Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 21(2): 3110–3128.
DOI: 10.3934/mbe.2024138
Received: 28 November 2023
Revised: 07 January 2024
Accepted: 11 January 2024
Published: 31 January 2024

Research article

A self-supervised fusion network for carotid plaque ultrasound image classification

Yue Zhang[1], Haitao Gan[1], Furong Wang[2], Xinyao Cheng[3], Xiaoyan Wu[4], Jiaxuan Yan[1], Zhi Yang[1,*] and Ran Zhou[1,*]

[1] School of Computer Science, Hubei University of Technology, Wuhan 430068, China

[2] Liyuan Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430074, China

[3] Department of Cardiology, Zhongnan Hospital, Wuhan University, Wuhan 430068, China

[4] Cardiovascular Division, Zhongnan Hospital, Wuhan University, Wuhan 430068, China

* **Correspondence:** Email: zyang631@hbut.edu.cn, ranzhou@hbut.edu.cn.

Abstract: Carotid plaque classification from ultrasound images is crucial for predicting ischemic stroke risk. While deep learning has shown effectiveness, it heavily relies on substantial labeled datasets. Achieving high performance with limited labeled images is essential for clinical use. Self-supervised learning (SSL) offers a potential solution; however, the existing works mainly focus on constructing the SSL tasks, neglecting the use of multiple tasks for pretraining. To overcome these limitations, this study proposed a self-supervised fusion network (Fusion-SSL) for carotid plaque ultrasound image classification with limited labeled data. Fusion-SSL consists of two SSL tasks: classifying image block order (Ordering) and predicting image rotation angle (Rotating). A dual-branch residual neural network was developed to fuse feature presentations learned by the two tasks, which can extract richer visual boundary shape and contour information than a single task. In this experiment, 1270 carotid plaque ultrasound images were collected from 844 patients at Zhongnan Hospital (Wuhan, China). The results showed that Fusion-SSL outperforms single SSL methods across different percentages of labeled training data, ranging from 10 to 100%. Moreover, with only 40% labeled training data, Fusion-SSL achieved comparable results to a single SSL method (predicting image rotation angle) with 100% labeled data. These results indicate that Fusion-SSL could be beneficial for the classification of carotid plaques and the early warning of a stroke in clinical practice.

Keywords: carotid plaque; ultrasound image; self-supervised learning; classification; deep learning

1. Introduction

Carotid atherosclerotic plaque rupture leading to the formation of a thrombus is one of the most significant causes of strokes [1, 2]. For individuals with a high risk of having a stroke, early detection and timely classification of carotid atherosclerotic plaques are clinically significant. Attributable to the convenience, efficiency, and affordability of ultrasound, carotid ultrasound examination has been widely used in carotid plaque classification [3]. In clinical practice, we primarily rely on doctors to manually identify carotid plaques based on neck images, which is time-consuming and labor-intensive. We need to make efforts to detect and classify plaques early to assess the risk level of ischemic stroke in patients and take appropriate preventive and treatment measures. Recently, artificial intelligence-aided carotid plaque diagnosis has shown great potential in helping to reduce the workload of doctors and alleviate the burden of medical care.

Over the past few years, deep learning (DL) algorithms have proven effective in carotid plaque ultrasound image classification. These algorithms enhance the classification capability by constructing multiple nonlinear neural network layers to discover the complex structures in high-dimensional data. Lekadir et al. [4] proposed a convolutional neural network (CNN) that can automatically extract the optimal information for identifying different plaque constituents from images. Zhan et al. [5] designed the clustered PCA network based on the PCAnet with the principal component analysis (PCA) vector as the convolution kernel to extract the features of patches effectively. Ma et al. [6, 7] employed a deep residual network for carotid plaque classification and the spatial pyramid pooling (SPP) was redesigned, and the multilevel strip pooling (MSP) was proposed for carotid plaque longitudinal echo classification. Zreik et al. [8] proposed a multitask recurrent convolutional neural network applied to coronary artery multiplanar reformatted (MPR) images to perform an automatic analysis. After a detailed analysis of the DL-based segmentation technique in the carotid artery ultrasound images [9], Huang et al. proposed a novel boundary-delineation network to extract the vascular wall in carotid ultrasound [10] and a nested attention-guided deep learning model (named NAG-Net) for accurate segmentation of carotid lumen-intima interface and media-adventitia interface [11]. Cai et al. [12] designed a machine-learning approach based on internal carotid artery blood flow to predict cerebral perfusion status. Although the above algorithms have achieved excellent performance, their performance heavily relies on the availability of a substantial amount of labeled images. However, it is challenging to acquire numerous labeled carotid plaque ultrasound images in clinical practice, and manual annotation costs are very high. Therefore, the shortage of labeled image quantity is a limitation for deep learning algorithms in the clinical carotid plaque classification.

To ease the issue of limited labeled image quantity, self-supervised learning (SSL) methods provide a new solution and show good performance [13]. For instance, Bai et al. [14] presented an SSL approach for cardiac magnetic resonance (MR) image segmentation based on the prediction of anatomical positions. Koohbanani et al. [15] demonstrated the SSL method called Self-Path that effectively leverages contextual, multi-resolution, and semantic features to enhance region adaptation for histopathological image patch classification. Abbet et al. [16] introduced an SSL method that integrates clustering metrics with the representation of organizational regions to acquire potential histopathological patterns. Hervella et al. [17] presented the SSL approach for multimodal reconstruction tasks and conducted a series of experiments using multimodal retinal and fluorescein angiography that provided complementary fundus information. In order to better exploit unlabeled images, Chen et al. [18] pro-

posed a strategy of incorporating context restoration as a pretext task and validated its effectiveness in medical imaging tasks. The above works demonstrate that SSL methods can effectively improve the performance of CNNs, but the key lies in how to extract more features through pretext tasks.

Most existing works are mainly focused on how to construct pretext tasks and they don't consider using multiple pretext tasks to pretrain the network. However, different pretext tasks can make the network acquire richer visual feature presentations from different aspects. The model should be more accurate and robust if these different features can be fused. In this study, we propose a self-supervised fusion network for carotid plaque ultrasound image classification named Fusion-SSL. The method utilizes two SSL tasks (classifying image block order [19] and predicting image rotation angle [20]) for the network pretraining. A dual-branch residual neural network is developed to fuse feature presentations learned by the two tasks, which can extract richer visual boundary shape and contour information than a single task. The Fusion-SSL method further enhances the classification performance with limited labeled training images. Investigation on carotid ultrasound images shows that the developed Fusion-SSL algorithm provides high accuracy and agreement in classification results, which is superior to the performance of a single SSL network. This suggests that Fusion-SSL may be suitable for clinical use.

2. Materials and methods

2.1. Acquisition and preprocessing of carotid plaque ultrasound image dataset

2.1.1. Data acquisition

In this study, ultrasound experts with decades of vascular imaging experience acquired 1270 carotid plaque ultrasound images from 844 patients at Zhongnan Hospital (Wuhan, China). Carotid ultrasound images were collected by an Acuson SC2000 (Siemens, Erlangen, Germany) ultrasound system with a 5–12 MHz linear array probe (9L4). The study obtained approval from the Institutional Review Board (IRB) of Zhongnan Hospital (Wuhan, China) and written informed consent from all patients.

