



OPEN

Integrating deformable CNN and attention mechanism into multi-scale graph neural network for few-shot image classification

Yongmin Liu^{1,3}✉, Fengjiao Xiao^{1,2}, Xinying Zheng^{2,3}, Weihao Deng^{1,3}, Haizhi Ma^{1,3}, Xinyao Su⁴ & Lei Wu^{1,4}

Graph neural networks have excellent performance and powerful representation capabilities, and have been widely used to handle Few-shot image classification problems. The feature extraction module of graph neural networks has always been designed as a fixed convolutional neural network (CNN), but due to the intrinsic properties of convolution operations, its receiving domain is limited. This method has limitations in capturing global feature information and easily ignores key feature information of the image. In order to extract comprehensive and critical feature information, a new CA-MFE algorithm is proposed. The algorithm first utilizes different convolution kernels in CNN to extract multi-scale local feature information, and then based on the global feature extraction ability of attention mechanism, parallel processing of channel and spatial attention mechanism is used to extract multidimensional global feature information. This paper provides a comprehensive performance evaluation of the new model on both mini-ImageNet and tiered ImageNet datasets. Compared with the benchmark model, the classification accuracy has increased by 1.07% and 1.33% respectively; In the 5-way 5-shot task, the classification accuracy of the mini-ImageNet dataset was improved by 11.41%, 7.42%, and 5.38% compared to GNN, TPN, and dynamic models, respectively. The experimental results show that compared with the benchmark model and several representative Few-shot classification algorithm models, the new CA-MFE model has significant superior performance in processing few-shot classification data.

Image classification is an important research topic in the field of computer vision, which utilizes large-scale labeled samples to optimize the parameters of neural network models, thereby significantly improving the accuracy of image classification. The success of deep learning can be attributed to three key factors: powerful computing resources (such as GPUs), complex neural networks (such as CNNs, LSTMs), and large-scale datasets. However, due to various factors such as privacy, security, or high cost of labeling data, many application scenarios such as medical, satellite, and military fields cannot obtain sufficient labeled training samples; To address these issues, Few-shot learning (FSL)^{1–3} methods has been proposed.

The background of FSL methods can be traced back to 2000, when Miller et al.⁴ first focused on the problem of learning from Few-shot and proposed a Congealing algorithm. Since then, more and more people have begun to pay attention to FSL methods. In 2015, Koch et al.⁵ first introduced deep learning methods into FSL, achieving the transformation of FSL from non-deep models to deep models, becoming a watershed for FSL. Subsequent researchers fully utilized the advantages of deep learning and combined it with FSL methods, proposing methods such as data augmentation⁶, metric learning⁷, and meta-learning⁸, thus promoting the rapid development of FSL. Among various existing FSL methods, the method based on Graph Neural Network (GNN)⁹ combines the advantages of metric learning and meta learning. Due to its excellent performance and powerful representation ability in transduction settings, it is now widely used in FSL tasks.

Zheng Q et al.¹⁰ proposed MR-DCAE: a deep convolutional autoencoder based on manifold regularization for unauthorized broadcast recognition. By combining manifold regularization, this method enhances the feature

¹School of Electronic Information and Physics, Central South University of Forestry Science and Technology, Changsha 410004, China. ²Business School of Hunan Normal University, Changsha 410000, China. ³Research Center of Smart Forestry Cloudy, Central South Forestry University of Science and Technology, Changsha 410004, China. ⁴Bangor College China, Central South University of Forestry and Technology, Changsha 410004, China. ✉email: T20040550@csuft.edu.cn

extraction capability of autoencoders in identifying illegal broadcast signals. This method effectively alleviates the overfitting problem in small sample scenarios. MobileViT¹¹ is a lightweight network architecture that combines Transformer and MobileNet, designed to perform efficient image classification tasks in resource-constrained environments. Its advantage is that it can reduce the consumption of computing resources while ensuring accuracy. This method demonstrates the potential of lightweight networks in Few-shot learning, and the design concept of MobileViT inspires how to improve the efficiency of image classification tasks through lightweight models. Zheng Q et al.¹² proposed MobileRaT, a lightweight radio Transformer method used for automatic modulation classification in UAV communication systems. MobileRaT is a lightweight radio Transformer model designed to solve the problem of automatic modulation classification in UAV communication systems. By utilizing the global feature extraction capability of Transformer architecture, it can efficiently handle complex patterns in radio signals. Zheng Q et al.¹³ proposed a real-time transformer discharge pattern recognition method based on Few-shot learning driven CNN-LSTM. This method combines CNN and LSTM networks to process time series and spatial features, and can accurately identify transformer discharge patterns under small sample conditions. It demonstrates the possibility of improving classification accuracy by combining convolutional and recursive neural networks with a small amount of data. Zheng Q et al.¹⁴ proposed a DropPath method based on the PAC-Bayesian framework for pruning 2D discriminative convolutional network. The PAC-Bayesian framework combines the DropPath pruning method, which can reduce the computational complexity of the model while maintaining network performance. This method greatly optimizes the computational efficiency of the model by eliminating unnecessary paths.

Currently, many graph neural network-based methods have been applied to Few-shot learning tasks. Among them, Edge Label Graph Neural Network (EGNN) and Hierarchical Graph Convolutional Network (HGCN) are two popular models¹⁵. In 2019, Kim et al.¹⁶ proposed an Edge Label Graph Neural Network (EGNN), which iteratively updates edge labels using intra cluster similarity and inter-cluster dissimilarity, and uses a deep neural network on the edge label graph to achieve Few-shot learning tasks. In 2021, Hu et al.¹⁷ proposed the Hierarchical Graph Convolutional Network (HGCN), which uses hierarchical graph convolution to improve the accuracy of node classification and provide support for it. In order to improve the stability of graph networks in semi supervised learning, Li et al.¹⁸ added residual networks and attention mechanisms on the basis of graph neural networks to enhance the information transmission capability of graph structures. HGCN adopts a hierarchical graph convolution structure to improve the accuracy of node classification by aggregating node information from different levels. However, this method has some limitations in capturing local features and complex relationships between nodes. HGCN has higher algorithm complexity and implementation difficulty, especially when dealing with small sample data. In contrast, EGNN explicitly uses an edge label update mechanism to handle the similarity and dissimilarity between nodes, allowing for iterative updating of edge labels in Model¹⁹. This method can not only better capture the local relationships between nodes, but also perform explicit clustering evolution in the graph structure. This clustering mechanism is particularly important in Few-shot learning because it can effectively enhance feature representation and classification performance under limited data conditions, and is easier to implement and optimize, making it more suitable for Few-shot image classification tasks²⁰. Therefore, the clustering ability and edge label update mechanism of EGNN in Few-shot learning tasks make it more flexible, accurate, and effective in handling complex graph structured data.

In Few-shot learning tasks, graph neural networks are divided into two main modules: feature extraction module and graph network module. The feature extraction module is usually executed by a convolutional neural network (CNN)²¹, which extracts features from input samples. The graph network module is used to update the feature information of the generated graph to obtain node-to-node similarity for subsequent classification implementation. CNN, as an effective feature extraction network model, has been validated in image classification, image segmentation, object detection, and language recognition research. It achieves local perception by sliding filters to extract local features. Feature extraction information is crucial for graph learning in subsequent graph modules, but convolutional neural modules can only extract features with fixed receptive fields, which will result in shallow details extracted into the network, incomplete feature information acquisition. It ultimately affects the classification accuracy of the image by preventing subsequent graph network modules from effectively learning and updating through node feature information.

