

Research Article

MTAL: A Novel Chinese Herbal Medicine Classification Approach with Mutual Triplet Attention Learning

Wangli Hao ¹, Meng Han ¹, Shancang Li ², and Fuzhong Li ¹

¹School of Software, Shanxi Agricultural University, Shanxi, China

²Department of Computer Science at University of the West of England, Bristol BS16 1QY, UK

Correspondence should be addressed to Fuzhong Li; lifuzhong@sxau.edu.cn

Received 22 September 2021; Revised 14 November 2021; Accepted 12 January 2022; Published 27 February 2022

Academic Editor: Ivan Lee

Copyright © 2022 Wangli Hao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chinese herbal medicine classification is a critical task in medication distribution and intelligent medicine, as well as a significant topic in computer vision. However, the majority of contemporary mainstream techniques are semiautomatic, with low efficiency and performance. To tackle this problem, a novel Chinese herbal medicine classification approach, Mutual Triplet Attention Learning (MTAL), is proposed. The motivation of our approach is to leverage a group of student networks to learn collaboratively and teach each other about cross-dimension dependencies throughout the training process, with the goal of quickly gaining strong feature representations and improving the outcomes. The results of the experiments show that MTAL outperforms other models in terms of accuracy and computation time. MTAL, in particular, improves accuracy by over 5.5 percent while reducing calculation time by over 50 percent.

1. Introduction

Two essential optimization goals for image classification, such as the Chinese herbal medicine (CHM) classification, are computational efficiency and classification accuracy. The CHM classification can be utilized for a variety of fields, including intelligent dispensing and pharmaceutical recommendations.

In order to perform Chinese herbal medicine classification and solve the high-dimensional nonlinear problem in the data of CHM, Luo et al. [1] developed a new approach that applying the linear discriminant analysis (LDA) and locally linear embedding algorithm (LLE) to conduct the CHM classification, whereas the dataset utilized in their paper is quite small that only includes six classes. To classify the leaves plant, Zhang et al. [2] applied a supervised local projection strategy and obtained promising results. Unger et al. [3] proposed a novel method that leverages the support vector machine (SVM) with morphometric and Fourier characteristics to perform the CHM classification based on two test datasets. These two datasets contain 17 and 26 classes separately, and each category has about 10 samples. Spe-

cifically, their method obtained the corresponding classification accuracy of 84% and 73.21% on two datasets, respectively. In the task of CHM classification, Luo et al. [4] compared two methods, PCA and SVM. And they found that SVM achieved better performance. Further, the SOM method has been applied to perform CHM classification [5]. Although promising performances have been achieved by these methods, the dataset they employed is quite small, and there are few samples in each category. However, these methods based on hand-crafted features with less robustness lead to poor classification performance. To achieve enhanced results, deep neural networks have been taken into consideration [6]. However, the model is large and has many parameters, which restricts their utilization in platforms or applications with fast execution or low memory demands, e.g., mobile phones.

Mutual learning is a practical method for achieving promising classification results, through tiny yet strong deep neural networks. The mutual learning begins with a group of students who synchronously learn to tackle the task together [7]. Concretely, each student is bound by two losses: a typical supervised learning loss, and a mimicry loss, which drives

each student's class probabilities to correspond with those of other students. Although mutual learning has obtained superior classification performance with less time consumption, it overlooks the interdependencies among spatial locations or channels.

Triplet attention is a simple yet effective attention strategy that may establish dependencies between weighted spatial or channels [8]. Triplet attention, in particular, utilizes a three-branch structure to calculate attention weights by capturing cross-dimension interaction. Triplet attention produces interdimensional relationships via the rotation operation and expresses spatial and interchannel information with low computational cost overhead when applied to an input tensor.

In order to further improve the Chinese herbal medicine classification performance in terms of accuracy and calculation time, this paper develops a novel Mutual Triplet Attention Learning (MTAL) approach by integrating the advantages of mutual learning and triplet attention. Specifically, MTAL allows two student networks to collaborate on parameter updates and learn interchannel and spatial dependencies from one another throughout the training process. These designs allow MTAL to achieve greater CHM classification efficiency and effectiveness by allowing the suggested model to gather more rich and robust features in a shorter amount of time.

In summary, our contribution can be listed as follows:

- (i) To accomplish CHM classification, a novel Mutual Triplet Attention Learning (MTAL) technique is initially developed. MTAL's mutual learning component enables our model to get superior classification results while using significantly less computing time than previous models, in particular, with a 50% reduction in computing time and a 5.5% increase in accuracy. Furthermore, the MTAL's triplet attention unit allows our model to attain spatial and channel attention, which improves the CHM classification results by 5.5% over the model without the triplet attention component. Those advantages of MTAL will allow it to be applied in mobile devices to perform Chinese herbal medicine classification more efficiently and effectively
- (ii) Several experiments have been designed and performed to verify the superiority of our model. Specifically, these experiments include the comparison of our MTAL model with different models, the evaluation of our del based on two identical basic student networks (two single ResNet18 or two single ResNet 50), and the evaluation of our MTAL model based on two distinct student works (one student network is ResNet18 and the other student network is ResNet50). Furthermore, MTAL has achieved promising CHM classification results with accuracy of 81.64%

The remainder of this paper is shown as follows. The materials and methods are described in Section 2. We list

our experimental results in Section 3 and illustrate some discussions in Section 4. Finally, the conclusions are presented in Section 5.