2.1.2. Data preparation

The appearance of carotid ultrasound images depends on the image acquisition and is affected by the device, operator, and patient. Moreover, the region of interest (ROI) of carotid plaques in ultrasound images varies in size, which does not satisfy the requirement of a uniform-sized input for CNN. Direct cropping and scaling of ultrasound images may lead to information loss, image distortion, scale inconsistency, and image artifacts, which may affect subsequent analysis and applications. To address the issue of changed tissue appearance and improve the ability of extracting detailed and overall information, the following image preprocessing steps are used in this study to meet the input requirements of CNN: (i) Obtain the ROI image for each plaque by the segmented patch boundary, and set the pixels outside the boundary to zero to get the preprocessed plaque image. (ii) Pad the preprocessed patch image with zeros and generate a modified square plaque image based on the longer side of the original plaque rectangle. (iii) Normalize the size of the square plaque image obtained in step (ii) to a fixed size of 224×224. We used a linear scaling operation between the minimum and maximum values of the images as a standard normalization method to enhance the comparability and reliability. The

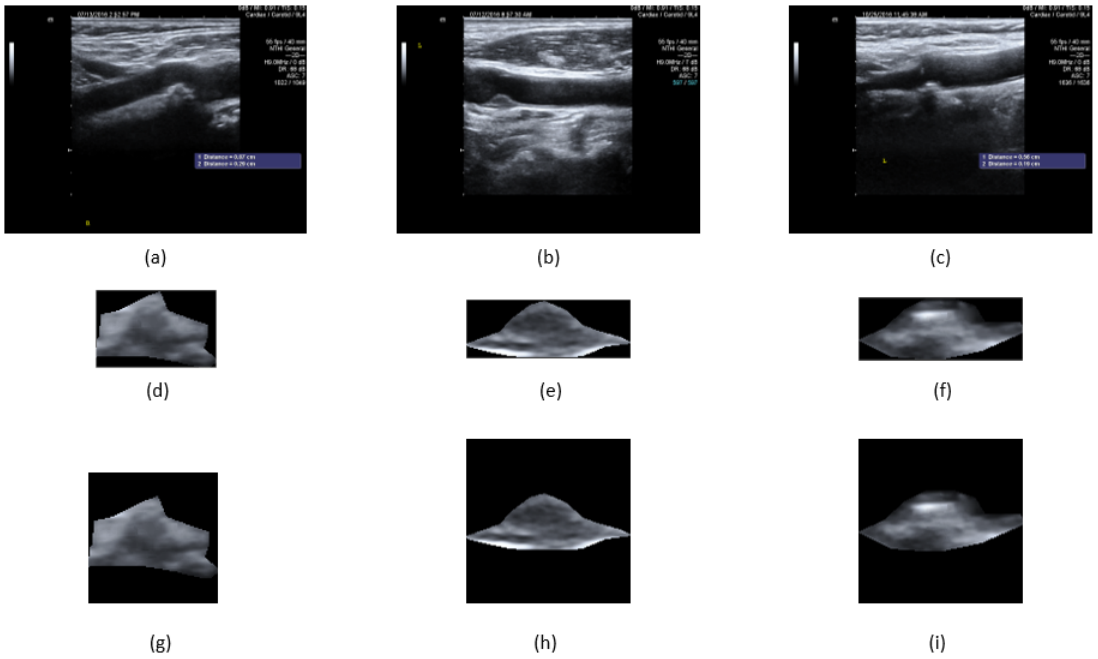

Figure 1. Three types of carotid plaque. (a–c) represent the original ultrasound images of hyperechoic plaque, hypoechoic plaque, and mixed-echoic plaque; (d–f) represent the three preprocessed plaque images; (g–i) represent the three square plaque images.

normalization formula is shown in Eq (2.1). The original and processed images are shown in Figure 1.

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2.1}$$

where x is the pixel value of the carotid plaque ultrasound image, x_{min} and x_{max} are, respectively, the minimum and maximum pixel values of that carotid plaque ultrasound image, and y is the normalized carotid plaque ultrasound image.

2.1.3. Real label data

The real-label data in this study was produced following the European Carotid Plaque Study Group standard, which classified carotid plaque echogenicity into three categories: hyperechoic plaque, hypoechoic plaque, and mixed-echoic plaque. Hypoechoic plaque and mixed-echoic plaque are prone to rupture [21]. This classification was executed by a clinical expert (co-author F. Wang) with over a decade of experience in using carotid ultrasound images to assess atherosclerosis. Initially, he classified 1270 plaques into three types according to echogenicity and reclassified them after three months. He calculated the kappa value (k = 0.747), which indicates a high degree of consistency between the two classifications. Finally, all carotid plaque ultrasound images were classified into three types: hyperechoic plaque, hypoechoic plaque, and mixed-echoic plaque. The resolution of these carotid plaque ultrasound images was 1024×768 with a pixel size of 0.13 mm.

2.2. Carotid plaque ultrasound image classification based on self-supervised fusion network

Although DL has demonstrated effectiveness in carotid plaque ultrasound image classification, its capability heavily depends on numerous labeled datasets. SSL has partially alleviated this issue, but extracting rich features through a single SSL pretext task is difficult. Therefore, we fused two pretext tasks to pretrain the network, then, a dual-branch residual neural network was constructed to extract feature maps from the two tasks and fuse them. We aim to obtain richer visual features of carotid plaques by learning from boundary shapes and contour information. This approach not only addresses the issue of insufficient labeled images, but also further enhances the capacity of carotid plaque image classification.

However, not all feature fusion is helpful in improving the performance of the model. Some features extracted by different SSL methods will produce redundancy and inconsistent problems. After comparing, we chose two reliable SSL methods, classifying image block order and predicting image rotation angle, which can provide diverse feature learning, complementary feature extraction, and enhanced robustness, thus improving the performance and reliability of the model. First, they both enrich the diversity of features. Classifying image block order randomly introduces different contextual information and predicting image rotation angle introduces different perspectives and transformations. Second, the features extracted by these two methods are complementary. Classifying image block order can emphasize the long-distance dependence in the image and investigate global information, helping the model to learn the semantic information of the image. Predicting image rotation angle can capture the local orientation information in the image and retain the information of the original image, helping the model to understand the geometry and structure in the image. Therefore, we propose the idea of combining these two SSL methods to leverage their strengths and compensate for their respective limitations. By simultaneously learning the tasks of classifying image block order and predicting image rotation angles, our network can extract richer image features from various auto-generated targets and better capture the semantic information and structure of the images. This fusion approach is expected to enhance the performance of SSL in image classification tasks and decrease the requirement for labeled images, thereby reducing the cost and time consumption of data labeling.

The Fusion-SSL method proposed in this study mainly consists of two stages. (1) Pre-training of self-supervised models: Model A and model B are obtained by training two different SSL methods with the same CNN, and then both model A and model B are migrated to the feature extraction layer in the dual-branch residual network for parameter initialization. (2) Dual-branch residual network feature fusion in plaque classification: Build a dual-branch residual network (DBResNet) to fuse the two self-supervised network models obtained in step (1). The DBResNet contains three parts: a feature extraction layer, a feature fusion layer, and a fully connected layer. 1) Transfer models A and B to the two parallel branch networks of the feature extraction layer in DBResNet (both the first and second branches are composed of ResNet-101, but the ResNet-101 architecture is only kept up to the conv5_x residual structure and the subsequent average pooling layer and fully connected layer are removed) to initialize the parameters for each branch and obtain two features (F_A and F_B) after training; 2) F_A and F_B are fused through the feature fusion layer to obtain the fused feature $\hat{F}$; 3) Input the fused feature $\hat{F}$ to the fully connected layer to obtain the final classification result. The overall flowchart of Fusion-SSL is shown in Figure 2.

