig. 1 (m_5), Fig. (m_3) and Fig. 1 (m_4), respectively. Besides, owing to the high similarity between benign and malignant masses, it is also a challenge to develop a strong classification model. For instance, Fig. 1 (m_1) and Fig. 1 (b_1), their appearances are very similar, but they belong to different categories.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042889, IEEE Access

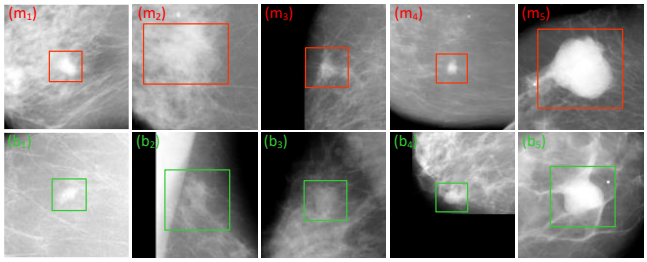

FIG. 1. Example image of BM breast masses in mammography. Note that, the green-labeled images (b_1) - (b_5) and the red-labeled images (m_1) - (m_5) represent five benign masses and five malignant masses, respectively.

To address the problems mentioned above, we proposed a Multi-Tasking U-shaped Network (MT-UNet) architecture, which contains two main modules: 1) classification module, mainly consists of an encoder based on the residual block; 2) segmentation module, mainly consists of a decoder based on the convolution block. The technical contributions of this paper can be included as follows:

(1) We add a decoder based on the conventional classification architecture to supervise the feature extraction of classification module, which can indirectly assist the classification module to extract the truly effective features.

(2) We stretch the image contrast through the proposed truncated normalization method and adaptive histogram equalization method, to extract features of breast masses with low contrast to the surrounding tissues better.

(3) We adopt the training method with label smoothing and transfer learning based on ImageNet to solve the problem of difficult convergence caused by insufficient training data and further improve generalization ability of the model.

II. RELATED WORK

Existing BM classification methods of breast masses can be classified into two categories in terms of feature construction: one is the method that requires manual design of features based on traditional machine learning; the other is CNN-based method, in which features can be automatically learned.

Before CNN was widely known, traditional machine learning-based methods were the mainstream in the BM classification of breast masses [12]–[21]. Traditional machine learning classification methods are generally divided into three steps: 1) manually constructing features, 2) filtering of effective features, and 3) sending them to a classifier for classification. There are many alternatives for the above three steps. In the research of Serifovictrbalic et al., they used the Zernike moment method to extract features, and selected 32 sets of low-order moment features that meet certain conditions as the input to the classifier. Finally, a single hidden layer neural network is applied to classify the BM of breast masses [13]. Xie et al., used the level set method to segment the mass region, the mass boundary region, and the background region, and then extracted the gray features and texture features of the three regions. After that, extreme learning machines are used to classify the BM of breast masses [14]. Al-Antari et al., extracted 28 first-order features, such as gray histogram, gray average, and smoothness. For the breast mass region, the Gray-Level Co-occurrence Matrices (GLCM) is utilized to extract another 304 higher-order features, and finally a three hidden layer deep belief network was adopted to classify the BM of breast masses [8]. Similarly, Punitha et al., [15] and Gautam et al., [16] both apply GLCM to extract the features of breast masses and utilize neural network methods for BM classification. Additionally, Muramatsu et al., proposed a new radial local ternary pattern and used different classifiers to classify different pattern features for the BM of breast masses [17]. Honda et al., designed 11 kinds of features of breast masses, such as irregular shape, edge smoothness, edge irregularity, etc., and utilized secondary discriminant analysis to classify the BM of breast masses [18]. Chaieb et al., studied various descriptors commonly used in the field of breast cancer, and then applied selection techniques to determine the best subset of features to improve the classification performance of breast masses [19]. Boumaraf et al., proposed a feature selection method based on genetic algorithm to select effective features from 130 handcrafted features. Then the back propagation neural network is used to classify breast masses [20]. Jebamony et al., adopted the Laws Texture energy metric for feature extraction, and then utilized the core vector machine classifier to determine breast cancer [21]. Analogously, Danala et al., extracted 109 features including shape, density, and wavelet in the breast mass regon, and finally used a multilayer perceptron to classify the BM of breast masses [22].

Conclusively, the BM classification of breast masses based on traditional machine learning methods mostly extract features such as size, shape, texture, and then use classic classifiers for classification. However, the extracted features must undergo rigorous analysis and experiments in advance, and the construction process is very complicated and tedious. In other words, the performance of the classifier depends largely on the design of the features.

Since 2013, the results of CNN-based methods in image classification tasks have gradually surpassed the traditional classification methods of hand-structured features [23]. In the field of medical imaging, more and more studies have been performed on the detection and classification of lesions using CNN technology [24]–[32]. In the BM classification tasks of breast masses, CNN-based methods can automatically learn effective features and perform end-to-end training. After training, the model will output BM probabilities of breast masses based on the input mammography image.