2. Materials and Methods

2.1. Dataset. A CHM classification dataset (CHMC) with 100 classes (see some samples in Figure 1), proposed in our prior work [9], is utilized in the experiments. Specifically, each class in the dataset has a total of 100 samples. Among the dataset, 80% images are employed for training, and 20% images are utilized for testing. Hence, there are a total of 10000 images in the dataset, 2000 samples for testing and 8000 samples for training. Furthermore, the classes of medicinal materials in this dataset are relatively rich.

In particular, from the natural properties of medicinal materials, CHMC contains botanical, mineral, and animal medicines. Among them, botanical Chinese herbal medicine includes cockscomb, rice bud, hematoxylin, and cinnamon twig, respectively, animal medicine (sea cuttlebone, sea dragon, earth dragon, corrugated fruit, and scorpion), and mineral medicine (red stone fat, alum). Moreover, considering the medicinal part, CHMC contains roots (*Asarum*, ginseng), bark (cork, pomegranate peel), seeds (lotus seeds, wild jujube kernels, and orange cores), etc. In addition, the majority of the examples in CHMC have natural backgrounds, which can help with real-world applications.

2.2. The Proposed MTAL Model

2.2.1. MTAL Structure. The proposed MTAL model (see Figure 2) contains two student networks. They perform mutual learning by teaching each other interactively throughout the training stage, which will boost their performance. Furthermore, each student network integrates the triplet attention module, which can capture the cross-dimension interaction by calculating attention weights based on a three-branch structure.

2.2.2. Problem Formulation. The proposed MTAL model with two student networks is formulated as follows (as shown in Figure 2). Assume M samples $\mathcal{X} = \{x_i\}_{i=1}^M$ from C classes; the corresponding label set is denoted as $\mathcal{Y} = \{y_i\}_{i=1}^M$ with $y_i \in \{1, 2, \dots, C\}$.

The predicted probability $p_1^c(x_i)$ of class c for sample x_i from *Network1* is calculated as

$$p_1^c(x_i) = \frac{\exp(g_1^c(x_i))}{\sum_{c=1}^C \exp(g_1^c(x_i))}, \quad (1)$$

where the logit $g_1^c(x_i)$ is obtained from the "softmax" layer of *Network1* for x_i .

The loss functions \mathcal{L}_{ϕ_1} and \mathcal{L}_{ϕ_2} for *Network1* and *Net work2* can be formulated as follows:

$$\mathcal{L}_{\phi_1} = (1 - \alpha) \times L_{l1} + \alpha \times D_{KL}(p_2 \| p_1) \quad \mathcal{L}_{\phi_2} = (1 - \alpha) \times L_{l2} + \alpha \times D_{KL}(p_1 \| p_2), \quad (2)$$



FIGURE 1: Samples of CHM dataset.