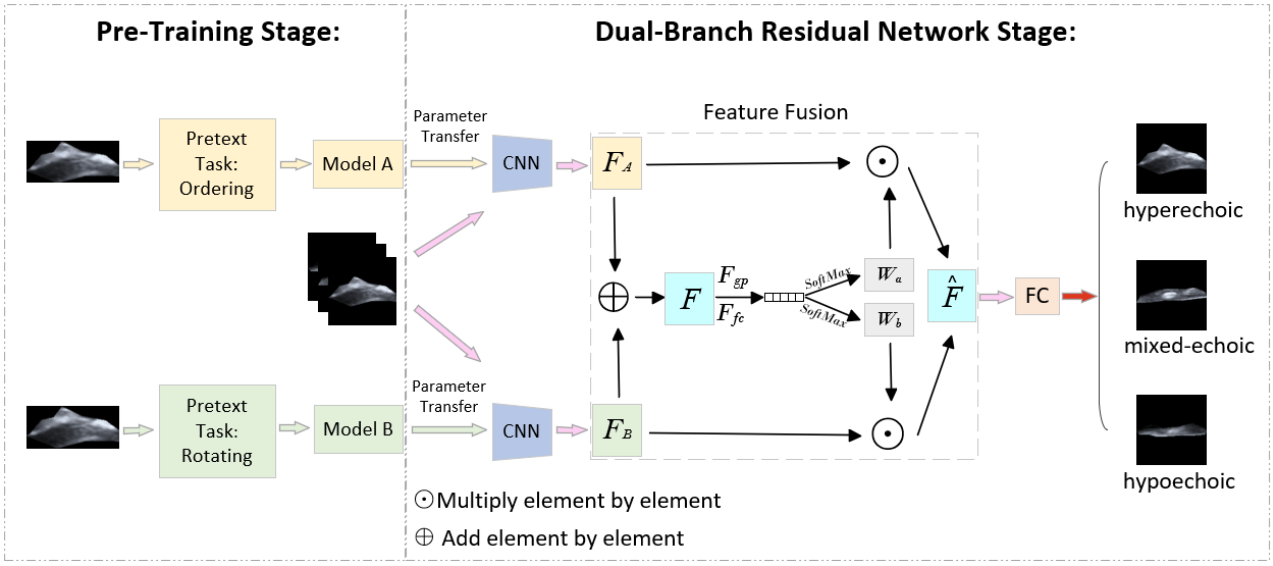

Figure 2. The overall flowchart of Fusion-SSL.

2.2.1. Pretraining of self-supervised models

To learn richer visual features about carotid plaques, the proposed Fusion-SSL method in this study combines two different SSL methods for feature learning.

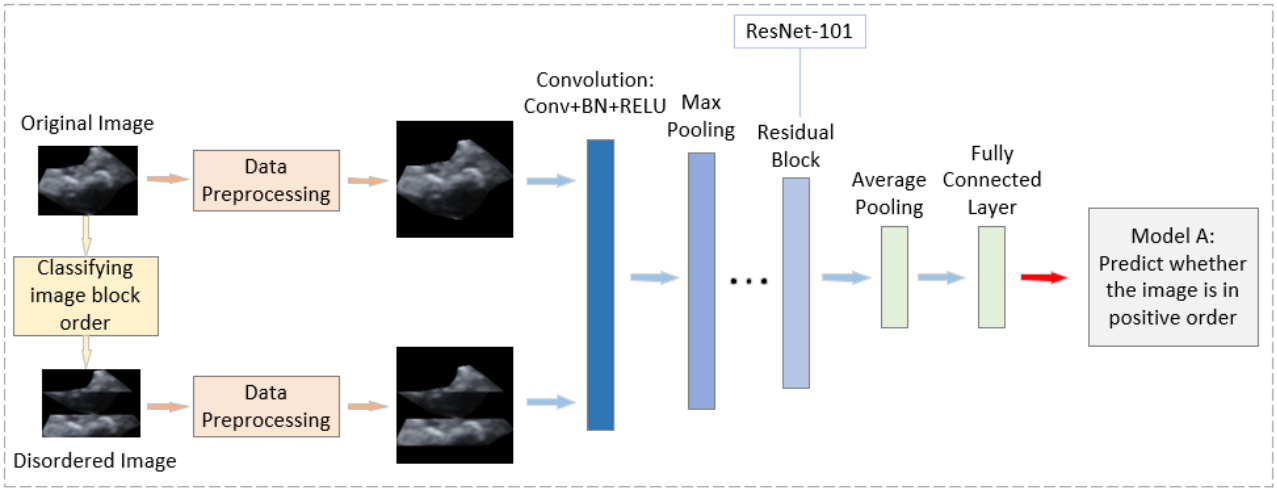

Figure 3. Flowchart of feature extraction for classifying image block order method.

(1) The first SSL method used in this study is the "classifying image block order" method. Its process is shown in Figure 3. First, the original carotid plaque ultrasound dataset X' is reorganized into a new carotid plaque ultrasound dataset R' by the self-supervised "classifying image block order" method, and corresponding pseudo-labels are generated. The label set of the original image dataset X' is set as 1, while the label set of the disordered image dataset R' is set as 0 (the real labels of the dataset

are not used here). Second, the $X^{'}$ dataset and $R^{'}$ dataset are preprocessed, and then the two datasets after the preprocessing operation are combined to call the expanded $X^{''}$ dataset. The label set of the $X^{''}$ dataset is $Y^{''}$ and its value is shown in Eq (2.2). The classification categories for this dataset are binary, which include the original images and the disordered images. Finally, the images of the dataset $X^{''}$ with their corresponding label set $Y^{''}$ are convolved and pooled to serve as the input for ResNet-101, and the optimal network model A is obtained after multiple training iterations.

$$y_i = \begin{cases} 1, & (X_i^{''} \in X') \\ 0, & (X_i^{''} \in R') \end{cases} \tag{2.2}$$

where y_i represents the value of $Y^{''}$ label set and $X_i^{''}$ represents the i-th image in $X^{''}$ dataset.

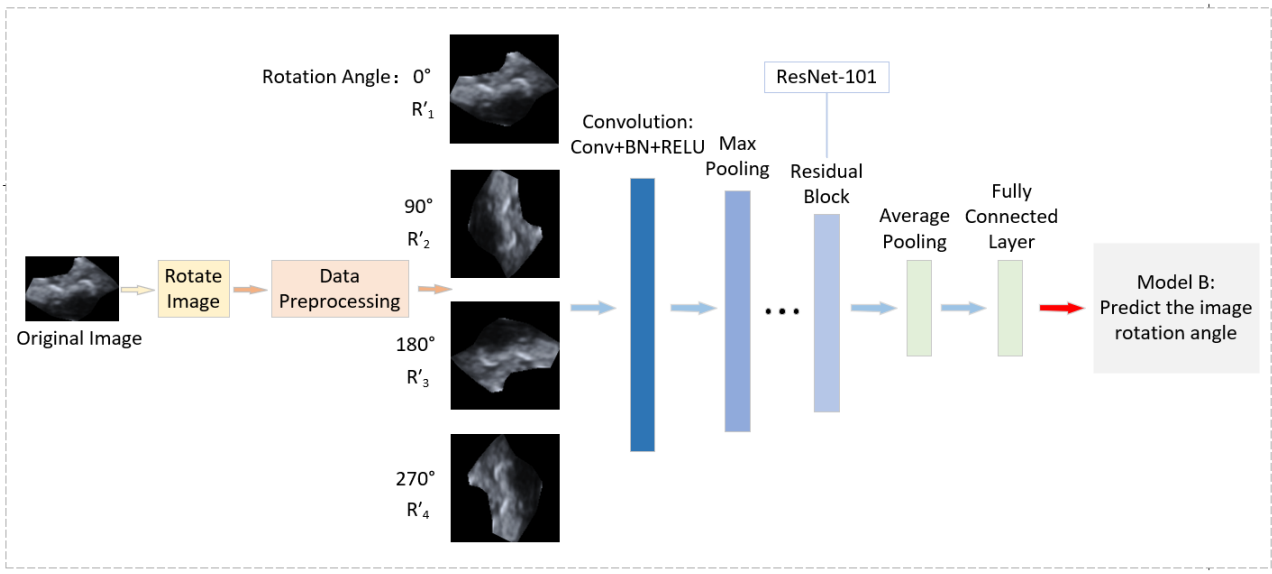

Figure 4. Flowchart of feature extraction for predicting image rotation angle method.