The research of BM classification of breast masses based on CNN is mainly focused on the improvement of CNN input [24], [25], CNN structure [26]–[28], and training methods.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042889, IEEE Access

In the improvement of CNN input, Dhungel et al., stitched the mammography and the mask image of the breast mass into a two-channel image for the input of CNN [24]. Similarly, Amit et al., combined the enhanced image, the residual image of the enhanced image and the original image, and the residual image of the delayed response image and the original image into a three-channel image as the input of the CNN [25].

In terms of CNN structure improvement, Jiao et al., removed the fully connected layers used for classification in CNN, and sent the middle-level features and high-level features to two linear support vector machines for BM classification of breast masses [26]. Sarkar et al., used Leaky ReLU as the activation function and applied dropout to the fully connected layer to build a CNN model with four convolutional layers and three fully connected layers for the BM classification [27]. Rampun et al., removed the local response normalization layer on the basis of AlexNet, and added a BN layer after each convolution layer, using the PReLU activation function instead of the ReLU activation function to obtain better classification performance than the original AlexNet Effect [28].

Great stride has been made in the training method of CNN. Shen et al., proposed a staged CNN training scheme, and applied different learning rates to train different parts of the model at different stages, and finally obtained a BM classification and localization model of breast masses based on the whole image [29]. In the study of Jiao et al., after the CNN model was trained, they frozen all convolutional layers and trained fully connected layers to maximize the feature distance between classes and minimize the feature distance within the class; Later, the same number of samples are randomly selected from the training set and exchanged with the misclassified samples in the validation set to further improve the robustness of the classification model [30].

Furthermore, there are methods for BM classification of breast masses that combine traditional features and depth features. For example, Wang et al., used a CNN with a feature selection mechanism to extract features from two views of the breast mass and its surrounding region, and fused 17 manually constructed features with features extracted from a convolutional neural network, and finally all features are processed using a recurrent neural network to construct a BM classification model of breast masses that can handle incomplete information [31]. Similarly, Arevalo et al., also explored the fusion of CNN's extracted deep features and hand-designed features, and experimentally proved that traditional features are flawed, and it is impossible to further improve the classification effect of CNN through fusion.

Overall, in terms of feature extraction of breast masses, CNN extracted features have a more complete and more abstract description than features constructed manually. Therefore, this paper uses CNN-based methods to classify the BM of breast masses to further improve the performance of classification.

III. THE PROPOSED METHOD

For the BM classification of breast masses, the classic CNN classification model ResNet-50 [33] is applied in this study. On this basis, in order to further improve the classification effect, we have made improvement in model structure, data pre-processing, the training method with label smoothing, and transfer learning.

A. MODEL STRUCTURE

According to the statistics on INbreast data, we found that the diameter difference of breast masses can reach 20 times. To alleviate this problem and indirectly let the model focus on the information of breast tumor regions, we improved the structure of the single-output classification model into model structure with multiple-output. In ResNet [33], the image input is encoded by a series of ResBlocks to obtain the feature map, and then the global average pooling operation is used to obtain the feature vector, which is then input to the fully connected layer for classification. Based on the prior assumption that the shape and location information of the mass is more important for the feature extraction of the mass, we add the decoding structure after the encoding structure of ResNet to obtain the segmentation mask output. During the training, the classification output and segmentation output are simultaneously trained, which indirectly enables the model to pay attention to the shape and location of breast masses during the down-sampling. Meanwhile, in order to alleviate the problem of gradient disappearance or gradient explosion during backpropagation caused by the segmentation branch being too deep, we use the short connection mechanism of U-Net [34] to connect the output of different levels in the encoding structure to the decoding structure.

TABLE I
IMPLEMENTATION DETAILS OF MODEL FOR THE BM CLASSIFICATION OF BREAST MASSES, WHERE K REPRESENTS THE CONVOLUTION KERNEL, S REPRESENTS THE STEP SIZE.

Structure name	Implementation details [a]
ResBlock-A@s	Conv K: 1×1 S: 1 Conv K: 3×3 S: s Conv K: 1×1 S: s Conv K: 1×1 S: 1
E1	MaxPooling S: 2 ResBlock-A@1 x3
E2	ResBlock-A@2 ResBlock-A@1 x3
E3	ResBlock-A@2 ResBlock-A@1 x5
E4	ResBlock-A@2 ResBlock-A@1 x2

As shown in Fig.2, we construct a Multi-Task U-shaped Network (MT-UNet) architecture, which can output the BM classification results and segmentation results of breast masses at the same time. The network parameters corresponding to MT-UNet are shown in Table I. During the training process, the model uses segmentation loss and classification loss to calculate gradients to promote the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042889, IEEE Access

model's attention to the shape and location information of breast masses, to extract more representative features and improve classification accuracy during the encoding stage.

In the inference process, the decoder structure used for segmentation can be deleted to speed up the inference process.