In order to extract deeper image feature information in graph neural networks, researchers have conducted corresponding studies in the fields of image classification^{21,22}, image segmentation^{23,24}, and target detection²⁵. Yao et al.²¹ combined a parallel structure consisting of CNN and Recurrent Neural Network (RNN) for image feature extraction in 2019, and introduced a special perceptron attention mechanism to unify the features extracted by the two different neural network structures of the model to enhance deep feature extraction. Zhang et al.²⁶ proposed a parallel multi-branch TransFuse structure in 2021, which combines Transformer and CNN in parallel and effectively captures low-level spatial features and high-level semantic features. In 2019, a dual attention network (DANet) with adaptive self-attention mechanism was proposed, which improves the feature representation of scene segmentation by aggregating contextual dependencies based on self-attention mechanism, adaptively integrating local features and global dependencies in spatial and channel dimensions, and capturing global information. Inspired by the DANet network, a dual branch dual-attention mechanism network (DBDA) for hyperspectral image classification was proposed²⁷. The network includes spectral and spatial branches that capture spectral and spatial features respectively, and uses channel and spatial attention mechanisms to refine the feature map. The fused spectral spatial features are obtained by connecting the outputs of the two branches. A parallel structure combines the advantages of local and global feature extraction, achieving more comprehensive feature representation and more efficient task processing.

Based on the advantages of CNN and attention mechanism, we propose a new CA-MFE algorithm. By combining local and global feature information, this method overcomes the limitations of traditional graph neural networks in capturing global feature information and achieves the task of multi-scale key feature

extraction. This algorithm utilizes different convolution kernel functions in CNN to obtain multi-scale local feature information, and based on the global feature extraction capabilities of attention mechanism, parallel processing channel, and spatial attention mechanism, extracts multidimensional global information, effectively extracting multidimensional global features from images. This algorithm runs a convolutional neural network in parallel with channel and spatial attention modules (kernel sizes of 3×3 and 5×5). CNN captures multi-scale local feature information through different kernel sizes, while attention mechanism extracts multidimensional global features from channel and spatial perspectives. By parallel fusion of two branches, the classification performance of the Few-shot classifier has been improved. Then, the fused feature information is constructed into a graph structure, and EGNN (Edge Gated Graph Neural Network) alternately updates node and edge features to learn the graph structure. The classifier predicts the classification results of the samples.

The experimental results show that the CA-MFE algorithm combines the advantages of deformable CNN and attention mechanism, significantly improving the classification performance of graph neural networks and promoting the classification of Few-shot images. Experiments conducted on mini-ImageNet and hierarchical ImageNet datasets have shown that this method is indeed superior to other representative Few-shot classification methods.

Related work

Deep learning method

In recent years, many deep learning methods have made significant progress in fields such as Few-shot learning, signal processing, and image classification. The following are some studies related to the CA-MFE model proposed in this paper, and some comparisons are made to highlight the advantages of the method proposed in this paper.

Firstly, MR-DCAE improves the feature extraction ability of illegal broadcast signal recognition through manifold regularization and performs well in small sample tasks. However, MR-DCAE relies on a fixed autoencoder structure and has limitations in multi-scale feature extraction. In contrast, the CA-MFE model proposed in this paper flexibly captures local and global features using multi-scale convolution kernels, and further enhances the global and importance of features through channel and spatial attention mechanisms, making it more advantageous in complex image scenes. Secondly, MobileViT achieves efficient image classification in resource constrained environments by combining lightweight convolutions with Transformers. Although its lightweight design performs well in real-time tasks, the global feature extraction of Transformer is computationally intensive. By using multi-scale convolution and parallel attention mechanism, the CA-MFE model not only maintains lightweight, but also flexibly processes multi-scale image features, demonstrating higher efficiency and accuracy in processing Few-shot images. In UAV (unmanned aerial vehicle) communication, MobileRaT uses Transformer structure for automatic modulation classification, mainly adept at processing global signals.

However, in image classification tasks, the processing of fine-grained features is relatively weak. In contrast, the CA-MFE model not only captures global information of images through parallel multi-scale convolution and attention mechanisms, but also accurately extracts local details, making it more comprehensive in Few-shot image classification. CNN-LSTM is applied for real-time recognition of transformer discharge patterns, which excels in processing combinations of temporal and spatial features, but has limited performance in pure image classification. In contrast, the CA-MFE model is specifically designed for Few-shot image classification, enhancing image feature extraction capabilities through multi-scale convolution and attention mechanisms. It can effectively capture local and global features without increasing excessive computational burden, making it suitable for static image classification tasks. Finally, the DropPath pruning method in the Pac Bayes framework reduces the complexity of the model and improves computational efficiency. However, the CA-MFE model aims to optimize the utilization of computational resources through multi-scale convolution and attention mechanisms, ensuring low computational overhead.

In future work, CA-MFE can be further combined with pruning techniques to achieve more efficient model performance. In summary, the CA-MFE model proposed in this paper has significant advantages in multi-scale feature extraction, lightweight design, and comprehensive processing of global and local features, making it particularly suitable for Few-shot image classification tasks in complex scenes.

Graph neural network

Marco Gori et al.²⁸ proposed the concept of GNN in 2005, and in 2009 Franco et al.²⁹ defined the theoretical foundation of GNN, inherited and developed the GNN algorithm, as shown in Fig. 1, and made some degree of improvement. Kipf et al.^{30,31} proposed the frequency and spatial domain based on Graph Convolutional Networks (GCN), which for the first time simply applies the convolution operation in image processing to graph structured data processing. In 2018 Petar et al.^{32,33} proposed Graph Attention Networks (GAT), which employs an attention mechanism to compute the similarity between nodes as a way of learning each node's different levels of importance. In 2019 Kim et al.³⁴ proposed an edge-labeling graph neural network (EGNN), which uses intra-cluster similarity and inter-cluster dissimilarity to iteratively update the edge labels, and a deep neural network was used to implement a small-sample learning task on an edge-labeled graph. In 2021 Hu et al.³⁵ proposed the Hierarchical Graph Convolutional Neural Networks (HGCN) which uses hierarchical graph convolution to improve the accuracy of node classification and supports semi-supervised learning.

Attention mechanism

The attention mechanisms attempt to mimic the human brain and vision to process data. During visual observation, human beings consider the key part of the visual field space as the focused part and on the contrary the surroundings as the non-focused part. In other words, the attention mechanisms give higher weights to the relevant parts while minimizing the irrelevant parts by giving them lower weights. This allows the brain to

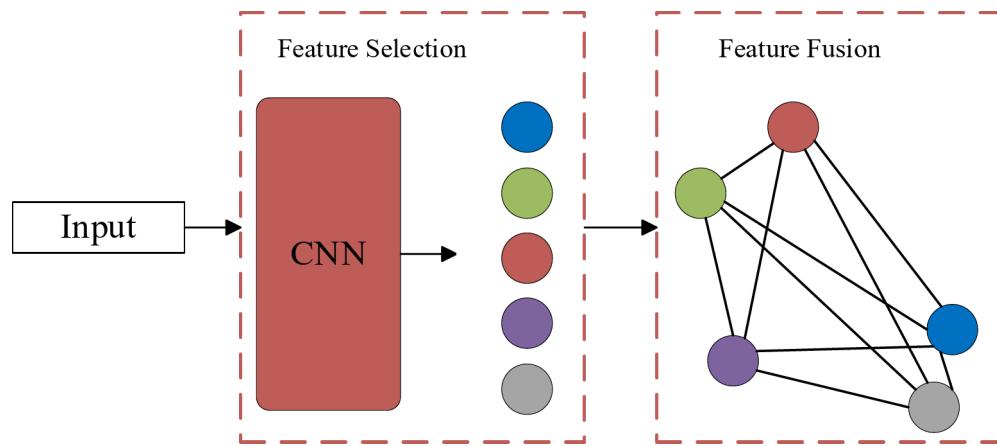


Fig. 1. Structure of the graph neural network.