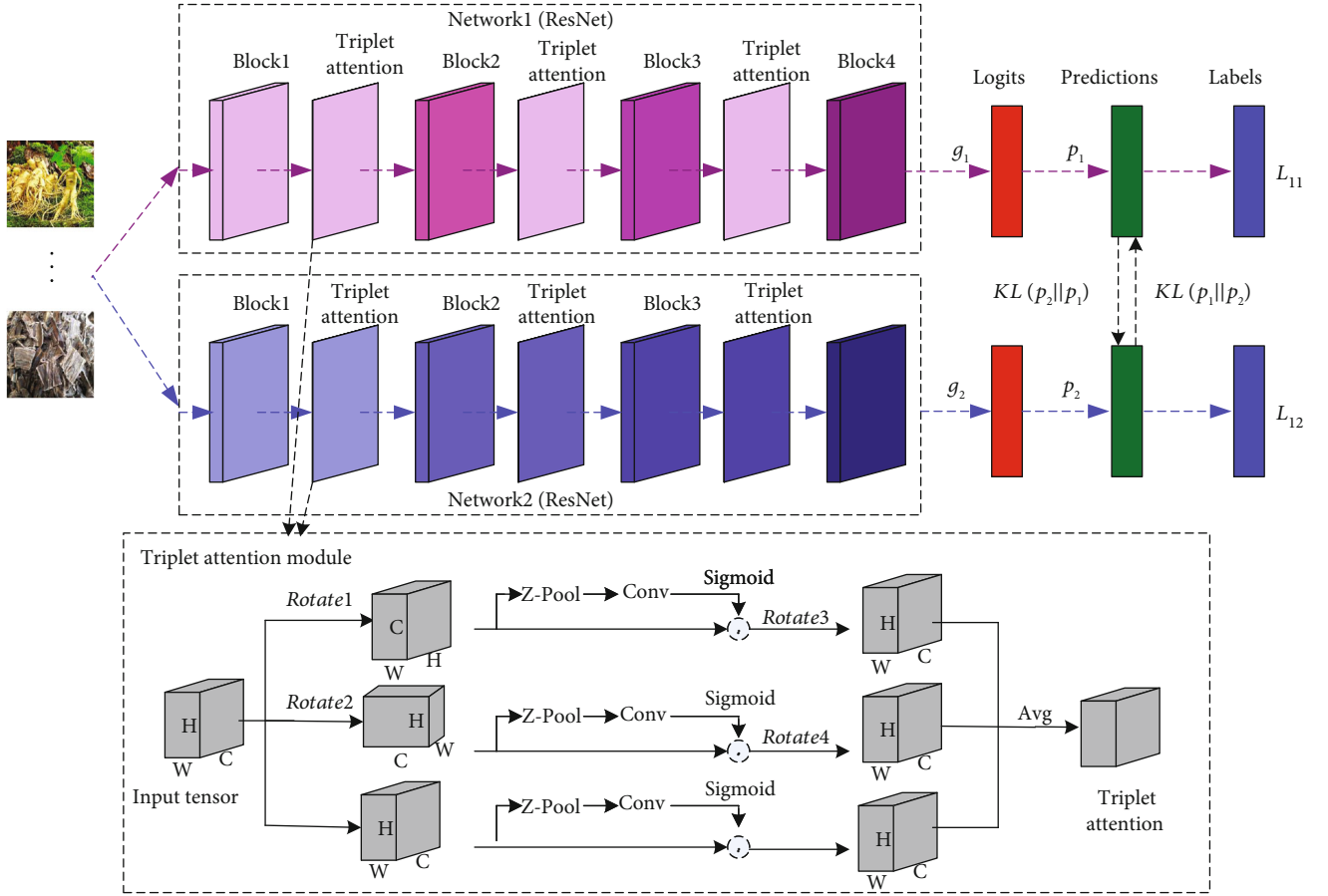


FIGURE 2: The framework of the MTAL model. The MTAL model contains two student networks. They perform mutual learning by teaching each other interactively throughout the training stage, which will boost their performance. Furthermore, each student network integrates the triplet attention module, which can capture the cross-dimension interaction by calculating attention weights based on a three-branch structure.

where L_{l1} and L_{l2} denote the conventional cross entropy losses in classification tasks and $D_{KL}(p_2||p_1)$ and $D_{KL}(p_1||p_2)$ indicate the Kullback-Leibler (KL) divergence losses. α indicates hyperparameters controlling the strengths of two loss terms. Among them, L_{l1} can be obtained by the following equation.

$$L_{l1} = - \sum_{i=1}^M \sum_{c=1}^C \mathbb{I}_{y_i^c} \times \log(p_1^c(x_i)), \quad (3)$$

where y_i is the true label for x_i , y_i^c is an indicator, if $y_i = c$, $y_i^c = 1$, $y_i \neq c$, and $y_i^c = 0$, and L_{l1} denotes the cross entropy error between the correct labels and the predicted values, which can enforce the model to predict the correct results for the training samples.

To enhance the generalization capacity of *Network1* on the testing samples, we employ another peer *Network2* to offer training experience via its posterior probability p_2 . In order to quantify the matching degree of the predictions p_1

and p_2 , the KL divergence is utilized. $D_{KL}(p_2||p_1)$ indicates the KL distance from p_1 to p_2 and can be achieved through

$$D_{KL}(p_2||p_1) = \sum_{i=1}^M \sum_{c=1}^C \sum_{\mathcal{X}} p_2^c(x_i)^c \log \frac{p_2^c(x_i)}{p_1^c(x_i)}, \quad (4)$$

where $p_2^c(x_i)$ has a similar meaning to that of $p_1^c(x_i)$ of Equation (1).

L_{l2} and $D_{KL}(p_1||p_2)$ can be obtained by the similar ways with those of L_{l1} and $D_{KL}(p_2||p_1)$.

Considering the triplet attention module (TAM) in our MTAL method, they are integrated after each block of the corresponding ResNet. The purpose of TAM is to capture the cross-dimensional dependencies via rotating the input feature maps followed by residual transformation. Specifically, triplet attention contains three branches. The top branch is used to achieve the dependencies between the channel and the spatial dimensions (C and W). The middle branch is utilized to obtain the dependencies between the channel and the spatial dimensions (C and H). The bottom branch is employed to compute attention weights across the spatial dimensions (W and H). Finally, the simple average of the three-branch weights is utilized as the final weight.

Rotate1 and *Rotate2* (see Figure 2) in TAM indicate the operations that rotate the input tensor 90° anticlockwise along the W and H axes, respectively. *Rotate3* and *Rotate4* represent the operations that rotate the input tensor 90° clockwise along the W and H axis, respectively.

The Z-pool layer in the TAM is employed for reducing the first dimensionality of the input tensor to two via concatenating the max pooled and average pooled features across the corresponding dimension. The Z-pool operation can be formulated as the following equation:

$$Z - \text{pool}(\mathcal{F}) = [\text{MaxPool}_{1d}(\mathcal{F}), \text{AvgPool}_{1d}(\mathcal{F})], \quad (5)$$

where $1d$ indicates the max and average pooling operations over the first dimension of the corresponding tensor.

The Conv layer in the TAM denotes the standard two-dimensional convolutional layers with kernel size k (k is an empirical parameter, specified 7 in our paper). The Sigmoid in the TAM indicates the sigmoid activation function, which has the following definition:

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)}. \quad (6)$$

2.2.3. Model Training. For a fair comparison, the models in the experiment uniformly adopt SGD to learn the parameters, with a batch size 32. The learning rate starts with 0.01 and then decreases to 1/10 per 80 epochs. The momentum is 0.9. α is 0.8. We stop our training at 200 epochs.