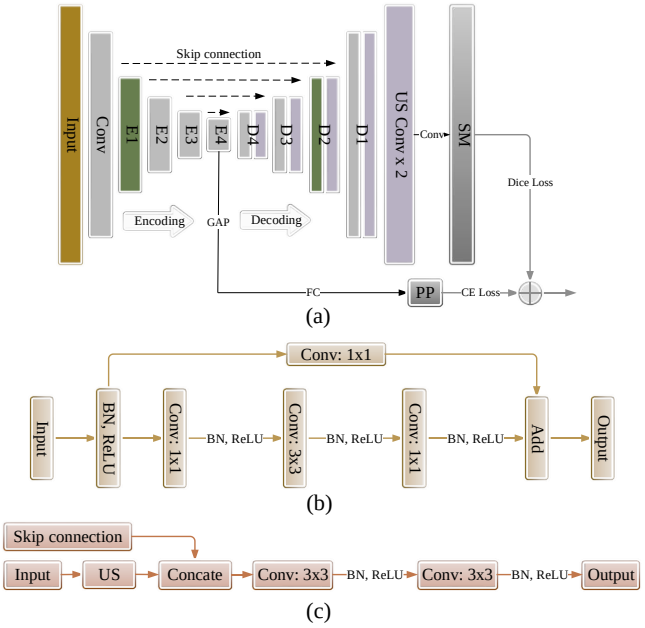

FIG. 2. MT-UNet architecture diagram. Fig. 2 (a) shows a schematic diagram of the overall structure, in which the main body of the CNN model is composed of four encoding modules (E1 ~ E4) and four decoding modules (D1 ~ D4). The input is a pre-processed three-channel image and the output is the mass segmentation mask (SM) and the BM prediction probability (PP) of the mass. The implementation details of E1 ~ E4 are shown in Table I, and the specific structure of ResBlock-A can be seen in Fig. 2 (b). After encoding structure, the output of E4 is first subjected to global average pooling (GAP), and then the classification output is obtained through the fully connected layer. To obtain the segmentation output, we added four decoding modules D1 to D4 after E4, whose structure is shown in Fig. 2 (c), and finally a single-channel segmentation mask output was obtained.

B. DATA PREPROCESSING

In the BM classification of breast masses, the gray value range of the original breast mammography image is 0 ~ 65535. Generally, a breast mammography image contains many meaningless zero-value regions.

In CNN-based methods, the pre-processing of model input data is mostly implemented using simple normalization operations [25]–[27]. However, the gray distribution corresponding to breast masses is mainly concentrated in the region with a gray level of more than 10,000, and the gray distribution will not change if a linear normalization operation is performed directly on the mammography.

To highlight the breast mass region, the pixels with gray level greater than 0 are sorted in ascending order, and the pixels located from the first 1% to the last 5 ‰ are normalized to achieve effective truncated normalization. Fig. 3 shows the effect diagram of the truncated normalization method. It can be seen from Fig. 3 that the contrast between the various tissues in the breast region becomes more obvious.

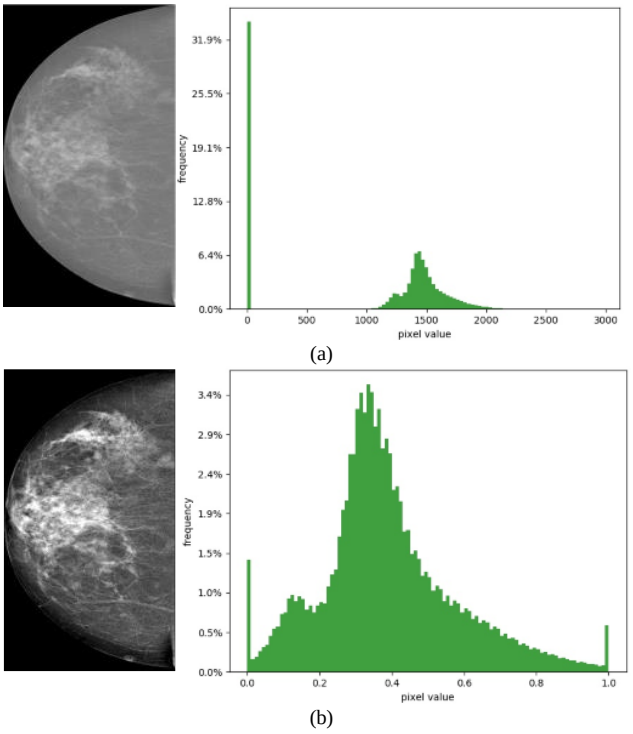

FIG. 3. The effect diagram of the truncated normalization method. Among them, Fig. 3 (a) is the original breast image and its corresponding gray histogram, the histogram shown in Fig. 3(b) is only the gray histogram corresponding to the breast region, not the gray histogram of the whole image.

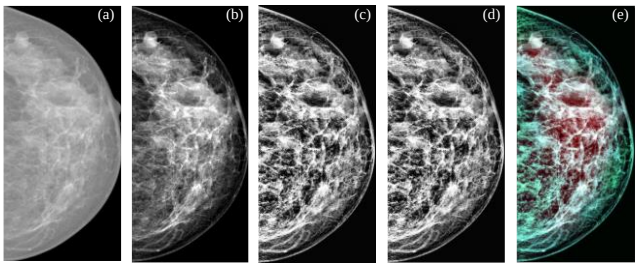

FIG. 4. The comparison of the original image and the pre-processed image. Where (a) represents the original image; (b) represents the image after truncation normalization; (c, d) represents the enhanced image when the clip-limit is 0.01 and 0.02, respectively. (e) represents the synthesized images of (b), (c) and (d).