(2) The second SSL method used in this study is the "predicting image rotation angles" method, and the process is shown in Figure 4. First, the original carotid plaque ultrasound dataset $X^{'}$ is transformed into rotational angles of 0, 90, 180, and 270 degrees by the self-supervised "predicting image rotation angle" method. At the same time, data preprocessing operations are performed on the new carotid plaque ultrasound dataset $R^{'}$ after angular transformation. Carotid plaque ultrasound datasets with four different rotation angle transformations $R_1^{'}$, $R_2^{'}$, $R_3^{'}$, and $R_4^{'}$ are obtained as shown in Figure 4. The label set of the dataset $R_1^{'}$ is set to 0, the label set of the dataset $R_2^{'}$ is set to 1, the label set of the dataset $R_3^{'}$ is set to 2, and the label set of the dataset $R_4^{'}$ is set to 3. $R_1^{'}$, $R_2^{'}$, $R_3^{'}$, and $R_4^{'}$ were then merged into a new dataset $R^{''}$ with four categories, namely, four different angles. The label set of the $R^{''}$ dataset is the $Y^{''}$ and its value is shown in Eq (2.3). Next, the images in the dataset $R^{''}$ with their corresponding label set $Y^{''}$ are convolved and pooled to serve as the input to ResNet-101, and the final network model B is

obtained after multiple training iterations.

$$y_i = \begin{cases} 0, & (R_i'' \in R_1') \\ 1, & (R_i'' \in R_2') \\ 2, & (R_i'' \in R_3') \\ 3, & (R_i'' \in R_4') \end{cases} \tag{2.3}$$

where y_i represents the value of Y'' label set and R_i'' represents the i-th image in R'' dataset.

2.2.2. DBResNet feature fusion in plaque classification

The overall framework of DBResNet consists of three parts: the feature extraction layer, the feature fusion layer, and the fully connected layer.

(1) The feature extraction layer of DBResNet has two parallel branches, each of which consists of the network structure after deleting the average pooling layer and the fully connected layer in ResNet-101, and its structure is shown in Figure 5. The first layer is a 7×7 convolutional layer with a stride of 2 and 64 convolutional filters. The second layer is a 3×3 max pooling layer with a stride of 2. After passing through the first two layers, a 56×56, 64-channel output is obtained. The subsequent layers are well-known residual modules, including 3 conv2 blocks, 4 conv3 blocks, 23 conv4 blocks, and 3 conv5 blocks. Due to their complexity and extensive utilization, they will not be elaborated here. Finally, two 7×7, 2048-channel feature maps (F_A and F_B) are obtained after multiple convolution operations.

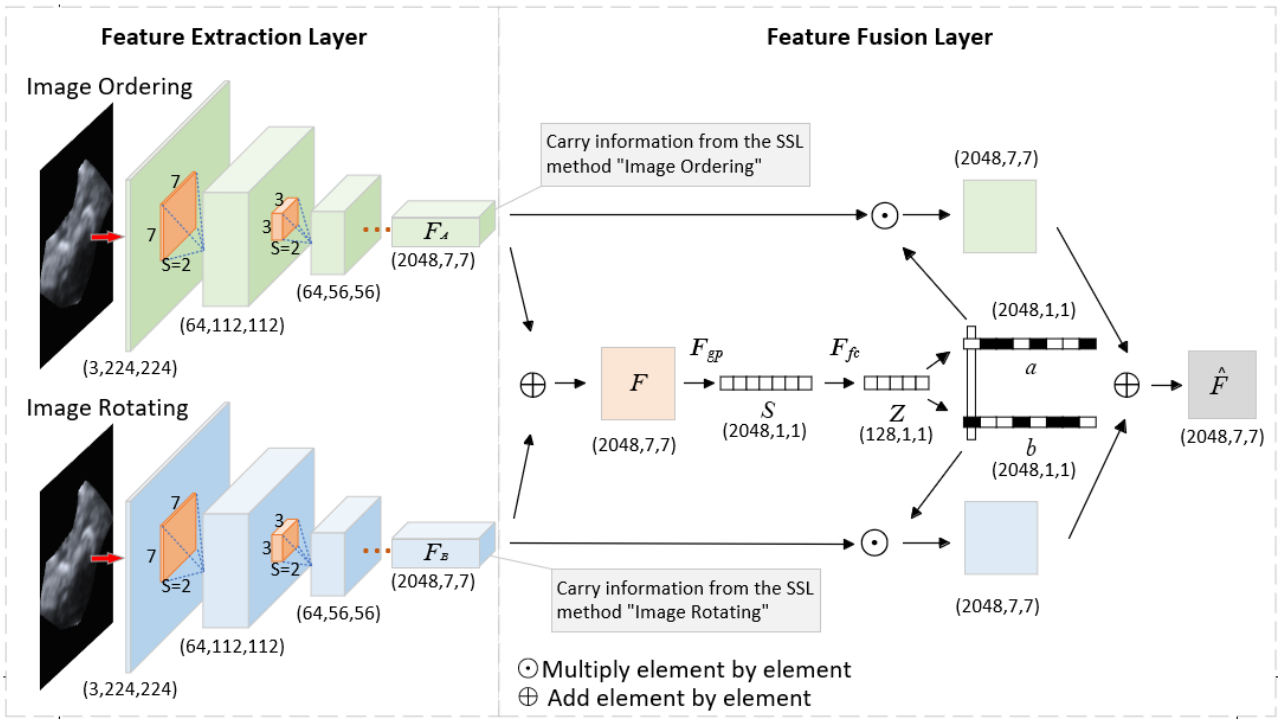

Figure 5. Flowchart of DBResNet.

(2) The selective kernel network (SKNet) [22] presented by Li et al. adopts a nonlinear method to fuse features from different kernels, achieving adjustment of receptive fields with different sizes. It contains three operations: The split operation generates several channels with varying sizes of kernel according to the different receptive field sizes of the neurons; the fuse operation integrates information from several channels to acquire a global and comprehensible representation for weight selection; and the select operation fuses the feature maps of different kernel sizes based on the weights obtained from the selection. Therefore, the Fusion-SSL method proposed in this study uses the SK module as a feature fusion layer of DBResNet to learn richer visual features. 1) As shown in Figure 5, the feature fusion layer of DBResNet differs from the SK module in the following locations: The network layer removes the Split operation in the SK module and combines the feature maps F_A and F_B after the feature extraction layer as inputs, and the feature F_A and F_B, respectively, carry information from two SSL methods. 2) The fusion result is fused from the two branches by elemental summation. First, the two feature maps are added to obtain F ($F = F_A + F_B$). Then, inspired by channel-attention, to achieve the weighted attention of different channel features and suppress redundant features, making the model pay more attention to features that are more important to the current task, the feature map is embedded with global information by applying a simple global average pooling operation (F_{gp}) to generate channel statistics s as shown in Eq (2.4), where C is the feature dimension of the s in Figure 5. The feature z is then obtained through a fully connected layer (F_{fc}). 3) Calculate the weight a_c and b_c of each layer of different sensory field information using the softmax function according to the direction of the channel, where b is a redundant matrix, and in the case of two branches, $b = 1 - a$, as shown in Eq (2.5). Finally, the elements of a_c and b_c are multiplied and summed with the corresponding features F_A and F_B to obtain the final output $\hat{F}$, as shown in Eq (2.6).

$$s_c = F_{gp}(F_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i, j) \tag{2.4}$$

$$a_C = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}}, b_C = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}} \tag{2.5}$$

$$\widehat{F} = a_c F_A + b_c F_B, a_C + b_C = 1 \tag{2.6}$$

(3) The fully connected layer comprises a global average pooling layer and a fully connected layer to predict the carotid plaque class (hyperechoic plaque, hypoechoic plaque, and mixed-echoic plaque).

2.3. Experiment setting

The hardware environment of this study is NVIDIA GeForce GT 710 graphics card and 24G memory. The software environment is the Windows 10 operating system, PyTorch 2.0.1 DL framework, Python 3.8.8, CUDA 11.7, and PyCharm 2023 community version.

For the pretraining part of the Fusion-SSL method, the details of the dataset division are as follows:

(1) The total number of carotid plaque ultrasound image dataset X'' expanded by the "classifying image block order" pretext task is 2540. Next, the expanded carotid plaque ultrasound image dataset X'' was randomly divided into a training set and a validation set in a ratio of 0.8:0.2. The number of images for the divided training sets and the validation sets were 2032 and 508.

(2) The total number of carotid plaque ultrasound image datasets R'' expanded by the "predicting image rotation angle" pretext task is 5080. Next, the expanded carotid plaque ultrasound image dataset R'' was randomly divided into a training set and a validation set in a ratio of 0.8:0.2. The number of images for the divided training sets and the validation sets were 4064 and 1016.