2.2.4. Evaluation Criteria. To compare the performance of different models, some evaluation criteria including accuracy, parameters, FLOPs [8], and loss are adopted. The accuracy reflects the performance of the model. Parameters denote the number parameters of the model; the smaller,

the better. FLOPs indicate the floating point operations needed to be performed per second; the smaller, the better. Parameters and FLOPs illustrate the efficiency of the model. Note that all experiments are executed on TITANX GPUs.

3. Results

In this section, we will describe our experiments in detail. The proposed model's efficiency is investigated through five separate experiments, including the evaluation of the performance of MTAL, evaluation of the effectiveness of mutual learning, evaluation of the effectiveness of triplet attention, evaluation of the MTAL model based on the identical student networks, and evaluation of the MTAL model based on two distinct student networks.

3.1. Evaluation of the Performance of the MTAL Model. To explore the performances of the proposed models, we compare various popular deep neural networks. The comparison models include MobileNet [10], ResNet [11], and Xception71 [12], with results presented in Table 1. Table 1 illustrates that the MTAL model obtains the best results, outperforming MobileNetV2, MobileNetV3, ResNet18, ResNet50, and Xception71 by 15.39%, 11.76%, 10.25%, 6.16%, and 4%, which validates the superiority of the MTAL model in CHM classification task.

3.2. Evaluation of the Effectiveness of Mutual Learning. To evaluate the effectiveness of mutual learning, the single ResNets without mutual learning and the mutual learning models are utilized to perform comparisons. These models are simply represented as SigRes and MulRes, respectively. Specifically, SigRes i indicates the ResNet model with i th layers, and MulRes $i(j)$ denotes the ResNet i model has received the knowledge from the other ResNet j model during training. Comparison results are shown in Tables 2–4.

Tables 2 and 3 show the comparison of models with or without mutual learning based on two identical student networks. Table 2 shows the comparison of ResNet18, and Table 3 denotes the comparison of ResNet50, respectively. From these two tables, we can see that the models with mutual learning obviously outperform those models without, with an accuracy increase of about 4.81% 5.13%. Those results validate the effectiveness of mutual learning.

Table 4 illustrates the comparison of models with or without mutual learning based on two distinct student networks, ResNet18 and ResNet50. From this table, we can see that the models with mutual learning obviously outperform those models without, with an accuracy increase of about 4.94% and 6.41%. Furthermore, the small student network with mutual learning (MulRes18(50)) has achieved better results than the large network without mutual learning (SigRes50), with accuracy increase of 2.47% and parameter decrease of 53.98%, which further verifies the effectiveness and efficiency of mutual learning.

3.3. Evaluation of the Effectiveness of Triplet Attention. To evaluate the effectiveness of triplet attention, the ResNet models with and without triplet attention modules are utilized to perform comparisons. These models are simply

TABLE 1: Comparison of various models based on 100 types of CHM.

Models	MobileNetV2	MobileNetV3	ResNet18	ResNet50	Xception71	MTAL
Accuracy	70.75	73.05	74.05	76.90	78.5	81.64

TABLE 2: Comparisons of different models with or without mutual learning based on ResNet18.

Model	Parameters	FLOPs	Accuracy
SigRes18	11.69 M	1.82 G	74.05
MulRes18(18)-1	11.69 M	1.82 G	77.63
MulRes18(18)-2	11.69 M	1.82G	77.85

TABLE 3: Comparisons of different models with or without mutual learning based on ResNet50.

Model	Parameters	FLOPs	Accuracy
SigRes50	25.56 M	4.12 G	76.90
MulRes50(50)-1	25.56 M	4.12 G	80.40
MulRes50(50)-2	25.56 M	4.12 G	80.60

TABLE 4: Comparisons of different models with or without mutual learning based on ResNet18 and ResNet50.

Model	Parameters	FLOPs	Accuracy
SigRes18	11.69 M	1.82 G	74.05
SigRes50	25.56 M	4.12 G	76.90
MulRes18(50)	11.69 M	1.82 G	78.80
MulRes50(18)	25.56 M	4.12 G	80.70

represented as SigRes and SigAttRes, respectively. Specifically, SigRes i /SigAttRes i indicates the ResNet model with i th layers; comparison results are shown in Tables 5 and 6.

Tables 5 and 6 show the comparison of models with or without triplet attention based on ResNet18 and ResNet50, respectively. From these two tables, we can see that the models with triplet attention are obviously superior to those models without, with an accuracy increase of about 1.42% and 2.01%, which validates the effectiveness of triplet attention.

3.4. Evaluation of the MTAL Model Based on Two Identical Student Networks. In this section, we verify the performance of the proposed MTAL based on two identical student networks. Specifically, MTAL utilizes the ResNet backbones. Furthermore, the backbone network with different layers is employed to validate the generalization of the proposed approach, including ResNet18 and ResNet50 separately here.

The single convolutional neural network ResNet [11], single ResNet with triplet attention [8], and mutual learning model [7] based on the same ResNet backbone, respectively, are adopted to compare with the proposed MTAL with two identical basic student networks. These four models are sim-

TABLE 5: Comparisons of different models with or without triplet attention based on ResNet18.

Model	Parameters	FLOPs	Accuracy
SigRes18	11.69 M	1.82 G	74.05
SigAttRes18	11.69 M	1.83 G	75.10

TABLE 6: Comparisons of different models with or without triplet attention based on ResNet50.