After the truncated normalization, the image is enhanced using the contrast limited adaptive histogram equalization (CLAHE) method [35] to further stretch the contrast between the breast mass region and the background region. In the experiment, we used two thresholds (0.01 and 0.02 times the number of image pixels) to perform CLAHE processing on the ones that under truncated normalization, and then stitched them with the truncated and normalized image as the input of CNN. The original image and the pre-processed image are shown in Fig. 4. The original mammography image is a single-channel image, the pre-processed image is a three-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042889, IEEE Access

channel image, the first channel is a truncation normalization image, and the remaining two channels are enhanced images processed by the contrast limited adaptive histogram equalization. It can be seen from Fig. 4 that the boundaries of various organizations in the latter are more obvious than in the former.

C. LABEL SMOOTHING

To alleviate the problem of overfitting caused by insufficient labeled data, we adopt label smoothing to process the labels used in training to suppress the model's confidence in the prediction results and enhance the generalization ability of the model. The conventional CNN model training process uses one-hot type labels, which is a vector containing only 0 or 1, its length is equal to the number of categories, and each position in the label vector corresponds to the corresponding category. The definition of label smoothing is shown in Eq.1., where δ represents the original label distribution, μ represents a distribution independent of sample x. The label distribution after label smoothing can be expressed as q'.

$$q'(k \mid x) = (1-\alpha)\delta_{k,v} + \alpha\mu(k) \tag{1}$$

Set a parameter α, the vector values of 1 in the one-hot type label can be changed to 1-α and the vector values of 0 can be changed to α/(K-1) (K is the number of categories), as shown in Fig. 5. Label smoothing has achieved good results in deep learning models in image classification, speech recognition, machine translation and other fields. Szegedy et al. [36] first used label smoothing technology during CNN training. They thought that if the model Predicting a complete probability of 0 or 1 for all training data, the generalization ability of the model cannot be well guaranteed. At the same time, using a one-hot form of the label will encourage the model to always try to make the difference between the maximum logit value and all other logit values larger, and because the gradient is a bounded value, this will cause the model to reduce its adaptability. In this study, due to insufficient training data, the model is prone to overfitting problems, so we use label smoothing, which introduces a certain amount of noise into the labels, which helps alleviate the model's overfitting problems.

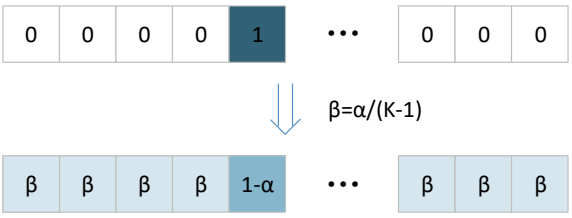

FIG. 5. Function μ takes a uniformly distributed label smoothing.

D. TRANSFER LEARNING

To alleviate the problem of insufficient labeled data, we use transfer learning technology in the BM classification of breast masses. During the training process of CNN models, the training data and testing data have independent and identical prior assumptions, thus data dependence is one of the most serious problems faced by deep learning methods. Different types of tasks require different types of data, and the size of the CNN model and the scale of the required training data are almost linear. In other words, training deep CNN models with insufficient data is very prone to overfitting. Transfer learning can alleviate this problem very well, which assumes that the underlying features learned by the CNN model on different tasks are not much different. We can train the model on data with a different distribution from the testing dataset during training, so that the model has a good initialization parameter, and then use the training data with the same distribution as the testing dataset to fine-tune the model. As shown in Fig. 6., the model is first trained on a large dataset with high labeling quality, then the new model is initialized with some of the pre-trained parameters (generally the feature extraction part), and finally fine-tune the new model on the target dataset. Transfer learning alleviates the need for independent and identical distribution of training and testing data in the way.

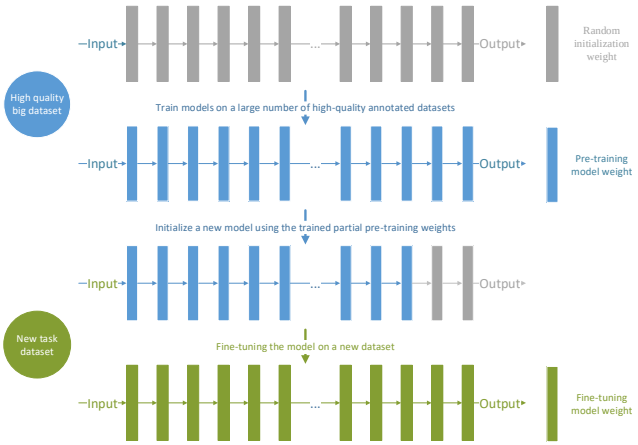

FIG. 6. Schematic diagram of the transfer learning.