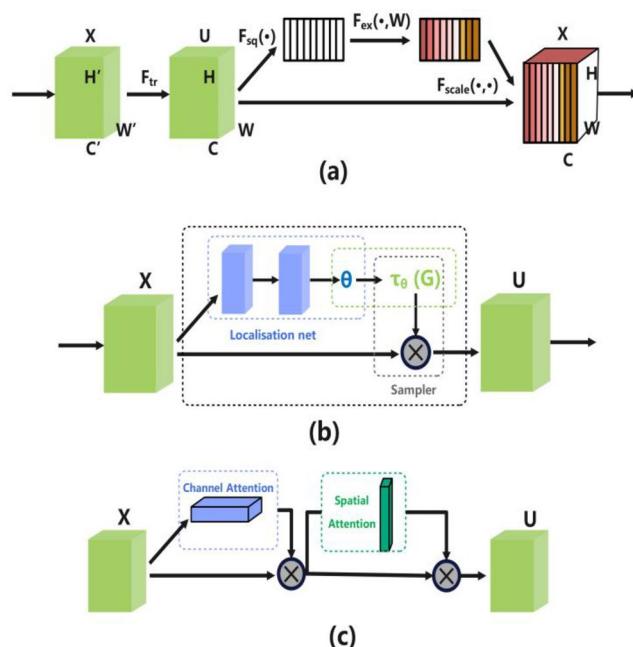


Fig. 2. Structure of the main representative attention mechanism models (SENet (a), STN (b), CBAM (c)).

accurately and efficiently process and attend to the most important parts, rather than processing the entire view space. The attention mechanisms were introduced into computer vision with the aim of mimicking the human visual system. This attention mechanism can be viewed as a dynamic weight adjustment process based on input image features. By introducing the attention mechanism, neural networks are able to automatically learn and selectively focus on the important information in the input, improving the performance and generalization of the model. The attention mechanism has been highly successful in many visual tasks, including image classification^{10,36}, target detection^{37,38}, semantic segmentation²⁴, image generation³⁹, and multimodal tasks^{34,40}.

We classify the attention mechanisms into six categories, including channel attention mechanisms, generating attention masks based on different channels and using it to select important channels, the representative network is the SENet model³⁷; spatial attention mechanisms, generating spatially attentional masks and using it to select important spatial regions, the typical spatial attention mechanism is the STN model³⁶; channel and spatial attention mechanisms, predicting the channel and spatial attention masks and selecting important features respectively, the representative model is CBAM⁴¹; In addition to this, there are temporal attention mechanisms, branching attention mechanisms and spatio-temporal attention mechanisms. The main representative models are shown in Fig. 2. Including SENet (a), STN (b), and CBAM (c), they all play unique roles in the classification of Few-shot images. Figure 2 (a): The SENet model enhances important features by introducing channel attention mechanism. Specifically, it compresses the global information of the feature map, then adjusts the importance of each channel through learned weights, and restores channel features through weighted combinations.

The STN model in Fig. 2 (b) achieves spatial transformation of features through local perception capability. It uses a localization network to generate transformation parameters, and then transforms input features to meet the requirements of the target task. The CBAM model in Fig. 2 (c) combines channel attention and spatial attention mechanisms, gradually focusing on the channel and spatial dimensions of the feature map, thereby effectively extracting global information and local features.

In order to improve the performance of the network for Few-shot image classification, the CNN with the attention mechanism were parallel fused, based on the local sensing of the CNN and the global information extraction characteristics of the attention mechanism to generate multi-scale feature information, and finally used the EGNN graph network module to realize the learning and updating of the generated nodes.

Methods

In early research on edge labeled graph neural networks (EGNN), the application of graph neural networks had not yet reached a mature stage due to a lack of in-depth understanding of the combination of parallel CNNs and attention mechanisms, as well as limitations in optimizing computational resources to improve model efficiency⁴⁰. With the continuous deepening of research and the advancement of technology, researchers are gradually able to better understand the synergistic effects between these mechanisms and redefine how attention mechanisms are applied to graph neural networks to achieve multi-dimensional feature aggregation at the node and edge levels⁴². Therefore, although EGNN has shown excellent performance in certain specific tasks, the complexity and uncertainty of combining it with CNN and attention mechanisms have led to less attention given to this ensemble approach in previous research. By designing a new framework based on CA-EGNN, this paper breaks through these limitations and fully demonstrates the adaptability and innovation capabilities in the development of this field.

The new graph neural network model framework (CA-EGNN) for solving Few-shot classification task through parallel CNN with attention mechanism was shown in Fig. 3. The inputs in the figure are two parts, the support set and the query set, and different colors indicate different sample categories. After the learning and updating of the EGNN graph network module for the node similarity, a series of changes occur in the distances between nodes of different colors. The network as a whole consists of two main parts: the parallel feature extraction module based on CNN and attention (CA-MEF) and the graph neural network module based on edge labeling (EGNN). The CA-MEF module consists of two parallel branches, i.e., the CNN branch and the attention branch, in order to realize that the two branches extract local and global features of the inputs and

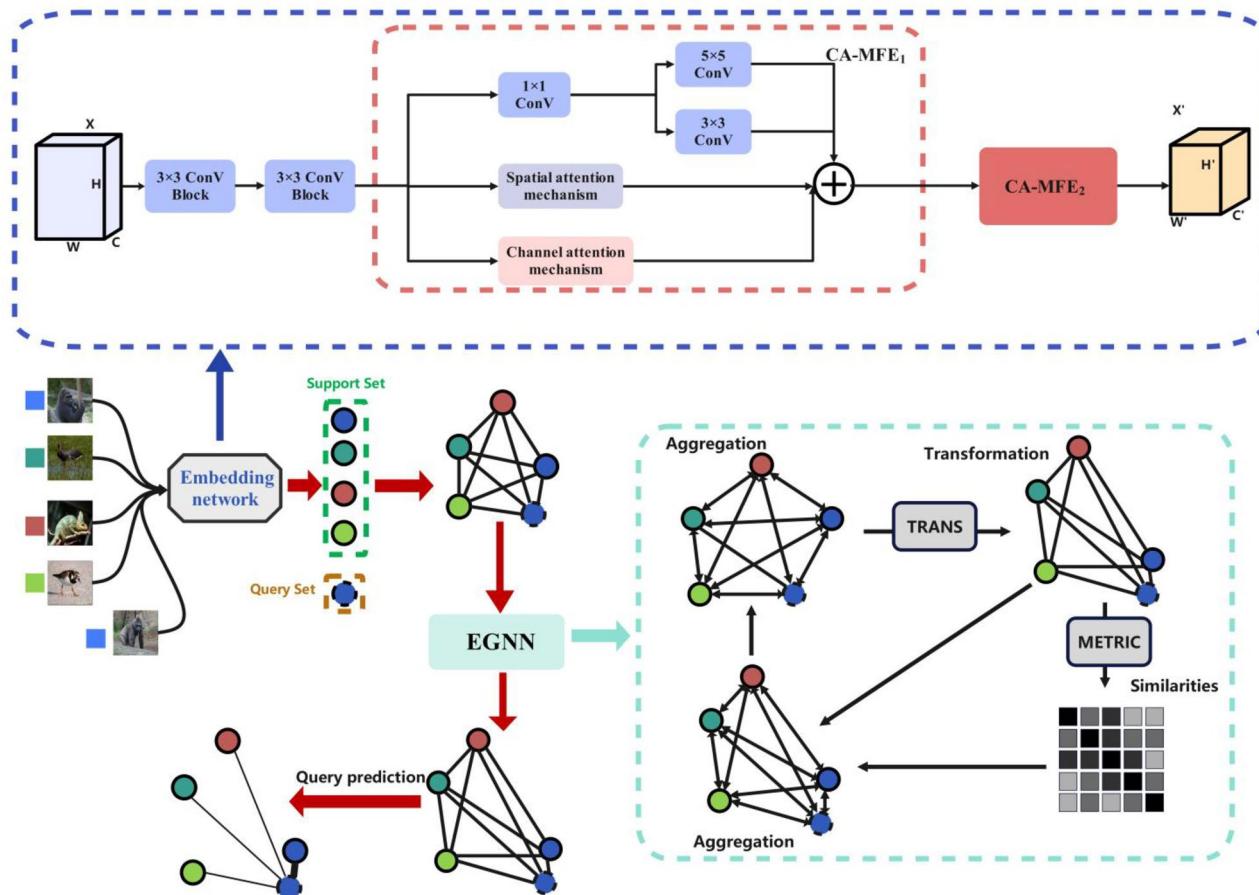


Fig. 3. Overall structure of CA-EGNN.

