IECE
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2025.146072

RESEARCH ARTICLE

A Few-shot Learning Method Using Relation Graph

Zijing Liu[1,*] **and Chenggang Wang**[1]

[1] No.10th Research Institute, China Electronics Technology Group Corporation, Chengdu 610036, China

Abstract

Few-shot learning aims to recognize new-class items under the circumstances with a few labeled support samples. However, many methods may suffer from poor guidance of limited new-class samples that are not suitable for being regarded as class centers. Recent works use word embedding to enrich the new-class distribution message but only use simple mapping between visual and semantic features during training. To solve the aforementioned problems, we propose a method that constructs a class relation graph by semantic meaning as guidance for feature extraction and fusion, to help the learning of the second-order relation information, with a light training request. In addition, we introduce two ways to generate pseudo prototypes for augmentation to resolve the lack of representation due to limited samples in novel classes: 1) A Generation Module(GM) that trains a small structure to generate visual features by using word embedding; 2) A Relation Module(RM) for training-free scenario that uses class relations in semantics to generate visual features. Extensive experiments on benchmarks including *miniImageNet*, *CIFAR-FS* and *FC-100* prove that our method achieves state-of-the-art results.

Keywords: few-shot learning, relation graph.

1 Introduction

In recent years, we have witnessed the success of deep learning in the field of image recognition, but such a breakthrough depends on a large amount of labeled data costing significant human labor. In contrast, humans are capable to master a certain knowledge with a few samples. Few-shot learning aims to mimic such an ability to extract class information by only viewing limited samples, which helps to solve the problem of large data requirements.

Meta-learning [1, 2] is one of the most popular methods for few-shot learning. It aims to learn a model that can extract prototypes for novel classes from a few data, based on which the test data are able to be distinguished by their feature similarities. Vinyals et al. [3] introduces two structures for support and query set to diversely extract the feature and compare the compound score for classification. ProtoNet methods [4, 5] use light structures to extract features and propose to measure the distance between prototypes and features in a shared metric space. Meta-Baseline [6] takes advantage of the base training and episodic training by combining them into two stages. In a word, These methods provide several simple and effective baselines with

Academic Editor:
Xiaoling Wang

Submitted: 04 March 2025
Accepted: 23 March 2025
Published: 27 March 2025

Vol. 2, **No.** 1, 2025.
10.62762/CJIF.2025.146072

***Corresponding author:**
✉ Zijing Liu
853000906@qq.com

Citation
Liu, Z., & Wang, C. (2025). A Few-shot Learning Method Using Relation Graph. *Chinese Journal of Information Fusion*, 2(1), 70–78.

relatively less computational cost. Despite performing well, the proto-based method lacks the ability to extract high-quality prototypes only from visual domain, especially when the support sample is not representative. It is vital to enrich class information from other aspects.

To enhance the feature representation, semantic information is introduced during feature extraction [7, 8]. Xing et al. [9] adaptively combines the visual and semantic features with a learnable factor. KTN [12] aligns the fully connected node for base classes from both modals, and obtains new knowledge from the semantic modal for novel classes. Chen et al. [11] matches the deepest visual features with the word embedding and then uses the former to augment the feature with the nearest neighbor distance. CVAE [10] proposes a generative-adversarial structure to generate better auxiliary features. These methods introduce semantic information into FSL, but with simple mapping between visual and semantic features, without digging further messages. Such mapping may be helpful to some extent, but may not generate an ideal embedding space, leading to the limitation of generalization. Attribute methods [13, 14] go further into semantic information, they use component information of a class to divide visual features into several segments, each aligning with attribute embeddings. But they need extra attribute information labels in datasets for guidance, making the training cost higher. Compared with these methods, our proposed method takes class relations into consideration, which is easy to obtain by using word embedding.

In our opinion, it is more appropriate to learn the relationship among classes, rather than only direct features, since the semantic structure contains the core embedding information. For such purpose, we propose a novel FSL method that establishes a class-relation graph to guide the feature extraction, hoping the network can learn second-order relation messages. Specifically, we design an absolute constraint for direct feature mapping and a relation constraint for graph mapping among second-order relations. Furthermore, to improve feature representation with limited seen samples, we propose a Generative module and a Relation module to produce pseudo prototypes for modifying the support features in few-shot case. The former bridges word embedding and learned features to directly generate pseudo features for novel classes, while the other uses the word correlation between base and novel classes to incorporate the base prototypes for generation. The performance of our model shows state-of-the-art behavior in *miniImageNet* and *CIFAR-FS*, and achieving an improvement of 1.12%–3.8% in *FC-100* for 1-shot case compared with the second-best method. It also has a first-class performance for 5-shot case. Our main contribution can be summarized as follows:

1. Using word embedding, we propose an absolute constraint a relation constraint to produce visual features with semantic information.
2. We propose a Generative Module and a Relation Module to produce pseudo features that can refine prototypes to class centers.
3. Our experiments show that our methods achieve state-of-the-art performance on three challenging datasets, *miniImageNet*, *CIFAR-FS* and *FC-100*.