For the DBResNet part of the Fusion-SSL method, the dataset is divided as follows:

The dataset used in this part is the original carotid ultrasound dataset X', which has a total number of 1270. The dataset was then randomly divided into training set, validation set, and test set based on the number of patients in the ratio of 0.6:0.2:0.2. The number of images in the partitioned training set, validation set, and test set were 764, 260 and 246.

The network training parameters for this experiment were epoch, batch size, and learning rate. The selection of these parameters is based on the datasets, task, model complexity, computational resources, and other conditions.

(1) Adjustment of epoch: Increasing the number of epochs increases the time of model training, but improves the model accuracy. However, excessive epochs may lead to overfitting of the model. Therefore, the most appropriate epoch can be selected by gradually increasing the number of epochs and observing the model performance. In our experiment, the "classifying image block order" pretext task was trained for 25 epochs, while the "predicting image rotation angle" pretext task was trained for 150 epochs. The remaining experiments were all trained for 50 epochs.

(2) Adjustment of batch size: A larger batch size can accelerate training speed, especially on the graphics processing unit (GPU), but may reduce accuracy. A smaller batch size may lead to a more stable convergent behavior, reducing the probability of overfitting. Therefore, multiple attempts on the model are needed to find the most appropriate batch size. As shown in Table 1, we finally chose 8 as the best batch size.

Table 1. Performance evaluation of the Fusion-SSL method on 100% labeled training images for different batch size.

batch size	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	G-Mean (%)	F1 score (%)	Kappa
4	86.0 ± 2.3	85.7 ± 1.9	85.8 ± 2.1	92.7 ± 1.1	89.1 ± 1.6	85.5 ± 2.1	0.78 ± 0.03
8	**87.6 ± 1.0**	**87.4 ± 1.2**	**87.2 ± 1.3**	**93.4 ± 0.7**	**90.2 ± 1.0**	**87.2 ± 1.1**	**0.80 ± 0.02**
16	87.2 ± 2.0	87.0 ± 1.7	86.9 ± 2.0	93.1 ± 1.3	89.8 ± 1.8	86.9 ± 2.0	0.79 ± 0.03

(3) Adjustment of learning rate: A smaller learning rate can make the model more stable but may require more iterations to converge. The large learning rate can accelerate the weight update speed but will cause the model to jump in the iteration and miss the optimal solution. Therefore, multiple experiments need to be performed to find a suitable learning rate. As shown in Table 2, we finally chose 0.0001 as the best learning rate.

Table 2. Performance evaluation of the Fusion-SSL method on 100% labeled training images for different learning rate.

learning rate	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	G-Mean (%)	F1 score (%)	Kappa
0.01	86.2 ± 1.9	85.4 ± 2.0	86.3 ± 2.2	92.9 ± 1.0	89.5 ± 1.6	85.7 ± 2.1	0.78 ± 0.03
0.001	85.8 ± 1.3	85.1 ± 1.8	85.2 ± 1.2	92.6 ± 0.6	88.7 ± 1.0	84.9 ± 1.1	0.77 ± 0.02
0.0001	**87.6 ± 1.0**	**87.4 ± 1.2**	**87.2 ± 1.3**	**93.4 ± 0.7**	**90.2 ± 1.0**	**87.2 ± 1.1**	**0.80 ± 0.02**
0.00001	84.8 ± 0.3	84.5 ± 1.0	84.0 ± 0.5	91.7 ± 0.4	87.5 ± 0.5	84.0 ± 0.3	0.76 ± 0.01

After multiple experiments, the final parameter settings for different experiments in this study were as follows:

(1) During the pretraining phase, the CNN was ResNet-101, the optimizer was Adam optimizer, with a batch size of 8 and a learning rate of 0.0001, and the loss function was cross-entropy loss. The "classifying image block order" pretext task was trained for 25 epochs, while the "predicting image rotation angle" pretext task was trained for 150 epochs.

(2) During the training of the DBResNet, the Adam optimizer was used with a batch size of 8 and a learning rate of 0.0001, the loss function was cross-entropy loss, and the training was performed for 50 epochs.

(3) During the training process of the other comparative experiments, the Adam optimizer was used with a batch size of 8 and a learning rate of 0.0001, the loss function was cross-entropy loss, and the training was performed for 50 epochs.

Lastly, all experimental results in this study were obtained through 5 random experiments.

2.4. Evaluation metrics

To evaluate the performance of the Fusion-SSL method, we adopted several evaluation metrics, including accuracy, precision, sensitivity, specificity, G-Mean, F1 score [23], and Kappa coefficient [23]. In addition, to compare the performance of different classifiers, we also evaluated the performance of the Fusion-SSL method using the receiver operating characteristic curve (ROC) and the area under the ROC curve, Precision-Recall (PR) curve and the area under the PR curve. Based on the multiclassification confusion matrix, we determined the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each of the three plaque classifications and then used them to obtain the following metrics. The equations are as follows:

$$Accuracy = \frac{1}{K}\sum_{k=1}^{K}\frac{TP_k + TN_k}{TP_k + TN_k + FP_k + FN_k} \times 100\% \tag{2.7}$$

$$Precision = \frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k + FP_k} \times 100\% \tag{2.8}$$

$$Sensitivity = \frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k + FN_k} \times 100\% \tag{2.9}$$

$$Specificity = \frac{1}{K}\sum_{k=1}^{K}\frac{TN_k}{TN_k + FP_k} \times 100\% \tag{2.10}$$

$$G - Mean = \sqrt{Sensitivity \times Specificity} \times 100\% \tag{2.11}$$

$$F1 = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \times 100\% \tag{2.12}$$

$$kappa = \frac{c \times s - \sum_k^K p_k \times t_k}{s^2 - \sum_k^K p_k \times t_k} \tag{2.13}$$

Here k is the k-th class, and K is the total number of classes. The calculation of Kappa is based on the confusion matrix C in Figure 8 of reference [23], where $c = \sum_k^K C_{kk}$ is the total number of correctly predicted elements, $s = \sum_i^K \sum_j^K C_{ij}$ is the total number of elements, $p_k = \sum_i^K C_{ki}$ is the number of times the class K was predicted (column sum), and $t_k = \sum_i^K C_{ik}$ is the number of times the class K truly occurred (row sum).

3. Results

3.1. Selection of the base CNN structure

This experiment evaluated the performance of popular network models in CNN (ShuffleNet-V2 [24], MobileNet-V3 [25], EfficientNet-B0 [26], and ResNet-101 [27]) to select the best model for plaque classification applications. To test the performance of CNNs on small training datasets, this experiment chose 40 and 100% of the labeled training images as examples. The experimental results are shown in Table 3. Among all the networks shown in Table 3, ResNet-101 has the best evaluation metrics for all the evaluation metrics in the case of 40 and 100% labeled training images. Therefore, ResNet-101 was used in this study as the base network for subsequent experiments. The deep structure, residual connection, pretrained model, and convolution pooling of ResNet-101 provide better feature extraction and initial weight, helping our model to reduce the load, accelerate convergence, and achieve better performance.

Table 3. Performance evaluation of different networks on 40 and 100% labeled training images.

Model	Proportion	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1 score (%)	Kappa
ShuffleNet-V2	40%	64.1 ± 1.9	61.9 ± 2.9	62.1 ± 2.6	79.4 ± 1.1	60.6 ± 2.7	0.41 ± 0.03
	100%	69.0 ± 2.2	68.7 ± 3.3	66.6 ± 2.6	82.0 ± 1.1	65.4 ± 3.5	0.49 ± 0.04
MobileNet-V3	40%	69.1 ± 1.3	71.3 ± 3.6	66.3 ± 2.0	81.8 ± 0.9	62.9 ± 3.7	0.48 ± 0.02
	100%	73.7 ± 0.7	73.9 ± 0.7	72.0 ± 0.6	85.0 ± 0.4	72.3 ± 0.4	0.57 ± 0.01
EfficientNet-B0	40%	70.0 ± 0.7	69.0 ± 1.6	67.8 ± 0.5	82.9 ± 0.4	66.7 ± 1.2	0.51 ± 0.01
	100%	79.3 ± 0.9	79.2 ± 0.7	78.4 ± 1.1	88.6 ± 0.6	78.6 ± 1.0	0.67 ± 0.02
ResNet-101	**40%**	**78.4 ± 1.0**	**77.4 ± 1.1**	**77.2 ± 2.1**	**88.3 ± 0.9**	**76.8 ± 1.7**	**0.65 ± 0.02**
	100%	**83.3 ± 0.9**	**83.1 ± 0.5**	**82.1 ± 1.6**	**91.1 ± 0.7**	**82.4 ± 1.2**	**0.73 ± 0.02**

3.2. *Performance comparison of Fusion-SSL and single SSL methods under a few labeled training images*

To show that Fusion-SSL can effectively relieve the issue of limited labeled images, this study compared Fusion-SSL with different SSL methods on diverse ratios of labeled training images, including 10, 20, 30, 40, and 100%.