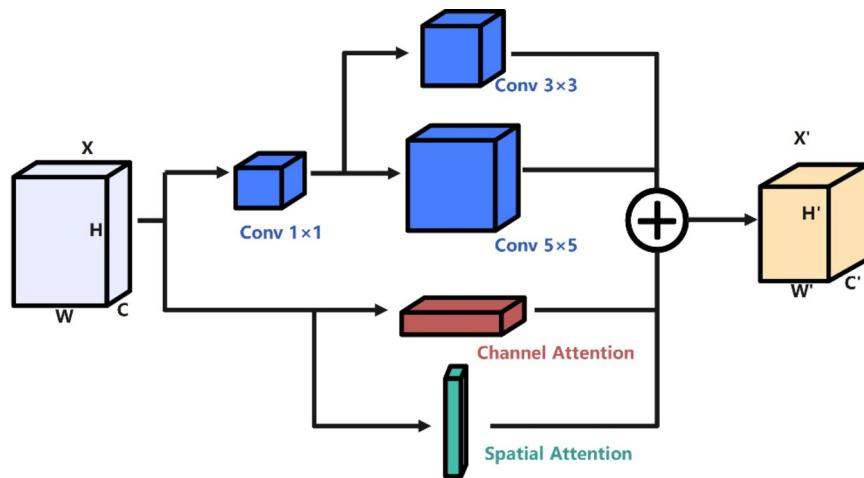


Fig. 4. Structure of CA-MFE mode.

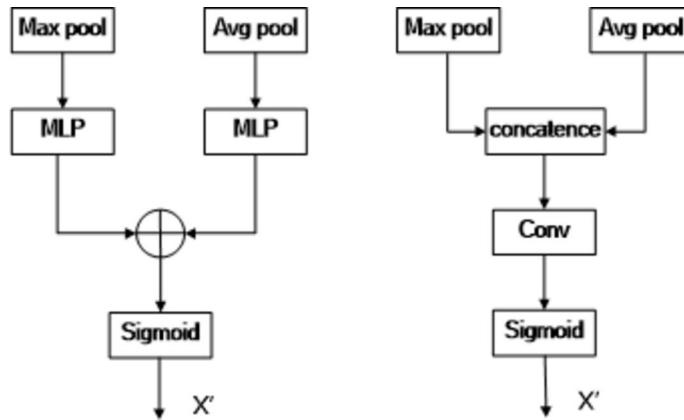


Fig. 5. Channel attention module (left) and spatial attention module (right).

merge them efficiently, and then extract the deeper information as node features; EGNN achieves similarity learning among nodes by alternately updating feature information of nodes and edges, and finally outputs the classification task of images.

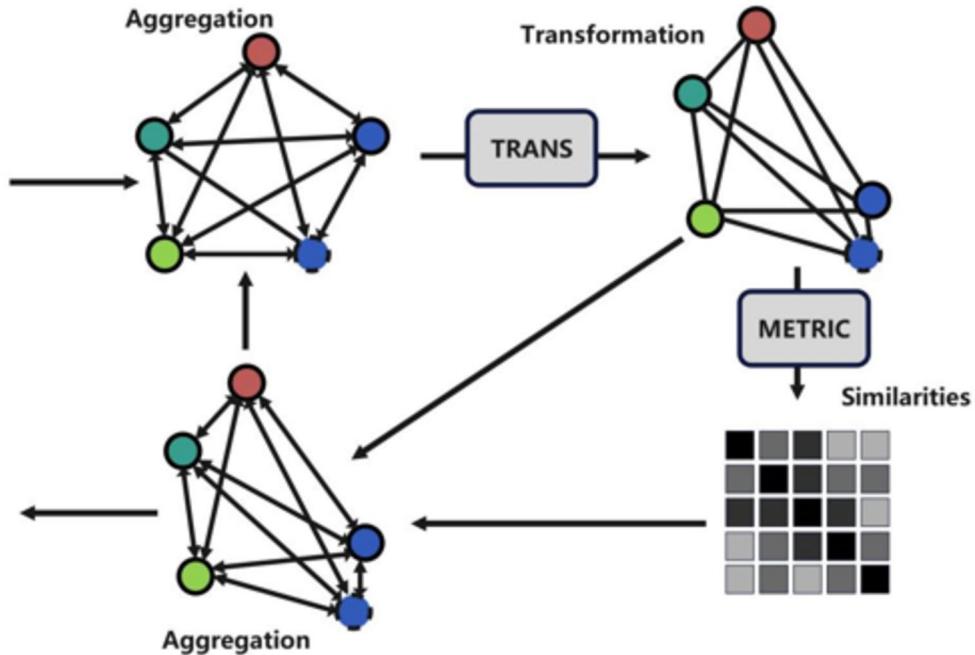
The CA-MFE algorithm

Detailed detail information enhances feature representation and establishes a good feature base for feature mapping in advanced spatial regions. A multi-scale feature extraction module (CA-MFE) was used to obtain deep feature details. As shown in Fig. 4, the module consists of four different branches. 1×1 convolution can be regarded as a simple transformation for dimensionality reduction, and the serial 3×3 convolution and 5×5 convolution after the dimensionality reduction made the feature extraction to different receptive domains to obtain a larger range of local details.

Another branch is the channel attention module^{24–37,[39,41,43,44,45]} and the spatial attention module⁴², which extract the global feature information of the image from both channel and spatial aspects, respectively. Their structure is shown in Fig. 5. In the channel attention module, the input feature map X is first subjected to global pooling and average pooling in the spatial dimension, which was used to compress the spatial dimensions and facilitate the network to learn the channel features, then the obtained pooling results are fed into the MLP to learn the features of the channel dimensions as well as the importance of each channel, and finally, the outputs are summed up and processed by mapping of the Sigmoid function to obtain the final result X' . In the spatial attention module, the input feature map X is first global pooling and average pooling in the channel dimension to compress the channel size for subsequent spatial feature learning, then the pooled results are spliced according to the channel, and finally the spliced results are convolved and processed by the activation function to obtain the result X'' .

In the overall CA-MFE module, firstly given the features of $X \in W \times H \times C$, X is input into 4 branches at the same time to generate 4 new feature maps, which have different sensory domains, and then the summing operation is performed on the 4 feature maps to obtain the new feature $X' \in W' \times H' \times C'$. The network

Output size	EGNN	CA-EGNN
41×41	$\left[\begin{array}{c} \text{Conv, } 3 \times 3, 64, \text{Padding 1} \\ \text{Max pool, } 2 \times 2 \end{array} \right]$	$\left[\begin{array}{c} \text{Conv, } 3 \times 3, 64, \text{Padding 1} \\ \text{Max pool, } 2 \times 2 \end{array} \right]$
20×20	$\left[\begin{array}{c} \text{Conv, } 3 \times 3, 96, \text{Padding 1} \\ \text{Max pool, } 2 \times 2 \end{array} \right]$	$\left[\begin{array}{c} \text{Conv, } 3 \times 3, 96, \text{Padding 1} \\ \text{Max pool, } 2 \times 2 \end{array} \right]$
10×10	$\left[\begin{array}{c} \text{Conv, } 3 \times 3, 128, \text{Padding 1} \\ \text{Max pool, } 2 \times 2 \end{array} \right]$	$\left[\begin{array}{c} \textbf{CA} - \textbf{MFE}, 128 \\ \text{Max pool, } 2 \times 2 \end{array} \right]$
5×5	$\left[\begin{array}{c} \text{Conv, } 3 \times 3, 256, \text{Padding 1} \\ \text{Max pool, } 2 \times 2 \end{array} \right]$	$\left[\begin{array}{c} \textbf{CA} - \textbf{MFE}, 256 \\ \text{Max pool, } 2 \times 2 \end{array} \right]$
1×1	FC,128	FC,128

Table 1. Architecture of EGNN and CA-EGNN feature extraction modules. Significant values are in bold.**Fig. 6.** The EGNN model framework.

architecture of the CA-MFE module is shown in Table 1, and the mathematical expression of the module can be expressed as Eq. (1):

$$X' = [H_{1 \times 1}(H_{4 \times 4}(X), H_{5 \times 5}(X)), C(X), S(X)], \quad (1)$$

where [...] denotes the connection, $H_{n \times n}(\cdot)$ denotes the $n \times n$ convolution operation, $C(\cdot)$ denotes the channel attention mechanism, and $S(\cdot)$ denotes the spatial attention mechanism.