2 Method

2.1 Problem Setup

Following the typical few-shot learning setting, a dataset contains a base set D^b consisting of base classes C^b, and a novel set D^n with novel classes C^n, in which $C^b \cap C^n = \phi$. Firstly the goal is to train a feature extractor $f_\theta(\cdot)$ on base classes. Based on that, a set of few-shot classification tasks are constructed for novel class recognition. For a N-way K-shot task, N classes are randomly chosen from C^n, and a fixed quantity of samples are equally chosen among each class. After that, the support set $S = \{(x_i, y_i)\}_{i=0}^{N\times K}$ is constructed by choosing K labeled samples for each class, while the rest of the samples are used to construct query set $Q = \{(x_i, y_i)\}_{i=0}^{M}$, the aim of few-shot learning is to use information from S to guide the classification of Q.

2.2 Backbone Training

As the setting in [15], we use a two-branch structure to train the feature extractor: Apart from keeping the standard classification architecture, another branch with a new logistic regression structure is added after the penultimate layer, which retrieves possible rotation of each sample. And manifold-mixup [18] method is included for augmentation to improve the model robustness.

While the design above has good performance in regular classification, the embedding space it constructs lacks semantic relations among classes. For example, the class 'chair' and class 'desk' may have high semantic similarity in real world, but such

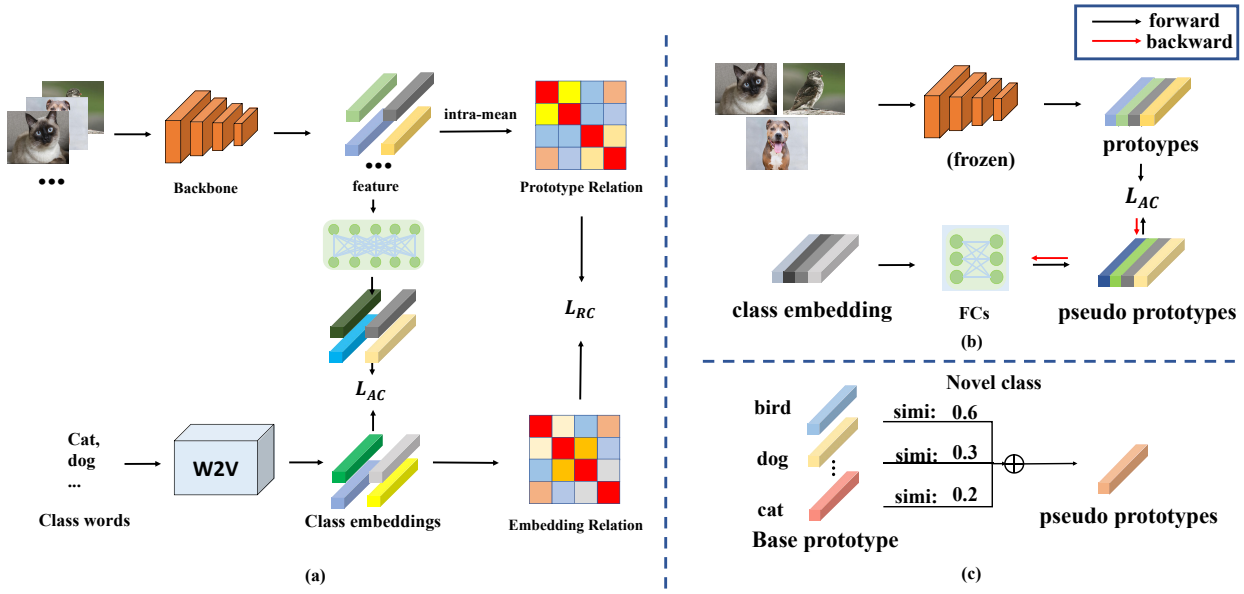


Figure 1. (a) Base training. the framework includes a classification branch and two extra branches for absolute and relative constraints. (b) Training of Generative Module(GM), for training-needing case. (c) Process of Relation Module(RM), for training-free case.

relation may differ in the embedding space. To solve the problem, we introduce using word embedding information to guide the training. The architecture is shown as Figure 1(a), in which the main structure consists of: 1) Absolute Constraint, which forces features to match the corresponding word embedding. 2) Relation Constraint, which forces relations between class prototypes to match the similarity between semantic meanings by building a relation graph.