Model	Parameters	FLOPs	Accuracy
SigRes50	25.56 M	4.12 G	76.90
SigAttRes50	25.56 M	4.17 G	78.45

ply named as SigRes, SigAttRes, MulRes, and MTALRes, respectively. Comparison results are shown in Figure 3.

Figure 3 illustrates the accuracy and loss of different comparison models based on two identical student networks, under different training epochs. Figures 3(a) and 3(b) show the comparison of accuracy of different models, and Figures 3(c) and 3(d) illustrate the comparison of loss of corresponding models in Figures 3(a) and 3(b). Specifically, in Figure 3, MulRes18(18)/MTALRes18(18) indicates the one ResNet18/ResNet18+triplet attention model learning experience from another model with the same architecture. For MulRes18(18)- i /MTALRes18(18)- i , i denotes the index of two mutual learning models. MulRes50(50)- i /MTALRes50(50)- i has a similar meaning to that of MulRes18(18)- i /MTALRes18(18)- i .

It is concluded that the results of SigAttRes18/SigAttRes50 exceed those of SigRes18/SigRes50 by 1.1%/2.01%, which validates the effectiveness of the triplet attention. MulRes18(18)/MulRes50(50) outperforms the corresponding Single ResNet models by 2.7%-5.1%, verifying the superiority of mutual learning. Moreover, both MTALRes18(18) and MTALRes50(50) obtain the best results in their corresponding counterparts at both performance and efficiency. Specifically, MTALRes18(18)/MTALRes50(50) surpasses SigRes18/SigRes50, SigAttRes18/SigAttRes50, and MulRes18(18)/MulRes50(50) by about 6.2%/6.1%, 4.7%/4.0%, and 1.3%/1.5% on accuracy and about 10%/19%, 78%/79%, and 80%/83% in terms of loss, respectively. The reason is that MTAL leverages the benefits of both mutual learning and triplet attention, which allows the network to learn cross-dimension dependencies from the other one.

Tables 7 and 8 summarize the comparisons of different models based on the corresponding evaluation criteria. The proposed MTAL models outperform SigRes, SigAttRes, and MulRes in terms of accuracy, with almost negligible parameters and FLOP increase (0.02% and 1%). In addition, MTALRes18(18) can acquire 2.3% better accuracy than