Table 4 and Figures 6 and 7 show the accuracy, precision, sensitivity, specificity, F1 score, and Kappa coefficient of SSL methods and the Fusion-SSL method trained with different proportions of labeled training images. The results indicate that the performance of the Fusion-SSL method is superior to individual SSL methods on all proportions of labeled training images.

For 10, 20, 30, 40, and 100% labeled training images, compared to the Ordering [20] method, the accuracy of Fusion-SSL was improved by 6.6, 5.8, 3.6, 3.3, and 3.1%. Compared to the Rotating [21] method, the accuracy of Fusion-SSL was improved by 7.2, 9.8, 6.3, 4.1, and 2.3%. Compared to the SimSiam [28] method, the accuracy of Fusion-SSL was improved by 1.9, 4.3, 4.9, 4.0, and 2.3%. Similarly, other evaluation metrics also yielded similar results.

Table 4. Performance comparison between Fusion-SSL and single SSL methods on labeled images of different proportions.

Method	Proportion	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	G-Mean (%)	F1 score (%)	Kappa
Ordering	10%	68.4 ± 2.4	67.4 ± 4.0	66.0 ± 2.2	81.8 ± 1.5	69.9 ± 3.7	64.3 ± 3.8	0.48 ± 0.04
	20%	75.1 ± 1.5	75.0 ± 1.3	73.8 ± 1.0	86.2 ± 0.8	79.1 ± 1.0	74.0 ± 0.8	0.60 ± 0.02
	30%	79.5 ± 2.3	79.0 ± 2.7	78.3 ± 3.5	89.0 ± 1.6	83.1 ± 2.8	78.2 ± 3.0	0.67 ± 0.04
	40%	81.2 ± 1.7	80.9 ± 1.5	80.0 ± 2.4	89.7 ± 1.2	84.3 ± 2.1	80.1 ± 2.2	0.70 ± 0.03
	100%	84.5 ± 1.1	84.4 ± 1.3	83.9 ± 1.5	91.6 ± 0.6	87.6 ± 1.1	84.1 ± 1.4	0.75 ± 0.02
Rotating	10%	67.8 ± 1.8	63.8 ± 9.5	63.8 ± 2.6	80.5 ± 1.5	63.5 ± 1.7	60.7 ± 5.9	0.45 ± 0.04
	20%	71.1 ± 3.0	70.5 ± 4.8	68.6 ± 4.3	83.0 ± 2.3	71.6 ± 7.0	67.3 ± 6.5	0.52 ± 0.06
	30%	76.8 ± 1.7	77.5 ± 1.3	75.8 ± 2.7	86.9 ± 1.4	80.2 ± 3.0	75.8 ± 2.8	0.62 ± 0.03
	40%	80.4 ± 1.7	80.4 ± 1.6	79.0 ± 1.9	89.2 ± 1.1	83.5 ± 1.6	79.4 ± 1.9	0.68 ± 0.03
	100%	85.3 ± 0.9	85.4 ± 1.0	84.5 ± 1.3	91.9 ± 0.6	88.0 ± 1.0	84.9 ± 1.0	0.76 ± 0.02
SimSiam	10%	73.1 ± 1.5	75.3 ± 2.1	72.6 ± 2.3	85.3 ± 0.9	77.9 ± 2.2	72.1 ± 2.2	0.57 ± 0.02
	20%	76.6 ± 1.5	76.4 ± 2.1	74.9 ± 2.0	86.9 ± 1.0	79.5 ± 2.1	74.7 ± 2.4	0.63 ± 0.03
	30%	78.2 ± 2.4	78.0 ± 2.5	78.2 ± 2.2	88.4 ± 1.1	83.0 ± 1.8	77.9 ± 2.3	0.65 ± 0.03
	40%	80.5 ± 1.2	81.0 ± 1.1	80.5 ± 2.0	89.3 ± 0.8	84.4 ± 1.2	80.5 ± 1.3	0.69 ± 0.02
	100%	85.3 ± 1.5	85.5 ± 1.2	85.2 ± 1.9	92.2 ± 0.8	88.5 ± 1.4	85.2 ± 1.6	0.77 ± 0.02
Fusion-SSL (Our)	**10%**	**75.0 ± 2.6**	**75.5 ± 3.4**	**72.5 ± 3.5**	**86.2 ± 1.6**	**77.9 ± 3.6**	**72.8 ± 4.3**	**0.59 ± 0.05**
	20%	**80.9 ± 0.4**	**80.6 ± 0.8**	**79.5 ± 0.9**	**89.4 ± 0.3**	**83.6 ± 0.9**	**79.3 ± 1.0**	**0.69 ± 0.01**
	30%	**83.1 ± 1.5**	**82.9 ± 1.1**	**81.6 ± 1.8**	**90.6 ± 1.0**	**85.6 ± 1.6**	**81.9 ± 1.6**	**0.73 ± 0.03**
	40%	**84.5 ± 1.3**	**84.2 ± 0.9**	**83.9 ± 2.1**	**91.7 ± 0.9**	**87.5 ± 1.6**	**83.9 ± 1.6**	**0.75 ± 0.02**
	100%	**87.6 ± 1.0**	**87.4 ± 1.2**	**87.2 ± 1.3**	**93.4 ± 0.7**	**90.2 ± 1.0**	**87.2 ± 1.1**	**0.80 ± 0.02**

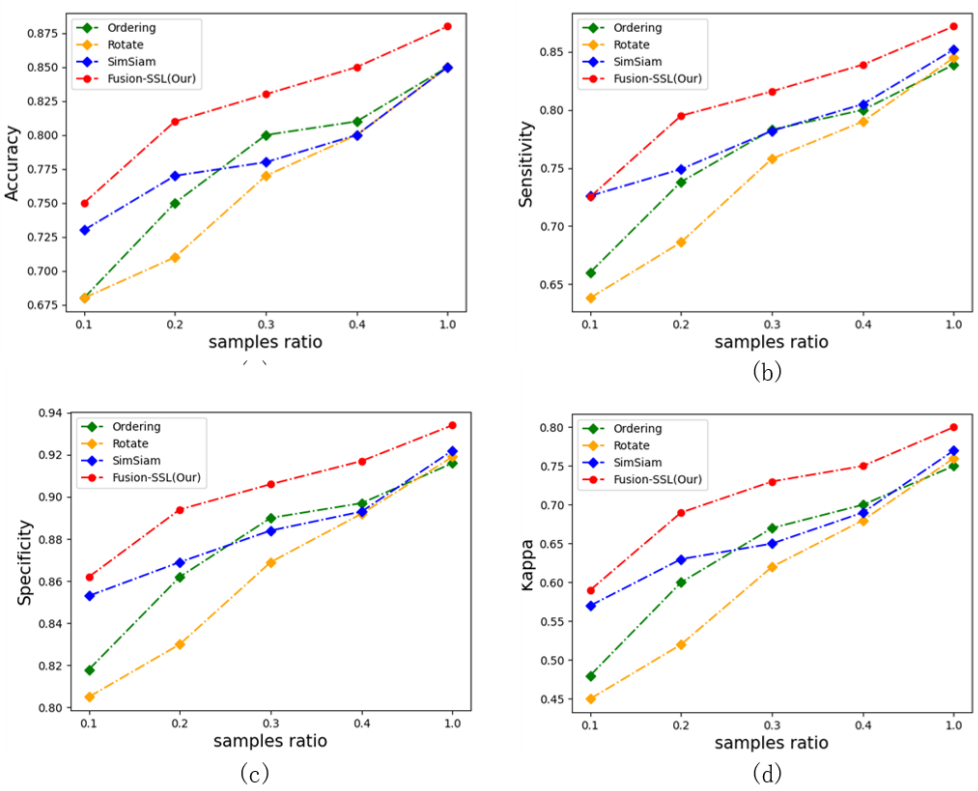

Figure 6. Evaluation metrics of Fusion-SSL and single SSL methods on 100% labeled images.