The edge labeling based graph neural network

The edge labeling graph neural network module (EGNN)¹⁵ exploits intra-cluster similarity and inter-cluster dissimilarity to achieve evolution of explicit clustering by iteratively updating the edge labels on the graph for better node similarity comparisons. The EGNN model framework is shown in Fig. 6.

The feature extraction network generates the corresponding node feature information for the input image and constructs a fully connected graph $G = (V, E; T)$, where V and E are the set of nodes and edges of the graph, and T is the total number of samples. G serves as the input to the graph neural network module, and in the node updating process, the features v_i^{l-1} and e_{ij}^{l-1} from nodes and edges of the layer $l-1$ are given, and the node features are updated by the Neighbor aggregation of other nodes (similar and dissimilar neighbors) and edge features to perform the update of node features v_i^l . The attention mechanism is shown in Eq. (2):

$$v_i^l = f_v^l \left(\left[\sum_j \tilde{e}_{ij1}^{l-1} v_i^{l-1} \mid \sum_j \tilde{e}_{ij2}^{l-1} v_i^{l-1} \right]; \theta_v^l \right). \quad (2)$$

Where f_v^l denotes the node feature transformation network, which consists of aggregation layer, connection layer and MLP layer with the parameter θ_v^l .

According to the newly updated node features then the edge features are updated by regaining the similarity and dissimilarity between each pair of nodes, combining the previous edge's feature values and the updated similarity and dissimilarity to update the features of each edge, which is expressed by the following formula:

$$\bar{e}_{ij1}^i = \frac{f_e^l(v_i^l, v_j^l; \theta_e^l) e_{ij1}^{l-1}}{\sum_k f_e^l(v_i^l, v_k^l; \theta_e^l) e_{ik1}^{l-1} / (\sum_k e_{ik1}^{l-1})}, \quad (3)$$

$$\bar{e}_{ij2}^i = \frac{(1 - f_e^l(v_i^l, v_j^l; \theta_e^l)) e_{ij2}^{l-1}}{\sum_k (1 - f_e^l(v_i^l, v_k^l; \theta_e^l)) e_{ik2}^{l-1} / (\sum_k e_{ik2}^{l-1})}, \quad (4)$$

Where f_e^l denotes the metric network for calculating similarity with the parameter set to θ_e^l , which consists of four 1×1 convolutional modules, a linear layer and an activation function.

Finally, the final obtained edge labeling information is used to classify and predict the query set nodes by Softmax classifier. Algorithm 1 demonstrates the overall algorithm process of CA-EGNN.

Input: $G = (N, E; T), T = S \cup Q, S = \{(x_i, y_i)\}_{i=1}^{N \times K}, Q = \{x_i\}_{i=N \times K+1}^{N \times K+T}$
Output: $\{P(x_i)\}_{i=N \times K+1}^{N \times K+T}$

1 : Process samples through the CA-MFE feature extraction module to obtain node features n_i^0 and edge label features of the graph structure $e_{ij}^0, i, j = 1, 2, 3 \dots T$, and construct graph network

2 : for $l = 1, \dots, L$ do

/* Node feature update */

3 : for $i = 1, \dots, N$ do

4 : $n_i^l = f_n^l([\sum_j \tilde{e}_{ij1}^{l-1} n_j^{l-1} || \sum_j \tilde{e}_{ij2}^{l-1} n_j^{l-1}]; \theta_n^l)$

5 : end for

/* Edge feature update */

6 : for $i, j = 1, \dots, E$ do

7 : Calculate the updated edge labels e_{ij}^l according to equations 3-4

8 : end for

9 : end for

10 : Using Softmax classifier to obtain query set prediction results $\{P(x_i)\}_{i=N \times K+1}^{N \times K+T}$

Algorithm 1. The CA-EGNN.

Result and discussion

Problem definition

Few-shot classification is the process of learning a classifier with only a small number of training samples for each data category. Where each Few-shot classification task T contains two parts, the support set S (a set of input samples that have labels, similar to a training set) and the query set Q (a set of unlabeled samples that serves to evaluate the learned classifier, similar to a test set). When the support set S contains N classes of samples with K labeled samples in each class, the problem is called an N -way K -shot classification problem.

During the training and testing process, Few-shot classification tasks generally use scene training (episode training) mechanism. That is, N categories are randomly selected from each set, K labeled samples are extracted from each category to form the support set S , and the remaining small number of samples from each category form the query set Q . Then, the model is trained on the support set S , and the prediction loss of the query set Q is optimized through parameter updates until the results converge. This is the training process of one set. Figure 7 shows the N -way K -shot Few-shot learning image task, where green represents the support set and purple represents the query set.

Experimental data and setup

In order to verify the validity and applicability of the proposed model, we conduct validation experiments using two public datasets, mini-ImageNet, and tiered-ImageNet, as shown in Table 2. mini-ImageNet¹¹ was a small-sample dataset proposed by the authors in 2016 in the context of the Matching networks network model, and was extracted from the original of the ILSVRC-2012 dataset. All are RGB color images with a size of 84×84 pixels,

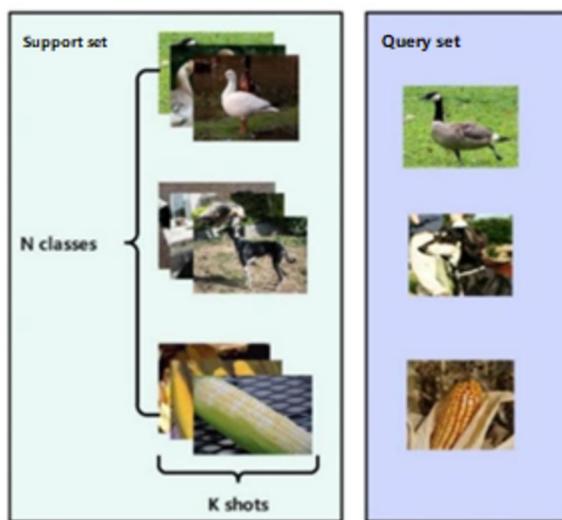


Fig. 7. Illustration of the few-shot learning image classification task.

Datasets		Training set	Validation set	Test set	Total
Mini-ImageNet	Classes	64	16	20	100
	Image	38,400	9600	12,000	60,000
Tiered-ImageNet	Categories	20	6	8	34
	Classes	351	97	160	608
	Image	448,695	124,261	206,209	779,165

Table 2. Experimental assignments of mini-ImageNet, tiered-ImageNet dataset categories and image.

and 100 different categories of images are extracted with 600 samples per category. The set used for training, validation and testing according to the ratio of 64:16:20, where the three sample sets were not duplicated with each other. tiered-ImageNet¹² on the other hand was proposed in 2018 in the context of meta-learning networks for semi-supervised learning, and similarly to the mini-ImageNet dataset, it was also extracted as a subset from the ILSVRC-2012. It extracts 34 hyper-categories (Categories), and each hyper-category contains 10~30 subcategories (Classes) ranging from 10 to 30, totaling 608 subcategories. Each subcategory has a varying number of images, totaling 779,165 images. Unlike the mini-ImageNet dataset, the tiered-ImageNet dataset is much larger in size, and the dataset division takes into account the category hierarchy of ImageNet by dividing the extracted super-classes into 20 super-classes (351 sub-classes) as the training set, 6 super-classes (97 sub-classes) as the validation set, and 8 super-classes (160 sub-classes) as the test set, and the average number of images in each class is 1281, while ensuring that there is no crossover between the three datasets.