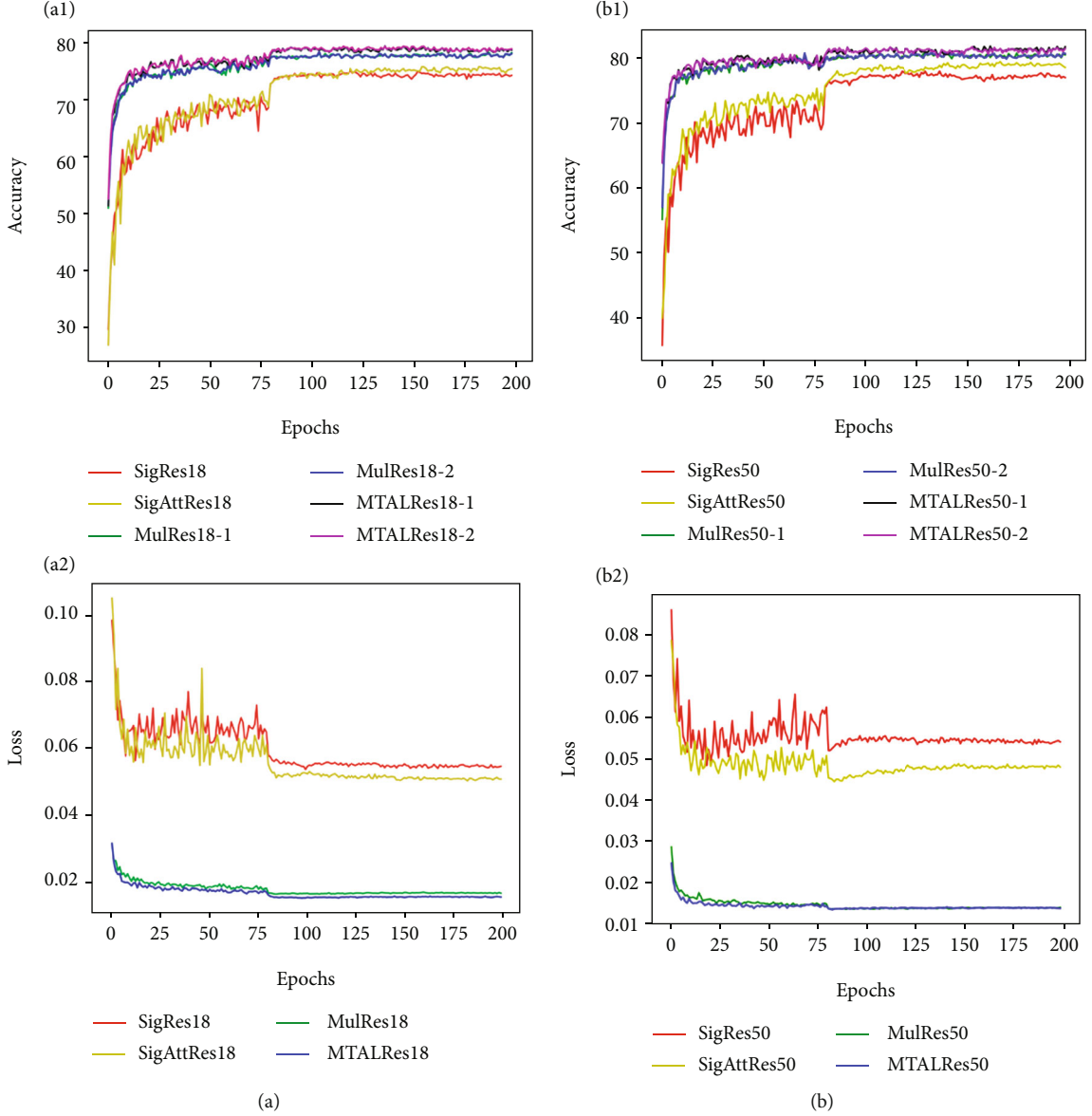


FIGURE 3: Comparison results of different models in terms of accuracy and loss based on the same basic network. (a) Comparison of accuracy of different models, including single ResNet18 model (SigRes18), single ResNet18+triplet attention (SigAttRes18), two single ResNet18 with knowledge learned from the other ResNet18 (MulRes18-1 and MulRes18-2), and two SigAttRes18 with knowledge learned from the other SigAttRes50 (MTALRes18-1 and MTALRes18-2). (b) Comparison of accuracy of different models, including single ResNet50 model (SigRes50), single ResNet50+triplet attention (SigAttRes50), two single ResNet50 with knowledge learned from the other ResNet50 (MulRes50-1 and MulRes50-2), and two SigAttRes50 with knowledge learned from the other SigAttRes50 (MTALRes50-1 and MTALRes50-2). (c) The comparison of losses of corresponding models of (a). (d) The comparison of losses of corresponding models of (b).

TABLE 7: Comparisons of different models for the classification of CHM based on the same basic student network ResNet18.

Model	Parameters	FLOPs	Accuracy
SigRes18	11.69 M	1.82 G	74.05
SigAttRes18	11.69 M	1.83 G	75.10
MulRes18(18)-1	11.69 M	1.82 G	77.63
MulRes18(18)-2	11.69 M	1.82 G	77.85
MTALRes18(18)-1 (ours)	11.69 M	1.83 G	78.50
MTALRes18(18)-2(ours)	11.69 M	1.83 G	78.70

TABLE 8: Comparisons of different models for the classification of CHM based on the same basic student network ResNet50.

Model	Parameters	FLOPs	Accuracy
SigRes50	25.56 M	4.12 G	76.90
SigAttRes50	25.56 M	4.17 G	78.45
MulRes50(50)-1	25.56 M	4.12 G	80.40
MulRes50(50)-2	25.56 M	4.12 G	80.60
MTALRes50(50)-1(ours)	25.56 M	4.17 G	81.64
MTALRes50(50)-2(ours)	25.56 M	4.17 G	81.35