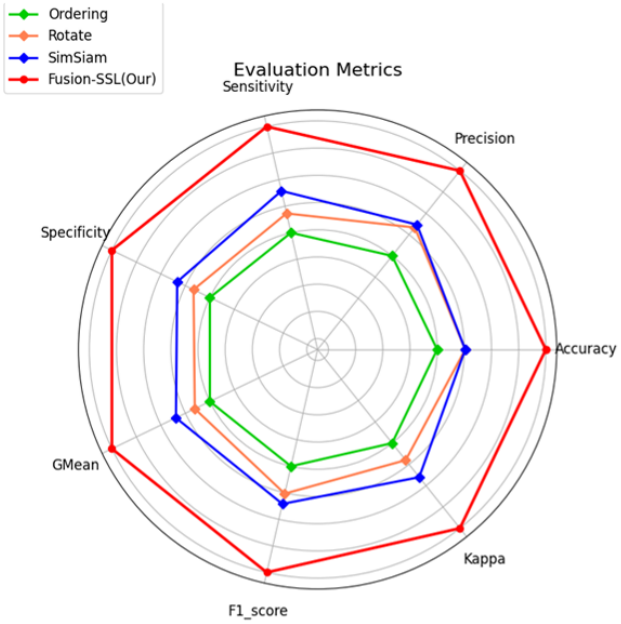

Figure 7. Radar chart of evaluation metrics of Fusion-SSL and single SSL methods on 100% labeled images.

Figures 8 and 9 demonstrate the ROC curves and PR curves, further indicating the superior performance of Fusion-SSL on different scales of labeled training images. Table 4 and Figures 8 and 9 also show that the performance of Fusion-SSL on 40% labeled training images (accuracy = 0.845, precision = 0.842, sensitivity = 0.839, specificity = 0.917, G-Mean = 0.875, F1 = 0.839, kappa = 0.75, ROC-AUC = 0.940, PR-AUC = 0.880) is very close to the performance of the predicting image rotation angle method (Rotating) on 100% labeled training images (accuracy = 0.853, precision = 0.854, sensitivity = 0.845, specificity = 0.919, G-Mean = 0.880, F1 = 0.849, kappa = 0.76, ROC-AUC = 0.940, PR-AUC = 0.890).

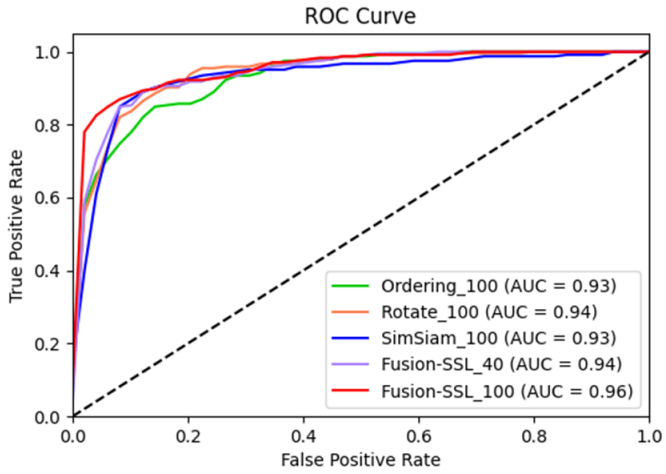

Figure 8. ROC curves of Fusion-SSL and single SSL methods on the labeled training images of different proportions, where the numbers 40 and 100 represent the proportions of the labeled training images and the numbers in parentheses represent the AUC.

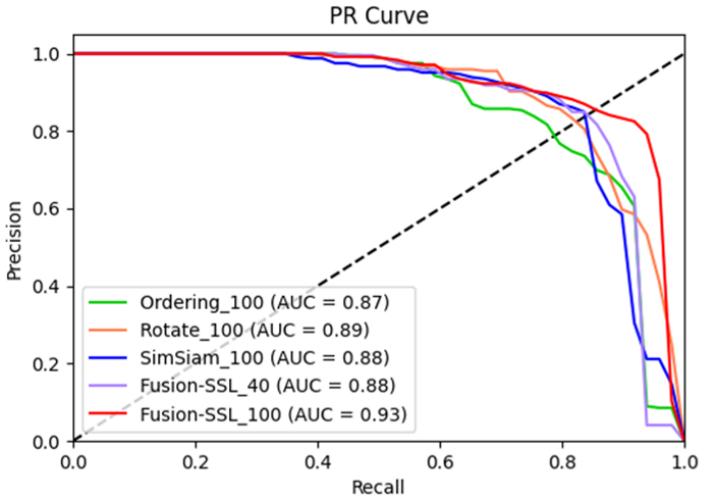

Figure 9. PR curves of Fusion-SSL and single SSL methods on the labeled training images of different proportions, where the numbers 40 and 100 represent the proportions of the labeled training images and the numbers in parentheses represent the AUC.

4. Discussion

Strokes are the primary cause of mortality and morbidity in upper-middle-income and high-income countries. The rupture of atherosclerotic plaques is one of the most significant causes of strokes and acute coronary syndrome. In this research, we proposed a Fusion-SSL method for automatic carotid plaque classification, which improved classification accuracy on a small set of labeled training images and demonstrated consistency with expert classification.

Regarding SSL and supervised learning in carotid plaque ultrasound image classification, although existing supervised neural networks have provided great classification performance for carotid plaques, their performance heavily relies on the availability of a substantial number of labeled training samples. Unfortunately, obtaining these samples was difficult due to the time and effort required for manual data annotation. The experiments have found that for popular CNNs (ShuffleNet-V2, EfficientNet-B0, MobileNet-V3, and ResNet-101), carotid plaque classification performance decreased as the number of labeled training images decreased (Table 3). To efficiently relieve the issue of limited labeled carotid plaque ultrasound images, our study employed the SSL method. For example, the classifying image block order method on 40% labeled training images (accuracy = 0.812, precision = 0.809, sensitivity = 0.800, specificity = 0.890, F1 = 0.801, kappa = 0.70) performed close to the ResNet-101 method on 100% labeled training images (accuracy = 0.833, precision = 0.831, sensitivity = 0.821, specificity = 0.911, F1 = 0.824, kappa = 0.73). Taking accuracy as an example, the classifying image block order method on 40% labeled training images(accuracy = 0.812) accounts for 97.5% of the accuracy produced by the ResNet-101 method with 100% labeled training images(accuracy = 0.833). These results suggest that SSL methods can reduce the workload of manual annotation and have great potential for clinical use.

We compared Fusion-SSL with a single SSL method. Although a single SSL method can effectively relieve the issue of limited labeled training images and enhance the performance of the basic CNN, the challenge lies in how to learn richer plaque features through pretext tasks. To address this issue and improve the performance of CNN in carotid plaque classification, this paper redirected the research focus toward the fusion of self-supervised networks. We proposed a Fusion-SSL method to fuse two different SSL methods. To validate the effectiveness of Fusion-SSL, a comparison was made between the Fusion-SSL method and single SSL methods. The experimental results in Table 4 showed that the performance of Fusion-SSL was superior to the single SSL methods in different cases (labeled training images ranging from 10 to 100%). This suggests that Fusion-SSL can acquire more valuable information about carotid plaques than a single SSL method.