Medical image data involves patient privacy and is difficult to collect, and the annotation of the data needs professional doctors. This results in not much annotation data in the medical image field, so transfer learning has been widely used in the medical image field and achieved good results [37]–[39]. Moreover, some studies have shown that transfer learning is effective for the BM classification of breast masses [40]. Specifically, we use the ResNet-50 classification model trained on the ImageNet dataset to initialize the encoding part of our classification model.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

During model training, in order to reduce the hardware cost of training and prediction, and to ensure that the model input image contains certain background information, we take the center point of the breast mass as the cropping center, cut a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042889, IEEE Access

square ROI region with a fixed side length of 512 pixels, and send it to the breast mass BM classification model for training and prediction. The Adam is used as the optimizer; The total number of training iterations is 200; Each sample performs label smoothing with α of 0.05 with a probability of 0.9; The batch data size is 8; The initial learning rate is 1e-4, and if the loss value does not decrease in ten consecutive iterations, the learning rate is reduced to the original 20%.

A. DATA

The experimental data we used in the study were derived from two public datasets: INbreast [41] and DDSM [42]. Among them, the INbreast dataset comes from the Porto Mammary Central Hospital in Portugal. It contains 115 women's breast mammography image. Since one breast generally corresponds to two views, namely the MLO (mediolateral oblique) view and the CC (craniocaudal) view, theoretically INbreast should contain 460 images. But because 25 of these 115 women underwent mastectomy, these people corresponded to only two images. Thus, the INbreast dataset only had 410 images. Among the 410 images, there were 107 lesions containing breast masses, of which 35 breast masses were malignant and 72 breast masses were benign. DDSM data were obtained from the Massachusetts General Hospital, Wake Forest University School of Medicine, Sacred Heart Hospital, and the University of Washington. There was a total of 2,620 cases, including 695 cases with no mass, 1011 cases with benign mass, and 914 cases with malignant mass. The labels in the DDSM dataset contain the mass mask and the BI-RADS tag corresponding to each image. There are fewer cases in the INbreast dataset, but mammography provided by INbreast are clearer than DDSM.

In addition, we divided the training set, validation set and testing set on the DDSM and INbreast datasets respectively, and then evaluated the performance on the corresponding test sets. On the DDSM and INbreast datasets, the ratio of training set, validation set and testing set is divided according to the ratio of 6:2:2. So for the DDSM dataset, the number of images in the training set, validation set, and testing set are 360, 120, and 120, respectively; for the INbreast dataset, the number of images in the training set, validation set, and testing set are 63, 22 and 22 , respectively.

It should be noted that the 3-channel image obtained by preprocessing is the input of the network. The first channel is a truncated normalized image, and the remaining two channels are enhanced images processed by the contrast limited adaptive histogram equalization.

B. EVALUATION CRITERIA

AUC (Area Under Curve), accuracy (ACC) and true positive rate (TPR) are commonly used evaluation indicators in binary classification models. In this study, we used them to evaluate the BM classification performance of breast masses. The corresponding formulas are defined in:

$$\mathrm{AUC} = S_{ROC} \tag{2}$$

$$\mathrm{ACC} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

where S_{ROC} represents the area under the ROC curve [43]. TP indicates the number of correct samples for predicting positive samples; TN indicates the number of correct samples for predicting negative samples; FP indicates the number of samples predicting negative samples as positive samples; FN indicates the number of samples predicting negative samples for positive samples.

C. OVERALL PERFORMANCE

To show the overall performance of the proposed method more intuitively, ROC curve of the MT-UNet architecture on the datasets DDSM and INbreast is drawn in Fig. 7. When the false positive rate is 0.01. the sensitivity of the MT-UNet architecture can reach over 80% on both the DDSM dataset and the INbreast dataset. In addition, when the false positive rate is 0.1, the recall rate corresponding to the DDSM dataset is almost close to 100%, and the one corresponding to INbreast dataset is also over 95%. This shows that the proposed MT-UNet classification architecture can extract features with strong characterization ability from mammography image.

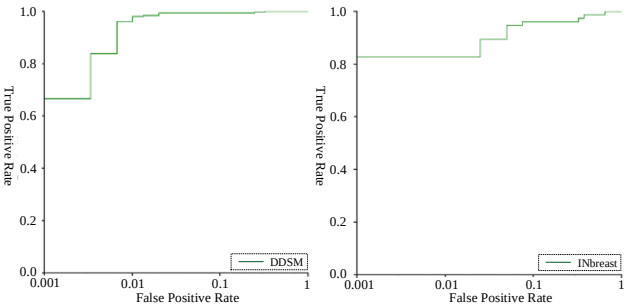

FIG. 7. Function μ takes a uniformly distributed label smoothing.

D. ABLATION STUDY

To verify the effectiveness of the proposed method's components, an ablation experiment was performed on the DDSM dataset. The results of the ablation experiments in Table II can prove the effectiveness of the various improved methods we have adopted for breast mass classification tasks. Among them, Baseline represents the use of resnet-50 to classify the original mammography directly; CN represents truncated normalized preprocessing; MC represents a multi-threshold CLAHE preprocessing method; SC represents a 0.01-threshold CLAHE processing method; Mask represents the mask image of the breast mass is added to the input data of the MT-UNet network; SegB represents the improvement structure that adds segmentation branches for simultaneous multi-task training; LS means using label smoothing technique for training.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042889, IEEE Access

After analyzing the ablation experiment results in Table II, the following conclusions can be drawn:

TABLE II
RESULTS OF ABLATION EXPERIMENTS PERFORMED ON THE DDSM DATASET.