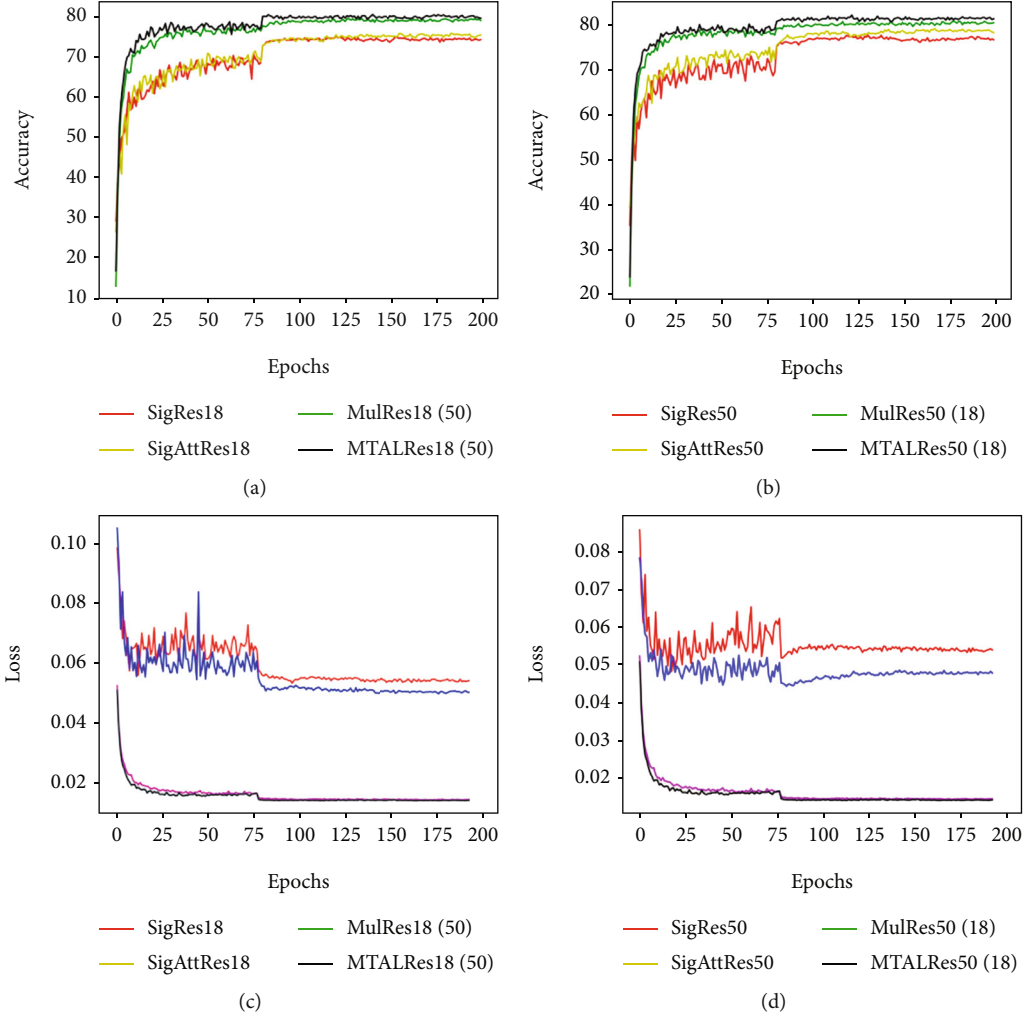


FIGURE 4: Comparison results of different models in terms of accuracy and loss based on the different basic student networks. (a) Comparison of accuracy of different models, including single ResNet18 model (SigRes18), single ResNet18+triplet attention (SigAttRes18), single ResNet18 with knowledge learned from ResNet50 (MulRes18(50)), SigAttRes18 with knowledge learned from SigAttRes50 (MTALRes18(50)). (b) Comparison of accuracy of different models, including single ResNet50 model (SigRes50), single ResNet50+triplet attention (SigAttRes50), single ResNet50 with knowledge learned from ResNet18 (MulRes50(18)), and SigAttRes50 with knowledge learned from SigAttRes18 (MTALRes50(18)). (c) The comparison of loss of the corresponding models of (a). (d) The comparison of loss of the corresponding models of (b).

TABLE 9: Comparisons of different models for the classification of CHM based on the basic network ResNet18.

Model	Parameters	FLOPs	Accuracy
SigRes18	11.69 M	1.82 G	74.05
SigAttRes18	11.69 M	1.83 G	75.10
MulRes18(50)	11.69 M	1.82 G	78.80
MTALRes18(50)(ours)	11.69 M	1.83 G	79.35

SigRes50 and achieve comparable accuracy with SigAttRes50, but its parameters and FLOPs are significantly lower than those of SigRes50 and SigAttRes50 (over 2 times less). Moreover, although MTALRes50(50) needs comparable parameters and FLOPs with other models with the same backbone, it achieves better results.

3.5. Evaluation of the Model Based on Two Distinct Student Networks. In this section, we validate the performance of the proposed MTAL based on two distinct student networks. Specifically, the two student networks are ResNet18 and ResNet50 separately.

The single convolutional neural network ResNet [11], single ResNet with triplet attention [8], and mutual learning model [7] based on different ResNet backbones, respectively, are adopted to compare with the proposed MTAL with different basic student neural networks. These four models are simply named as SigRes, SigAttRes, MulRes, and MTALRes, respectively. MulRes $i(j)$ denotes the model ResNet i with knowledge learned from the model ResNet j . MTALRes $i(j)$ has a similar meaning to that of MulRes $i(j)$. Comparison results are presented in Figure 4.

Figure 4 shows the accuracy and loss of different comparison models based on two distinct student networks,

TABLE 10: Comparisons of different models for the classification of CHM based on the basic network ResNet50.

Model	Parameters	FLOPs	Accuracy
SigRes50	25.56 M	4.12 G	76.90
SigAttRes50	25.56 M	4.17 G	78.45
MulRes50(18)	25.56 M	4.12 G	80.70
MTALRes50(18)(ours)	25.56 M	4.17 G	81.35

under different training epochs. Figures 4(a) and 4(b) show the comparison of accuracy of different models, and Figure 4(c) and 4(d) illustrate the comparison of loss of corresponding models in Figures 4(a) and 4(b).

From Figure 4, we can see that SigAttRes18/SigAttRes50 exceeds SigRes18/SigRes50 by 1.42%/2.01%, which validates the effectiveness of the triplet attention. MulRes18(50)/MulRes50(18) outperforms the corresponding single ResNet (SigRes18/SigRes50) by 6.41%/4.94%, verifying the superiority of mutual learning (even the smaller student can further boost the larger one student). Moreover, both MTALRes18(50) and MTALRes50(18) obtain the best results in their corresponding counterparts at both performance and efficiency. Specifically, MTALRes18(50)/MTALRes50(18) surpasses SigRes18/SigRes50, SigAttRes18/SigAttRes50, and MulRes18(50)/MulRes50(18) by 7.16%/5.79%, 5.65%/3.70%, and 0.69%/0.80% on accuracy and about 9.1%/9.7%, 77.5%/79.2%, and 80.2%/80.7% in terms of loss, respectively, which validates the effectiveness of the proposed model.

Tables 9 and 10 summarize the comparisons of different models based on the corresponding evaluation criteria. The proposed MTAL models outperform SigRes, SigAtt, and MulRes in terms of accuracy, with almost negligible parameters and FLOPs increase (0.02% and 1%). Additionally, MTALRes18(50) can acquire 3.18% and 1.15% better accuracy than those of SigRes50 and SigAttRes50, but its parameters and FLOPs are significantly lower than SigRes50 and SigAttRes50 (over 2 times less). Moreover, although MTALRes50(18) needs comparable parameters and FLOPs with other models with the same backbone (SigAttRes50), it achieves better results.