Although feature fusion has advantages in improving the model performance, it also has some disadvantages and limitations, which need to be balanced and processed in practical application. With the increase in dataset size and network depth, feature fusion may increase memory usage and computation time. Additionally, if there are too many features or a high correlation among features, it may lead to model overfitting, which consumes time and computational resources to perform multiple experiments to avoid. SSL methods have enormous potential in carotid plaque ultrasound image classification and deserve more in-depth studies. However, due to inherent limitations, the proposed Fusion-SSL method in this study only combines two simple SSL methods, and the complex features of the plaque have not been fully understood yet. In subsequent research, we need to explore more pretext tasks to extract the visual features of carotid plaques and integrate more SSL methods to enhance the classification

performance of carotid plaques with a few labeled training images.

5. Conclusions

In this paper, we proposed a self-supervised fusion network for carotid plaque ultrasound image classification, which enabled us to obtain richer features of carotid plaque to enhance classification performance. The experimental results demonstrated that the proposed method effectively mitigates the impact of limited labeled images on the accuracy of carotid plaque classification and outperforms single SSL methods with significant improvements in classification performance under different proportions of labeled training images. This indicates that our method is beneficial for the clinical recognition of carotid plaques and stroke warning, providing a new research approach in the field of carotid plaque ultrasound image classification.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant No. 62201203, the Natural Science Foundation of Hubei Province, China under grant No. 2021CFB282, and the Doctoral Scientific Research Foundation of Hubei University of Technology and China under grant No. BSDQ2020064.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. P. J. Modrego, M. A. Pina, M. M. Fraj, N. Llorens, Type, causes, and prognosis of stroke recurrence in the province of teruel, spain. a 5-year analysis, *Neurol. Sci.*, **21** (2000), 355–360. https://doi.org/10.1007/s100720070050
2. V. L. Feigin, R. V. Krishnamurthi, P. Parmar, B. Norrving, G. A. Mensah, D. A. Bennett, et al., Update on the global burden of ischemic and hemorrhagic stroke in 1990–2013: the GBD 2013 study, *Neuroepidemiology*, **45** (2015), 161–176. https://doi.org/10.1159/000441085
3. S. S. Ho, Current status of carotid ultrasound in atherosclerosis, *Quant. Imaging Med. Surg.*, **6** (2016), 285–296. https://doi.org/10.21037/qims.2016.05.03
4. K. Lekadir, A. Galimzianova, A. Betriu, M. D. M. Vila, L. Igual, D. L. Rubin, et al., A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound, *IEEE J. Biomed. Health Inf.*, **21** (2017), 48–55. https://doi.org/10.1109/JBHI.2016.2631401

5. J. Zhan, J. Wang, Z. Ben, H. Ruan, S. Chen, Recognition of angiographic atherosclerotic plaque development based on deep learning, *IEEE Access*, **7** (2019), 170807–170819. https://doi.org/10.1109/ACCESS.2019.2954626

6. W. Ma, X. Cheng, X. Xu, F. Wang, R. Zhou, A. Fenster, et al., Multilevel strip pooling-based convolutional neural network for the classification of carotid plaque echogenicity, *Comput. Math. Methods Med.*, **2021** (2021). https://doi.org/10.1155/2021/3425893

7. W. Ma, R. Zhou, Y. Zhao, Y. Xia, A. Fenster, M. Ding, Plaque recognition of carotid ultrasound images based on deep residual network, in *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, (2019), 931–934. https://doi.org/10.1109/ITAIC.2019.8785825

8. M. Zreik, R. W. van Hamersvelt, J. M. Wolterink, T. Leiner, M. A. Viergever, I. Isgum, A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary ct angiography, *IEEE Trans. Med. Imaging*, **38** (2019), 1588–1598. https://doi.org/10.1109/TMI.2018.2883807

9. Q. Huang, H. Tian, L. Jia, Z. Li, Z. Zhou, A review of deep learning segmentation methods for carotid artery ultrasound images, *Neurocomputing*, **545** (2023), 126298. https://doi.org/10.1016/j.neucom.2023.126298

10. Q. Huang, L. Jia, G. Ren, X. Wang, C. Liu, Extraction of vascular wall in carotid ultrasound via a novel boundary-delineation network, *Eng. Appl. Artif. Intell.*, **121** (2023). https://doi.org/10.1016/j.engappai.2023.106069

11. Q. Huang, L. Zhao, G. Ren, X. Wang, C. Liu, W. Wang, NAG-Net: Nested attention-guided learning for segmentation of carotid lumen-intima interface and media-adventitia interface, *Comput. Biol. Med.*, **156** (2023), 1588–1598. https://doi.org/10.1016/j.compbiomed.2023.106718

12. L. Cai, E. Zhao, H. Niu, Y. Liu,T. Zhang, D. Liu, et al., A machine learning approach to predict cerebral perfusion status based on internal carotid artery blood flow, *Comput. Biol. Med.*, **164** (2023). https://doi.org/10.1016/j.compbiomed.2023.107264

13. S. Gidaris, A. Bursuc, N. Komodakis, P. Perez, M. Cord, Boosting few-shot visual learning with self-supervision, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), 8058–8067. https://doi.org/10.1109/ICCV.2019.00815

14. W. Bai, C. Chen, G. Tarroni, J. Duan, F. Guitton, S. E. Petersen, et al., Self-supervised learning for cardiac MR image segmentation by anatomical position prediction, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019*, **11765** (2019), 541–549. https://doi.org/10.1007/978-3-030-32245-8_60

15. N. A. Koohbanani, B. Unnikrishnan, S. A. Khurram, P. Krishnaswamy, N. Rajpoot, Self-path: self-supervision for classification of pathology images with limited annotations, *IEEE Trans. Med. Imaging*, **40** (2021), 845–2856. https://doi.org/10.1109/TMI.2021.3056023

16. C. Abbet, I. Zlobec, B. Bozorgtabar, J. P. Thiran, Divide-and-rule: self-supervised learning for survival analysis in colorectal cancer, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020*, **12265** (2020), 480–489. https://doi.org/10.1007/978-3-030-59722-1_46

17. A. S. Hervella, J. Rouco, J. Novo, M. Ortega, Self-supervised multimodal reconstruction of retinal images over paired datasets, *Expert Syst. Appl.*, **161** (2020). https://doi.org/10.1016/j.eswa.2020.113674
18. L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, D. Rueckert, Self-supervised learning for medical image analysis using image context restoration, *Med. Image Anal.*, **58** (2019). https://doi.org/10.1016/j.media.2019.101539
19. J. Yan, H. Gan, X. Xu, Z. Yang, Z. Ye, SSCPC-Net: Classification of carotid plaques in ultrasound images using a self-supervised convolutional neural network, in *2022 China Automation Congress (CAC)*, (2022), 4504–4509. https://doi.org/10.1109/CAC57257.2022.10055587
20. S. Gidaris, P. Singh, N. Komodakis, Unsupervised representation learning by predicting image rotations, preprint, arXiv:1803.07728. https://doi.org/10.48550/arXiv.1803.07728
21. E. Picano, M. Paterni, Ultrasound tissue characterization of vulnerable atherosclerotic plaque, *Int. J. Mol. Sci.*, **16** (2015), 10121–10133. https://doi.org/10.3390/ijms160510121
22. X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2019), 510–519. https://doi.org/10.1109/CVPR.2019.00060
23. M. Grandini, E. Bagli, G. Visani, Metrics for multi-class classification: an overview, preprint, arXiv:2008.05756. https://doi.org/10.48550/arXiv.2008.05756
24. N. Ma, X. Zhang, H. Zheng, J. Sun, Shufflenet v2: practical guidelines for efficient CNN architecture design, in *Proceedings of the European Conference on Computer Vision (ECCV)*, **11218** (2018), 116–131. https://doi.org/10.48550/arXiv.1807.11164
25. A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, et al., Searching for mobilenetv3, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), 1314–1324. https://doi.org/10.48550/arXiv.1905.02244
26. M. Tan, Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks, in *International Conference on Machine Learning*, **97** (2019), 6105–6114. https://doi.org/10.48550/arXiv.1905.11946
27. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770–778. https://doi.org/10.48550/arXiv.1512.03385
28. X. Chen, K. He, Exploring simple siamese representation learning, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), 15750–15758. https://doi.org/10.48550/arXiv.2011.10566

AIMS Press