Methods	ACC	AUC
Baseline	0.9600	0.9934
Baseline+CN	0.9667	0.9932
Baseline+CN+MC	0.9667	0.9935
Baseline+CN+SC+Mask	0.9317	0.9634
Baseline+CN+MC+SegB	0.9767	0.9953
Baseline+CN+MC+SegB+LS	0.9817	0.9963

1) EFFECTIVENESS OF TRUNCATED NORMALIZATION AND MULTI-THRESHOLD CLAHE PROCESSING

From the comparison of the two methods of Baseline and Baseline + CN, although the AUC of the experiment with input data using truncated normalization is almost unchanged, the ACC has improved. After cropping into a 512×512 square ROI region, the number of pixels with zero gray value will be greatly reduced, so the effect of truncated normalization method is not obvious. Base on this, we added two gray images processed by CLAHE with different thresholds together with the truncated normalized gray images form a three-channel color image input: Baseline + CN + MC. From the comparison of the results of Baseline and Baseline + CN + MC, it can be seen that the truncated normalized pre-processing and multi-threshold CLAHE pre-processing methods used for the input image, while ensuring the Acc increase, also solve the problem of AUC decline. Therefore, our subsequent experiments use truncated normalization and CLAHE processed images as input.

2) EFFECTIVENESS OF SEGMENTATION BRANCHES

To further improve the classification performance, we used the mask labeling of breast masses and tried two different usage methods. One is to use the mask image as the input of the network model, denoted as Baseline+CN+SC+Mask, so that the model can better learn the size and shape features of the breast mass; the other is to add a segmentation branch, denoted as Baseline+CN+MC+SegB, thereby using segmentation tasks to assist the training of classification tasks. Comparing the experimental results of these two methods, the introduction of segmentation branches under the classification network framework is beneficial to the BM classification of breast masses.

3) EFFECTIVENESS OF LABEL SMOOTHING

From the comparison of Baseline + CN + MC + SegB and Baseline + CN + MC + SegB + LS, it can be seen that the training method of label smoothing can further improve the accuracy of the BM classification model of breast masses. In other words, adding some noise to the training data achieves the purpose of reducing the polarization degree of the model's prediction probability on the training set, and further alleviating a series of problems caused by the lack of training data.

E. EXPERIMENTAL COMPARISON

To quantitatively measure the performance of the proposed method, we used the ACC and AUC as evaluation indicators. We performed experiments on DDSM dataset and INbreast dataset, and compared other papers we found.

TABLE III
COMPARATIVE EXPERIMENTAL RESULTS WITH OTHER PAPERS

Methods	Datasets	ACC	AUC
Jiao et al. [26]	DDSM(600)	0.9670	--
Xie et al. [14]	DDSM(300)	0.9602	0.9659
Jiao et al. [30]	DDSM(600)	0.9740	--
Ours	DDSM(600)	0.9817	0.9963
Shen et al. [29]	INbreast	--	0.96
Dhungel et al. [24]	INbreast	0.9500	0.91
Ours	INbreast	0.9391	0.9767
Jiao et al [30]	MIAS(322)	0.9670	--
Xie et al. [14]	MIAS	0.9573	0.9659
Šerifović-Trbalić et al. [13]	MIAS	0.9040	0.9608

The results are shown in Table III. From the comparison in Table III, the proposed method is used in both the DDSM dataset and the INbreast dataset, which have an overall effect that is superior to other methods. For example, in the comparison of the DDSM dataset with the method of Jiao et al., our method is a percentage point higher than theirs in the ACC indicator. On the INbreast dataset, compared with the method of Dhungel et al., although the Acc indicator is slightly inferior, our method has obvious advantages in terms of the AUC indicator that reflects the overall performance. Besides, some papers also used the MIAS dataset for model training and testing, but because the MIAS dataset is no longer available for download, we cannot verify the proposed method on this dataset.

V. CONCLUSION

In the study of BM classification of breast masses, we propose a model structure for multi-task training using image mask annotation, that is, to improve the classification performance by training two tasks of classification and segmentation at the same time. To alleviate problems such as model overfitting and weaker generalization performance due to the small amount of labeled data, in addition to using transfer learning technology to initialize the model, we also innovatively applied label smoothing technology in the training process, the accuracy was greatly improved. Compared with other methods in published papers, the proposed method achieved an ACC of 98.17% and an AUC of 99.63% on the DDSM dataset; The ACC of 93.91% and AUC of 97.67% was obtained on the INbreast dataset. The experimental results show that our approach has achieved competitive results.

ACKNOWLEDGMENT

Project funded by China Postdoctoral Science Foundation under Grant No. 2020TQ0086.

REFERENCES

[1] C. Allemani et al., "Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries," Lancet, vol. 391, no. 10125, pp. 1023–1075, 2018, doi: https://doi.org/10.1016/S0140-6736(17)33326-3.

[2] A. Jemal et al., "Cancer Statistics, 2008," CA. Cancer J. Clin., vol. 58, no. 2, pp. 71–96, Mar. 2008, doi: 10.3322/CA.2007.0010.