For the above two datasets, we had performed a 5-way 5-shot standard Few-shot learning setup. The experiment was trained using the Adam optimizer, with the initial learning rate set to 5×10^{-4} and the weight decay factor to 10~6. The small batch size of the task for meta-training in the experiment was set to 40. For the mini-ImageNet dataset, the learning rate will decay by 0.1 after every 15,000 iterations, while the learning rate will decay by 0.1 after every 30,000 iterations because the tiered-ImageNet is a larger dataset than the mini-ImageNet, which requires more iterations to converge.

Result and analysis of CA-MFE

The proposed CA-EGNN network model for the Few-shot image classification task, we had compared its classification results would be compared with several most commonly used Few-shot image classification models on mini-ImageNet and tiered-ImageNet datasets from two aspects.

Different transduction settings

Depending on the transductive setup¹³, the models are classified into the following three types: non-Transductive means that each query sample prediction was performed separately; Transductive means that all query samples were processed and predicted simultaneously in a single graph; and BatchNorm can be thought of as transformational inference at test time, using query batch statistics instead of global batch normalization parameters.

The experimental task of 5-way 5-shot is carried out by the above three ways, and the experimental results are shown in Tables 3 and 4, which show that the CA-EGNN network improves the classification accuracy by 1.02% over the EGNN under transductive inference on mini-ImageNet dataset, and improves the accuracy by

Model	Trans.	5-way 5-shot
Matching networks	NO	55.30
Reptile	NO	62.74
Prototypical net	NO	65.77
GNN	NO	66.41
EGNN	NO	66.85
CA-EGNN	NO	68.02
MAML	BN	63.11
Reptile	BN	65.99
Relation Net	BN	67.07
MAML	YES	66.19
TPN	YES	69.43
TPN(Higher K)	YES	69.86
EGNN	YES	76.37
CA-EGNN	YES	77.39

Table 3. Classification accuracy of mini-ImageNet dataset under different transduction settings. Significant values are in bold.

Model	Trans.	5-way 5-shot
Reptile	NO	66.47
Prototypical net	NO	69.67
MAML	NO	69.89
EGNN	NO	70.98
CA-EGNN	NO	72.31
MAML	BN	70.30
Reptile	BN	71.03
Relation Net	BN	71.31
MAML	YES	70.83
TPN	YES	72.58
EGNN	YES	80.15
CA-EGNN	YES	80.61

Table 4. Classification accuracy of tiered-ImageNet dataset under different transduction settings. Significant values are in bold.

5.28%, 1.61%, and 1.17% over the Reptile, GNN, and EGNN models in non-transductive inference. In the tiered-ImageNet dataset, the classification accuracy under transductive inference is improved by 0.46% over EGNN, and the accuracy in non-transductive inference is improved by 5.85%, 2.42%, and 1.33% over Reptile, MAML, and EGNN models. The experimental results show that the new CA-EGNN network exhibits better classification results relative to other Few-shot classification networks in both mini-ImageNet and tiered-ImageNet dataset classification tasks, and its accuracy has achieved a good improvement.

Different modules for feature extraction

The CA-MFE improves the feature extraction module for the original four-layer convolutional neural network composition of graph neural networks, so we implement the network in various types of Few-shot image classification networks with convolutional neural network as the base network on mini-ImageNet and tiered-ImageNet datasets with 5-way 1-shot and 5-way 5-shot classification tasks, and their experimental results are shown in Tables 5 and 6. CA-EGNN demonstrates better classification accuracy regardless of the classification task and the dataset. Among them, it improves 11.41%, 7.42%, and 5.38% in the 5-way 5-shot task over GNN, TPN, and Dynamic, respectively. 8%, 2.8%, and 2.21% in the 5-way 1-shot task over GNN, TPN, and Dynamic, respectively. In contrast, CA-EGNN has more obvious accuracy improvement under the 5-way 5-shot task. In the case of a large number of sample images in the input query set, CA-MFE is able to extract richer and more critical feature information, and efficiently complete the learning of graph similarity measure in the graph network module, resulting in a more obvious improvement in the final classification accuracy compared to the 5-way 1-shot.

Semi-supervised few-shot learning results

In real life, a large amount of data is often unlabeled and only a small portion of data is labeled. Semi-supervised learning is to combine labeled and unlabeled data to make full use of the information in the unlabeled data to

Model	Backbone network	5-way 1-shot	5-way 5-shot
Matching net	Conv	43.67	54.71
Relation net	Conv	50.21	64.65
MAML	Conv	49.56	56.17
Proto net	Conv	49.26	64.78
Dynamic	Conv	56.13	72.01
GNN	Conv	50.34	65.98
TPN	Conv	55.54	69.97
EGNN	Conv	–	75.78
CA-EGNN	CA-PFE	58.34	77.39

Table 5. Classification accuracy of mini-ImageNet dataset for different classification tasks. Significant values are in bold.

Model	Backbone network	5-way 1-shot	5-way 5-shot
Relation net	Conv	54.31	71.22
MAML	Conv	51.83	70.83
Proto net	Conv	53.48	72.64
TPN	Conv	59.61	72.58
EGNN	Conv	–	80.15
CA-EGNN	CA-PFE	61.11	80.61

Table 6. Classification accuracy of tiered-ImageNet dataset for different classification tasks. Significant values are in bold.

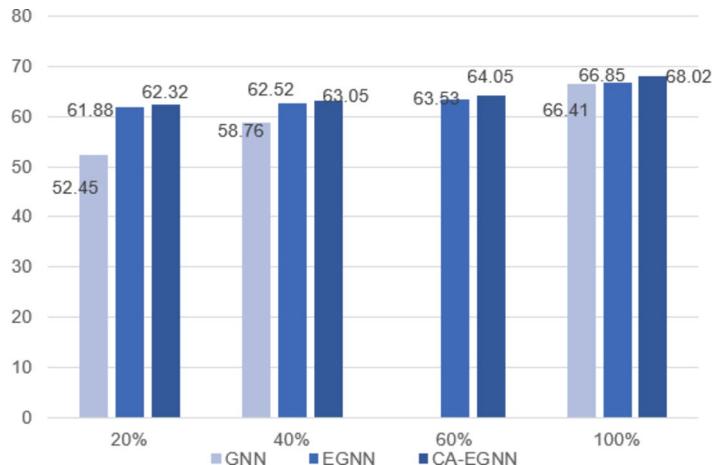


Table 7. Classification accuracy of mini-ImageNet dataset with different proportions of labeled samples ablation experiment.

improve the generalization ability of the model. In the semi-supervised setup, the project team uses a 5-way 5-shot iterative setup for training experiments on the mini-ImageNet dataset with 20%, 40%, and 60% proportion of labeled data, and the labeled samples are balanced across categories, with the same number of labeled and unlabeled samples across all categories. 20% labeled setup has 1 labeled sample and 1 unlabeled sample for each category. samples contain 1 input with labeled samples and 4 inputs with unlabeled samples; 40% labeled setup contains 2 inputs with labeled samples and 3 inputs with unlabeled samples per category; 60% labeled setup contains 3 inputs with labeled samples and 2 inputs with unlabeled samples per category; the specific experimental results are shown in Table 7. In the case of sparse labeled data, CA-EGNN still demonstrates good classification results compared to the benchmark model.

We had conducted two sets of ablation experiments on the mini-ImageNet dataset to validate the effectiveness of the CA-EGNN model by changing the Attention Mechanism module and adjusting the network layer where the CA-MFE module is located to conduct Few-shot image classification experiments.

Experiment	Model			
	SAM	CAM	Serial S&C	Parallel S&C
5-way 5-shot	66.51	66.86	67.15	68.02

Table 8. Classification results under different combinations of attention mechanisms. Significant values are in bold.