4. Discussion

The proposed MTAL leverages both mutual learning and triplet attention modules, which may gain cross-dimension knowledge of the spatial dimensions and channel dimension from them and the other student network in an interactive manner. This interactive learning and the cross-dimension dependency capturing capacity allows our model MTAL to achieve promising performance at both accuracy and efficiency. In order to validate the superiority of MTAL, several experiments are designed and conducted, including evaluation of the performance of the MTAL model, evaluation of the MTAL model based on two identical student networks, and evaluation of the MTAL model based on two distinct student networks.

When compared to other popular models, such as MobileNetV2, MobileNetV3, ResNet18, ResNet50, and Xception71, MTAL achieves state-of-the-art performance, demonstrating MTAL's efficacy.

The MTAL model obtains better performance and efficiency when comparing with their corresponding counterparts, including SigRes, SigAttRes, and MulRes, respectively, which validates the generalization performance of our model.

Through triplet attention-enhanced mutual learning, the small student network of our model can even obtain better performance than those of large student networks. For example, MTALRes18(50) achieves 3.18% and 1.15% better accuracy than those of SigRes50 and SigAttRes50, saving even two times complexity and time cost. On the other hand, when the scale of the network is the same with other networks, our model obtains better performance than other models.

In conclusion, our model outperforms the competition in terms of accuracy and efficiency in the CHM classification task.

5. Conclusions

This paper has developed a novel MTAL approach for CHM classification, which combines mutual learning with the triplet attention module for transferring the cross-dimension dependencies from one network to another one. With the help of deep mutual learning, an ensemble of basic student neural networks of our model can update parameters collaboratively and gain information from each other during the whole training process. As a benefit from the triplet attention module, our model can collect interdimensional information via the rotation operation and accomplish interchannel and spatial dependencies with nearly no increase in computational overhead. Leveraging the mutual learning and triplet attention module, our MTAL model has achieved excellent classification performance of Chinese herbal medicines with higher efficiency and effectiveness. When compared to other models, the experimental findings show that MTAL can greatly improve CHM classification performance with minimally compromised training settings and FLOPs. The MTAL model, in particular, delivers 3.18% greater accuracy and 80.7% lower loss than other models.

Further work would explore more efficient mutual learning-based methodologies and a more promising attention-based feature extraction approach to boost the effectiveness and the efficiency of the CHM classification.

Data Availability

The datasets used to support the findings of this study can be obtained by sending email to the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Shanxi Province Higher Education Innovation Project of China (2020L0154), Shanxi Province Educational Science “14th Five-Year Plan” 2021 General Planning Project + Research on Intelligent Agricultural Talent Cultivation Model Driven by “Industry, Learning and Research” of Agricultural and Forestry Colleges (GH-21006), Shanxi Agricultural University 2021 “Neural Network” Course Ideological and Political Project (KCSZ202133), Shanxi Agricultural University Doctoral Research Startup Project (2021BQ88), Intelligent Information Processing, Shanxi Provincial Key Laboratory Open Project Fund (CICIP2021005), and Shanxi Agricultural University Academic Recovery Research Project (2020xshf38). We thank Jian zhou for helpful conversations, Dr. Li for assistance with the samples, and Zhiwu Dong for producing several of the figures.

References

- [1] J. Wang, R. K. W. Wong, and T. C. M. Lee, “Locally linear embedding with additive noise,” *Pattern Recognition Letters*, vol. 123, pp. 47–52, 2019.
- [2] S. Zhang, Y. Lei, T. Dong, and X. P. Zhang, “Label propagation based supervised locality projection analysis for plant leaf classification,” *Pattern Recognition*, vol. 46, no. 7, pp. 1891–1897, 2013.
- [3] J. Unger, D. Merhof, and S. Renner, “Computer vision applied to herbarium specimens of german trees: testing the future utility of the millions of herbarium specimen images for automated identification,” *BMC Evolutionary Biology*, vol. 16, no. 1, p. 248, 2016.
- [4] D. Luo, W. Jia, Y. Chen, and G. H. Hamid, “Classification of chinese herbal medicines based on svm,” in *2014 International Conference on Information Science, Electronics and Electrical Engineering*, vol. 1, pp. 453–456, Sapporo, Japan, April 2014.
- [5] M. Wang, L. Li, C. Yu et al., “Classification of mixtures of chinese herbal medicines based on a self-organizing map (som),” *Molecular Informatics*, vol. 35, no. 3-4, pp. 109–115, 2016.
- [6] S. Liu, W. Chen, and X. Dong, “Automatic classification of chinese herbal based on deep learning method,” in *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Huangshan, China, July 2018.
- [7] T. H. T. L. H. Zhang and Y. Xiang, “Deep mutual learning,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4320–4328, Salt Lake City, UTAH USA, 2018.
- [8] D. Misra, T. Nalamada, A. U. Arasanipalai, and Q. Hou, “Rotate to Attend: Convolutional Triplet Attention Module,” in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, January 2021.
- [9] W. Hao, M. Han, H. Yang, F. Hao, and F. Li, “A novel chinese herbal medicine classification approach based on efficientnet,” *Systems Science and Control Engineering*, vol. 9, no. 1, pp. 304–313, 2021.
- [10] A. Howard, M. Sandler, G. Chu et al., “Searching for mobilenetv3,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324, Seoul, Korea (South), October 2019.
- [11] 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)*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [12] F. Chollet, “Xception: deep learning with depthwise separable convolutions,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.