[3] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA. Cancer J. Clin., vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[4] R. T. Chlebowski et al., "Influence of Estrogen Plus Progestin on Breast Cancer and Mammography in Healthy Postmenopausal Women: The Women's Health Initiative Randomized Trial," J. Am. Med. Assoc., vol. 289, no. 24, pp. 3243–3253, 2003, doi: 10.1001/jama.289.24.3243.

[5] T. Kooi et al., "Large scale deep learning for computer aided detection of mammographic lesions," Med. Image Anal., vol. 35, pp. 303–312, 2017, doi: https://doi.org/10.1016/j.media.2016.07.007.

[6] H.-D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey.," Pattern Recognit., vol. 36, no. 12, pp. 2967–2991, 2003, doi: 10.1016/S0031-3203(03)00192-4.

[7] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du, "Approaches for automated detection and classification of masses in mammograms," Pattern Recognit., vol. 39, no. 4, pp. 646–668, 2006, doi: https://doi.org/10.1016/j.patcog.2005.07.006.

[8] M. A. Al-antari et al., "An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network," J. Med. Biol. Eng., vol. 38, no. 3, pp. 443–456, 2018, doi: 10.1007/s40846-017-0321-6.

[9] H. D. Nelson, M. Pappas, A. Cantor, J. Griffin, M. Daeges, and L. Humphrey, "Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation," Ann. Intern. Med., vol. 164, no. 4, pp. 256–267, Feb. 2016, doi: 10.7326/M15-0970.

[10] C. Dromain, B. Boyer, R. Ferré, S. Canale, S. Delaloge, and C. Balleyguier, "Computed-aided diagnosis (CAD) in the detection of breast cancer," Eur. J. Radiol., vol. 82, no. 3, pp. 417–423, 2013, doi: https://doi.org/10.1016/j.ejrad.2012.03.005.

[11] H. R. Roth et al., "Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1170–1181, 2016, doi: 10.1109/TMI.2015.2482920.

[12] G. J. S. Litjens et al., "A survey on deep learning in medical image analysis.," Med. Image Anal., vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[13] A. Serifovic-Trbalic, A. Trbalic, D. Demirovic, N. Prljaca, and P. C. Cattin, "Classification of benign and malignant masses in breast mammograms. BT - 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2014, Opatija, Croatia, May 26-30, 2014." pp. 228–233, 2014, doi: 10.1109/MIPRO.2014.6859566.

[14] W. Xie, Y. Li, and Y. Ma, "Breast mass classification in digital mammography based on extreme learning machine.," Neurocomputing, vol. 173, pp. 930–941, 2016, doi: 10.1016/j.neucom.2015.08.048.

[15] S. Punitha, R. Subban, M. A. Devi, and J. Vaishnavi, "Particle swarm optimized computer aided diagnosis system for classification of breast masses.," J. Intell. Fuzzy Syst., vol. 32, no. 4, pp. 2819–2828, 2017, doi: 10.3233/JIFS-169224.

[16] A. Gautam, V. Bhateja, A. Tiwari, and S. C. Satapathy, "An Improved Mammogram Classification Approach Using Back Propagation Neural Network BT - Data Engineering and Intelligent Computing," 2018, pp. 369–376.

[17] C. Muramatsu, T. Hara, T. Endo, and H. Fujita, "Breast mass classification on mammograms using radial local ternary patterns.," Comp. Bio. Med., vol. 72, pp. 43–53, 2016, doi: 10.1016/j.compbiomed.2016.03.007.

[18] E. Honda, R. Nakayama, H. Koyama, and A. Yamashita, "Computer-Aided Diagnosis Scheme for Distinguishing Between Benign and Malignant Masses in Breast DCE-MRI.," J. Digit. Imaging, vol. 29, no. 3, pp. 388–393, 2016, doi: 10.1007/s10278-015-9856-7.

[19] R. Chaieb and K. Kalti, "Feature subset selection for classification of malignant and benign breast masses in digital mammography.," Pattern Anal. Appl., vol. 22, no. 3, pp. 803–829, 2019, doi: 10.1007/s10044-018-0760-x.

[20] S. Boumaraf, X. Liu, C. Ferkous, and X. Ma, "A New Computer-Aided Diagnosis System with Modified Genetic Feature Selection for BI-RADS Classification of Breast Masses in Mammograms.," CoRR, vol. abs/2005.0. 2020.

[21] J. Jebamony and D. Jacob, "Classification of Benign and Malignant Breast Masses on Mammograms for Large Datasets using Core Vector Machines," Curr. Med. Imaging Rev., vol. 16, no. 6, pp. 703–710, 2020.

[22] G. Danala, F. Aghaei, M. Heidari, T. Wu, B. Patel, and B. Zheng, "Computer-aided classification of breast masses using contrast-enhanced digital mammograms. BT - Medical Imaging 2018: Computer-Aided Diagnosis, Houston, Texas, USA, 10-15 February 2018." p. 105752K, 2018, doi: 10.1117/12.2293136.