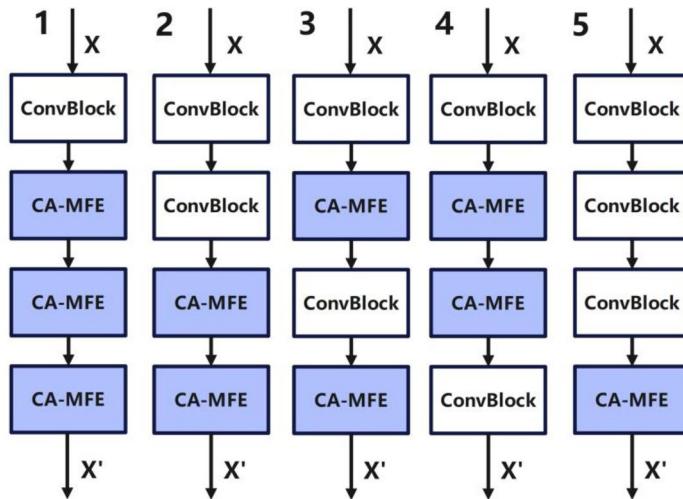


Fig. 8. Model of different network layers where CA-MFE is located.

Experiment	Model				
	1	2	3	4	5
5-way 5-shot	66.95	68.02	66.28	67.02	66.55

Table 9. Classification results for different layers of CA-MFE. Significant values are in bold.

Adjustment of attention mechanisms

The new CA-MFE model uses channel and spatial attention mechanisms in parallel to realize the extraction of global feature information of Few-shot images, however, the composition of different forms of attention mechanisms will affect the efficiency of image extraction, in order to verify which combination of attention mechanisms can make the Few-shot classification better, we compared four structures of Spatial Attention Mechanism Only (SAM), Channel Attention Mechanism Only (CAM), spatial attention mechanism and channel attention mechanism serial (Serial S&C), and spatial attention mechanism and channel attention mechanism parallel (Parallel S&C) in the EGNN base network, and the results are shown in Table 8.

The experimental results show that the combination of parallel attention mechanisms in the 5-way 5-shot task improves the accuracy by 1.51% and 1.16% over the individual attention mechanisms, respectively, and also improves the accuracy by 0.87% compared to the serial attention structure. It can be seen that the parallel module of spatial and channel attention mechanisms can better capture the global information in the image as a way to improve the overall classification accuracy.

Adjustment of the CA-MFE network layer

The original feature extraction network of EGNN consists of four convolutional modules, and the project team replaces the original convolutional modules with CA-MFE modules, replacing the convolutional modules with different layers will affect the final classification results of the model, in order to verify which module replacement can better improve the classification accuracy of Few-shot, the project team carries out experiments on the replacement of CA-MFE modules according to the CA-MFE modules as shown in Fig. 8, with the Few-shot classification task with the EGNN network, and the final experimental results are shown in Table 9.

The experimental results show that the different number of CA-MFE modules and the number of layers in which they are located affect the final classification effect, in which the best classification effect is achieved when replacing the last two layers of convolutional blocks in the original model, and the classification accuracy is 2% higher than the average value of the other models.

CA-EGNN shows better classification accuracy than other models on mini-ImageNet and tiered-ImageNet datasets. This indicates that the performance of the new model has been improved, which may lead to better classification results in practical applications. The ablation experiment shows that introducing attention mechanism and parallel processing of CNN can significantly improve the classification performance of the model. This indicates that the structural design of the new model is reasonable and may be more practical in practical applications. Due to its excellent performance in Few-shot classification tasks, the CA-EGNN model can be applied in many fields such as medical image diagnosis, satellite image classification, and military target recognition. These fields often face the problem of scarce labeled data, so the application of new models will have higher practical value. The experimental results show that this method has significant advantages in processing Few-shot classification data, ultimately promoting the development of Few-shot learning. The research results of this paper provide new ideas and methods for Few-shot learning. Helps promote further development in this field.

Conclusions

To address the problem of Few-shot image classification in graph neural networks, where feature extraction has limitations in capturing global features, we propose a parallel multi-branch feature extraction method (CA-MFE algorithm) based on CNN and attention mechanism. By introducing multi-scale convolution kernels and parallel channel space attention mechanisms, the CA-MFE model can extract multidimensional features from both local and global perspectives, significantly improving classification performance. This design utilizes CNN to obtain multi-scale local features through different convolution kernels, as well as global features extracted through spatial and channel attention mechanisms, and combines the above two parts in parallel to achieve rich and critical feature information acquisition. Experiments on mini-ImageNet and hierarchical ImageNet datasets show that compared to traditional base models, the CA-MFE model improves classification accuracy by 1.07% and 1.33%, respectively. In addition, in the 5-way 5-shot task, CA-MFE showed excellent performance under different inference settings, especially in the case of a large number of query samples, where CA-MFE significantly improved classification accuracy. This proves that the model can effectively combine the advantages of CNN and attention mechanism in Few-shot image classification, significantly improving classification accuracy.

Implement the CA-MFE algorithm for Few-shot image classification by adding multiple feature extraction models, integrating attention mechanism into CNN. Although it improves classification accuracy, it inevitably has increased the number of model parameters, and the computational complexity of the algorithm to some extent. Therefore, in order to reduce the number of model parameters, improve the learning rate, and shorten the learning process of the model, the research team is preparing to use lightweight model frameworks, model pruning, and other methods to make the model easier to use in practical applications. Meanwhile, the combination of lightweight network technology with more Few-shot learning methods such as MR-DCAE, MobileViT, and MobileRaT may provide new directions for improving model performance.

In addition, we will consider integrating meta-learning mechanisms to further improve the generalization ability in Few-shot learning. Future work will focus on optimizing the structure of feature extraction networks to reduce computational complexity and improve classification efficiency, such as introducing lightweight designs (e.g. separable convolutions). At the same time, multimodal information fusion will be a key direction for improving model performance, exploring how to further improve the robustness and adaptability of the model in complex environments is also an important topic for future research. Finally, enhancing the interpretability of the model will make it more transparent and noise resistant in practical applications. These optimizations the structure of the feature extraction network will further enhance the value and efficiency of the CA-MFE algorithm in Few-shot image classification in future.

Data availability

The data that support the findings of this study are openly available in the papers Vinyals, O., Blundell, C., et al.³⁸ "Matching Networks for One Shot Learning" and Triantafillou, E. et al.⁴⁶ "Meta-Learning for Semi-Supervised Few-Shot Classification".

Received: 16 January 2024; Accepted: 3 January 2025

Published online: 08 January 2025

References

1. Tian, Y., Zhao, X. & Huang, W. Meta-learning approaches for learning-to-learn in deep learning: a survey. *Neurocomputing* **494**, 203–223 (2022).
2. Hospedales, T. M., Antoniou, A., Micaili, P. & Storkey, A. J. Meta-learning in neural networks: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 5149–5169 (2021).
3. Wang, Y., Yao, Q., Kwok, J. T. & Ni, L. M. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv.* **53**, 1–34 (2020).
4. Miller, E. G., Matsakis, N. E. & Viola, P. A. Learning from one example through shared densities on transforms. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **1**, 464–471 (2002).
5. Koch, G., Zemel, R. & Salakhutdinov, R. Siamese neural networks for one-shot image recognition. *ICML Deep Learn. Workshop* **2**.
6. Tanner, M. & Wong, W. The calculation of posterior distributions by data augmentation. *JASA J. Am. Stat. Assoc.* **82**, 528–540 (1987).
7. Xing, E. P., Ng, A. Y., Jordan, M. I. & Russell, S. J. Distance metric learning with application to clustering with side-information. In *International Conference on Neural Information Processing Systems*, 521–528 (2002).
8. Vilalta, R. & Drissi, Y. A perspective view and survey of meta-learning. *Artif. Intell. Rev.* **18**, 77–95 (2002).
9. Garcia, V. & Bruna, J. Few-shot learning with graph neural networks. *Int. Conf. Learn. Represent.* <https://doi.org/10.48550/arXiv.1711.04043> (2017).