2.2.1 *Absolute Constraint*

In this section, we hope that the extract features can directly learn the class representation ability of word embeddings, thus projecting features into word embeddings dimension with a simple 2-layer FC structure. Then we design a loss function. In details, given the features $F = f_\theta(x)$ extracted from the backbone in a training batch, we assign word embeddings of base classes loaded from a Natural Language Processing model to form a direct restriction, to imitate its distribution. The word embedding label of each feature is chosen according to its class label: $y_j^w = E_b[y_j]$, where y_j is the label of the j-th sample and E_b is the word embedding of all base classes. An embedding layer is then applied to match the channel dimension. The process can be represented as the following constraint:

$$\mathcal{L}_{AC}(F', y^w) = \frac{1}{N}\sum_{j=0}^{N}(W(F_j) - y_j^w)^2 \qquad (1)$$

where W stands for a fully-connected layer and N is the batch size. F_j is the j-th visual feature in a batch. $F' = W(F_j)$ stands for the projected features by fully-connected steucture

2.2.2 *Relation Constraint*

Although features extracted by backbone can represent class to some extent, the feature distribution, such as class relations, may not have the same structure as in real word. In this case, we need to use a good second-order structure to guide training, supposing that the extracted features not only have good ability of class representation. To achieve this, we use a class relation graph in each batch to guide the similarity of class prototypes. To use the relations among semantic information, we first calculate the prototypes of batch classes according to the labels:

$$\mu_i = \frac{\sum_{j=0}^{N} F_j * \mathbb{I}(y_j = i)}{N_i} \qquad (2)$$

where N_i is the number of samples that belongs to class i, $\mathbb{I}(\cdot)$ equals to 1 if the inner condition is true. Then we establish the similarity relations between class semantics by using cosine similarity, and so done as for prototypes to calculate visual embedding relations:

$$R_\mu(m,n) = cos(\mu_m, \mu_n) = \frac{\mu_m \cdot \mu_n}{||\mu_m||\ ||\mu_n||} \qquad (3)$$

$$R_w(m,n) = cos(w_m, w_n) = \frac{w_m \cdot w_n}{||w_m||\ ||w_n||} \qquad (4)$$

where m,n are class index, $\mu\_$ and $w\_$ are corresponding prototypes and word embeddings. After that we use

the following loss to keep the word relations among extracted visual prototypes:

$$\mathcal{L}_{RC} = \frac{1}{MN}\sum_{m=0}^{M}\sum_{n=0}^{N}(R_\mu(m,n) - R_w(m,n))^2 \quad (5)$$

Considering the classification loss, the overall objective for training the base model is as follows:

$$\mathcal{L} = \mathcal{L}_C + \beta\mathcal{L}_{AC} + \gamma\mathcal{L}_{RC} \quad (6)$$

where β and γ are hyperparamters to, and L_C stands for the cross-entropy loss for regular classification restriction.

2.3 Pseudo generation

In order to further utilize the knowledge in semantic information for classes that is helpful to modify the support feature with a few samples, we extracted features from all trained samples after the base training, and calculate their class prototypes similarly as in Eq(2). We then apply them together with their word embeddings to generate pseudo prototypes for novel classes. Here we introduce two ways to achieve it: 1) A Generative Module that trains a structure that uses novel class embeddings to generate its visual features. 2) A Relation-based module that compounds similar base class prototypes according to the semantic relations between novel and base classes. By fusing the pseudo and the origin prototypes, the representation in few-shot case can be improved.

Generative Module(GM). One way to generate features is directly training a structure that is able to derive visual features according to their word embeddings. We use the base class word embedding w_b as input for an FC layer, its output $\hat{\mu}$ is set to approximate the corresponding prototypes. The process is illustrated in Figure 1(b). Once the training is finished, we use this structure to extract novel class pseudo prototypes according to the novel word embeddings.

Relation Module(RM). Another training-free approach is to aggregate the base prototypes according to the novel and base class relations. First, we construct the correlation matrix as done in (4) to find class correlations in semantic space, then we choose the base classes that have top-K similarities with a certain novel class to synthesize a pseudo prototype, as shown in Figure 1(c):

$$f_i = \sum_{j=0}^{K} w_{ij}\cdot\mu_j, \quad j \in \{\bar{C}_i\} \quad (7)$$

$$w_{ij} = \frac{exp(m_{ij})}{\sum_{j=0}^{K} exp(m_{ij})} \quad (8)$$

where i is the index of novel classes and C_i is the corresponding selected most similar base classes. μ_j stands for the prototype of the base class j, f_i represents the synthesized prototypes, and m_{ij} is the semantic similarity between novel class i and base class j, which is also computed with cosine similarity as in Eq(4). In order to compute the aggregation factor w_{ij}, we first compute similarities between novel class i and all base classes, then choose the top-K similarity scores with the novel class for normalization to computed to compound factors.