[23] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning.," Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[24] N. Dhungel, G. Carneiro, and A. P. Bradley, "The Automated Learning of Deep Features for Breast Mass Classification from Mammograms. BT - Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016 - 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part ." pp. 106–114, 2016, doi: 10.1007/978-3-319-46723-8_13.

[25] G. Amit, R. Ben-Ari, O. Hadad, E. Monovich, N. Granot, and S. Y. Hashoul, "Classification of breast MRI lesions using small-size training sets: comparison of deep learning approaches. BT - Medical Imaging 2017: Computer-Aided Diagnosis, Orlando, Florida, United States, 11-16 February 2017." p. 101341H, 2017, doi: 10.1117/12.2249981.

[26] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature based framework for breast masses classification.," Neurocomputing, vol. 197, pp. 221–231, 2016, doi: 10.1016/j.neucom.2016.02.060.

[27] P. R. Sarkar, D. Mishra, and G. R. K. S. Subrahmanyam, "Classification of Breast Masses Using Convolutional Neural Network as Feature Extractor and Classifier. BT - Proceedings of 2nd International Conference on Computer Vision & Image Processing - CVIP 2017, Roorkee, India, September 9-12, 2017, Volume 2." pp. 25–36, 2017, doi: 10.1007/978-981-10-7898-9_3.

[28] A. Rampun, B. W. Scotney, P. J. Morrow, and H. Wang, "Breast Mass Classification in Mammograms using Ensemble Convolutional Neural Networks. BT - 20th IEEE International Conference on e-Health Networking, Applications and Services, Healthcom 2018, Ostrava, Czech Republic, September 17-20, 2018." pp. 1–6, 2018, doi: 10.1109/HealthCom.2018.8531154.

[29] L. Shen, "End-to-end Training for Whole Image Breast Cancer Diagnosis using An All Convolutional Design.," CoRR, vol. abs/1708.0. 2017.

[30] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A parasitic metric learning net for breast mass classification based on mammography," Pattern Recognit., vol. 75, pp. 292–301, 2018, doi: https://doi.org/10.1016/j.patcog.2017.07.008.

[31] H. Wang et al., "Breast mass classification via deeply integrating the contextual information from multi-view data.," Pattern Recognit., vol. 80, pp. 42–52, 2018, doi: 10.1016/j.patcog.2018.02.026.

[32] J. E. A. Ovalle, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. Á. Guevara-López, "Representation learning for mammography mass lesion classification with convolutional neural networks.," Comput. Methods Programs Biomed., vol. 127, pp. 248–257, 2016, doi: 10.1016/j.cmpb.2015.12.014.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015," 2015, pp. 234–241.

[35] Z. Karel, "Contrast Limited Adaptive Histogram Equalization," Graphics Gems IV. pp. 474–485, 1994.

[36] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision. BT - 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016." pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[37] D. Lévy and A. Jain, "Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks.," CoRR, vol. abs/1612.0. 2016.

[38] R. Wang et al., "Multi-level nested pyramid network for mass segmentation in mammograms.," Neurocomputing, vol. 363, pp. 313–320, 2019, doi: 10.1016/j.neucom.2019.06.045.

[39] L. Tsochatzidis, L. Costaridou, and I. Pratikakis, "Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study," Journal of Imaging , vol. 5, no. 3. 2019, doi: 10.3390/jimaging5030037.

[40] Q. Chen, J. Liu, K. Luo, X. Zhang, and X. Wang, "Transfer deep learning mammography diagnostic model from public datasets to clinical practice: a comparison of model performance and mammography datasets," in Proc.SPIE, 2018, vol. 10718.

[41] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a Full-field Digital Mammographic Database," Acad. Radiol., vol. 19, no. 2, pp. 236–248, 2012, doi: https://doi.org/10.1016/j.acra.2011.09.014.

[42] R. S. Lee, F. Gimenez, A. Hoogi, and D. Rubin, "Curated breast imaging subset of ddsm. The Cancer Imaging Archive," Sci. Data, vol. 4, pp. 1–9, 2017.

[43] C. Sammut and G. I. Webb, Eds., "Area Under Curve," in Encyclopedia of Machine Learning and Data Mining, Boston, MA: Springer US, 2017, p. 61.

Haichao Cao received his Ph.D. degree from Huazhong University of Science and Technology in 2020. He is currently an algorithm engineer at Hikvision Research Institute. His research areas include digital image processing, medical image analysis, machine learning and computer vision.

Shiliang Pu received his Ph.D. degree from University of Rouen and Zhejiang University in 2005 and 2007, respectively. He currently is Chief Research Scientist at Hikvision, and President of Hikvision Research Institute. His research interests include AI, machine perception and robotics.

Wenming Tan received received his master degree from University of Science and Technology of China in 2009. He currently is a Senior Research Scientist at Hikvision. His research interests include machine learning and computer vision.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3042889, IEEE Access

Junyan Tong received her master degree from Nanjing University of technology in 2009. She currently is an algorithm application expert of Hikvision Research Institute, and her main research direction is AI and its application.

Di Zhang received his master degree from Xidian University. He is currently working on application of intelligent algorithm at Hikvision as algorithm manager. His research interests include image processing, artificial intelligence, and computer vision.