10. Zheng, Q. et al. MR-DCAE: Manifold regularization-based deep convolutional autoencoder for unauthorized broadcasting identification. *Int. J. Intell. Syst.* <https://doi.org/10.1002/int.22586> (2021).
11. Zheng, Q. et al. A real-time constellation image classification method of wireless communication signals based on the lightweight network MobileViT. *Cogn. Neurodyn.* **18** (2), 659–671. <https://doi.org/10.1007/s11571-023-10015-7> (2024).
12. Zheng, Q. et al. MobileRaT: a lightweight radio transformer method for automatic modulation classification in drone communication systems. *Drones* **7** (10), 596 (2023).
13. Zheng, Q. et al. A real-time transformer discharge pattern recognition method based on CNN-LSTM driven by few-shot learning. *Electr. Power Syst. Res.* **219**, 109241 (2023).
14. Zheng, Q. et al. PAC-Bayesian framework based on drop-path method for 2D discriminative convolutional network pruning. *Multidimens. Syst. Signal Process.* **31** (3), 793–827 (2020).
15. Wang Haiquan, G. et al. Review of tool wear detection technology. *Automation Technology and Application* **43**(07), 1–6. [https://doi.org/10.20033/J.1003-7241.\(2024\)07-0001-06](https://doi.org/10.20033/J.1003-7241.(2024)07-0001-06) (2024).
16. KIM, J. et al. Edge-labeling graph neural network for few-shot learning. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11–20 (2019).
17. HU, F. et al. Hierarchical graph convolutional networks for semi-supervised node classification. In *Proc. of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19)*, 4532–4539 (2019).
18. Li, F. et al. A small sample of graph convolution combining residuals and self-attention mechanisms image classification network. *Comput. Sci.* **50** (S1), 376–380 (2023).
19. Zhang, Y. & Xun, C. Classification method of tea diseases based on two-node and two-sided graph neural network. *Trans. Chin. Soc. Agric. Mach.* **55** (03), 252–262 (2024).
20. Chen Yaoqi. Study of the cognitive learning model based on the concept of attention. Southwest university (2023).
21. Kim, Y. Convolutional neural networks for sentence classification. *Eprint Arxiv*. <https://doi.org/10.3115/v1/D14-1181> (2014).
22. Ma, W., Yang, Q., Wu, Y., Zhao, W. & Zhang, X. Double-branch multi-attention mechanism network for hyperspectral image classification. *Remote Sens.* **11**, 1307 (2019).
23. Wang, J. et al. DualSeg: Fusing transformer and CNN structure for image segmentation in complex vineyard environment. *Comput. Electron. Agric.* **206**, 107682 (2023).
24. Fu, J. et al. Dual attention network for scene segmentation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3146–3154. <https://doi.org/10.1109/cvpr.2019.00326> (2019).
25. Gao, Z., Li, Y. & Wang, S. An image recognition method using parallel deep CNN. *Acad. J. Comput. Inf. Sci.* **2**, 34–48 (2019).
26. Zhang, Y., Liu, H. & Hu, Q. TransFuse: fusing transformers and CNNs for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, Lecture Notes in Computer Science* 14–24. https://doi.org/10.1007/978-3-030-87193-2_2 (2021).
27. Li, R., Zheng, S., Duan, C., Yang, Y. & Wang, X. Classification of hyperspectral image based on double-branch dual-attention mechanism network. *Remote Sens.* **12**, 582 (2020).
28. Gori, M., Monfardini, G. & Scarselli, F. A new model for learning in graph domains. In *Proc. IEEE International Joint Conference on Neural Networks, 2005* **2**, 729–734 (2006).
29. Dai, J. et al. Deformable convolutional networks. In *IEEE International Conference on Computer Vision (ICCV)*, 764–773. <https://doi.org/10.1109/iccv.2017.89> (2017).
30. Wu, F. et al. Semi-supervised multi-view graph convolutional networks with application to webpage classification. *Inf. Sci.* **591**, 142–154 (2022).
31. He, P. et al. Single shot text detector with regional attention. In *IEEE International Conference on Computer Vision (ICCV)*, 3047–3055. <https://doi.org/10.1109/iccv.2017.331> (2017).
32. Veliković, P. et al. *Graph Attention Networks* <https://doi.org/10.48550/arXiv.1710.10903> (2017).
33. Kim, J., Kim, T., Kim, S. & Yoo, C. D. Edge-labeling graph neural network for few-shot learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11–20. <https://doi.org/10.1109/cvpr.2019.00010> (2019).
34. Su, W. et al. VL-BERT: pre-training of generic visual-linguistic representations. *Int. Conf. Learn. Represent.* <https://doi.org/10.48550/arXiv.1908.08530> (2020).
35. Hu, F., Zhu, Y., Wu, S., Wang, L. & Tan, T. Hierarchical graph convolutional networks for semi-supervised node classification. In *Proc. of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. <https://doi.org/10.24963/ijcai.2019/630> (2019).
36. Jaderberg, M., Simonyan, K., Zisserman, A. & Kavukcuoglu, K. Spatial transformer networks. *Neural Inf. Process. Syst.* **28** (2015).
37. Hu, J., Shen, L., Sun, G. & Albanie, S. Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* <https://doi.org/10.1109/TPAMI.2019.2913372> (2017).
38. Vinyals, O., Blundell, C., Lillicrap, T. P. & Kavukcuoglu, K. Wierstra, D. Matching networks for one shot learning. *ArXiv Learn.* **29**, 3630–3638 (2016).
39. Zhang, H., Goodfellow, I., Metaxas, D. N. & Odena, A. Self-attention generative adversarial networks. *Int. Conf. Mach. Learn.*, 7354–7363. <https://doi.org/10.48550/arXiv.1805.08318> (2019).
40. Xu, Z. Fundus image classification method based on graph neural network research and implementation. Ningxia University (2023).
41. Woo, S., Park, J., Lee, J. Y. & Kweon, I. S. CBAM: Convolutional block attention module. In *Computer Vision – ECCV 2018, Lecture Notes in Computer Science*, 3–19. https://doi.org/10.1007/978-3-03-01234-2_1 (2018).
42. Evaluation of training efficiency. and performance of graph neural network model based on distributed environment [J/OL]. *Comput. Appl.*, 1–10. <http://kns.cnki.net/kcms/detail/51.1307.TP.20240925.1715.010.html> (2024).
43. Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M. & Monfardini, G. The graph neural network model. *IEEE Trans. Neural Netw.* **20**, 61–80 (2009).
44. Carion, N. et al. End-to-end object detection with transformers. In *Computer Vision – ECCV 2020, Lecture Notes in Computer Science*, 213–229. https://doi.org/10.1007/978-3-03-58452-8_13 (2020).
45. Xu, T. et al. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1316–1324. <https://doi.org/10.1109/cvpr.2018.00143> (2018).
46. Triantafillou, E. et al. Meta-learning for semi-supervised few-shot classification. *Int. Conf. Learn. Represent.* <https://doi.org/10.48550/arXiv.1803.00676> (2018).

Acknowledgements

The authors extend their appreciation to the National Natural Science Foundation of China under Grant (31870532), and the Natural Science Foundation of Hunan Province under Grant funding, and Changsha Science and Technology Plan Project under Grant funding (kq2402265) in this research work through the project.

Author contributions

LIU.YM. planned supervision in the study, provided conceptualization and methodology, prepared funding acquisition and conducted interviews, and lead the revision. Xiao.FJ. was responsible for software design and operation, edited, conducted interviews and drafted the original manuscript. Zheng.XY. was responsible for

writing, revising papers, drawing charts, validation, and edited the manuscript, and supervised the manuscript. Deng WH. validated the data, review the manuscript, and put forward the main research problems of this paper, literature reviewed and project administration. Ma HZ. control the writing progress of the paper, write the original draft preparation and participate in the revision of the paper. Su XY, and Wu L. participated in the later revision work of the paper. All authors have read the paper and approved the submitted version, and improved the manuscript by responding to the review comments.

Funding

This work was supported by the National Natural Science Foundation of China under Grant (31870532), Changsha Science and Technology Plan Project (kq2402265). This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025