2.4 Inference

During the testing stage, after we extract features both for support and query samples, we treat the novel class as input, either using GM or RM, to generate pseudo novel prototypes, which we incorporate with the support features for modifying the final class centers, where we directly calculate the mean of the two prototypes.

3 Experiments

3.1 Experiment Setup

Dataset. We adopt three standard benchmark datasets that are widely used in few-shot learning: *miniImageNet* [3], a small subset extracted from ImageNet [19], consists of 100 classes with 600 images for each class. Following the setting, We split the data set into 64 classes for training, 16 classes for validation, and 20 classes for testing. *CIFAR-FS* [21], a subset randomly sampled from CIFAR-100 [22], is composed of 100 classes and 600 images for each class. It follows the same settings of split as miniImagNet and all samples have the same resolution of 32×32. *FC-100* [23], another subset chosen from CIFAR-100 with the same data size as CIFAR-FS, but has a different way of splitting settings. Rather than being divided according to classes, It is partitioned into 20 superclasses in total, with 12 for training, 4 for validation, and the rest 4 for testing.

Word-Embedding. Word2Vec [20] is a word embedding model to generate word vectors. It is trained with billions of online articles and sentences to

Table 1. Comparison with SOTA works on *mini*ImageNet and cifar-fs.

Method	Backbone	*mini*ImageNet		cifar-fs	
		1-shot	5-shot	1-shot	5-shot
Matching Net [3]	ResNet-12	65.64 ± 0.20	78.72 ± 0.15	-	-
MAML [16]	ResNet-18	64.06 ± 0.18	80.58 ± 0.12	-	-
SimpleShot [17]	ResNet-18	62.85 ± 0.20	80.02 ± 0.14	-	-
S2M2R [18]	ResNet-18	64.93 ± 0.18	83.18 ± 0.11	63.66 ± 0.17	76.07 ± 0.19
DeepEMD [24]	ResNet-12	65.91 ± 0.82	82.41 ± 0.56	74.58 ± 0.29	86.92 ± 0.41
DSN [25]	ResNet-12	62.64 ± 0.66	78.83 ± 0.45	72.30 ± 0.80	85.10 ± 0.60
MetaOptNet [26]	ResNet-12	62.64 ± 0.61	78.63 ± 0.46	72.80 ± 0.70	85.00 ± 0.50
RFS [27]	ResNet-12	62.02 ± 0.63	79.64 ± 0.44	71.50 ± 0.80	86.00 ± 0.50
Inv-equ [28]	ResNet-12	67.28 ± 0.80	84.78 ± 0.50	77.87 ± 0.85	89.74 ± 0.57
R2-D2 [29]	ResNet-12	64.79 ± 0.45	81.08 ± 0.32	76.51 ± 0.47	87.63 ± 0.34
EASY [15]	ResNet-12	70.63 ± 0.20	86.28 ± 0.12	75.24 ± 0.20	88.38 ± 0.14
EASY $3 \times$ [15]	ResNet-12	71.75 ± 0.19	87.15 ± 0.12	76.20 ± 0.20	89.00 ± 0.14
FewTRUE [30]	ResNet-12	72.40 ± 0.78	86.38 ± 0.49	77.76 ± 0.81	88.90 ± 0.59
HCTransformer [31]	ResNet-12	74.74 ± 0.17	85.66 ± 0.10	78.89 ± 0.18	87.73 ± 0.11
Ours-baseline	ResNet-12	70.56 ± 0.20	86.23 ± 0.12	76.36 ± 0.20	88.98 ± 0.14
Ours-RM	ResNet-12	72.41 ± 0.19	86.45 ± 0.12	79.19 ± 0.21	89.48 ± 0.14
Ours-GM	ResNet-12	76.46 ± 0.18	86.71 ± 0.12	82.69 ± 0.18	89.56 ± 0.14

Note: Average 5-way 1-shot and 5-way 5-shot accuracy (%) with 95% confidence intervals. Ours-RM denotes our method using Relation Module to create the pseudo support sample. Ours-GM denotes our method using Generative Module to create the support sample. The best and the second performance is highlighted in red and blue.

extract word relations and creates an embedding for each word, so the embeddings for semantically-close words keep a similar correlation in the higher dimension. Here we use the pre-trained model offered by Google to directly generate embeddings for classes for simplicity.

Implementation Details. We set both β and γ to 0.5. For fair comparisons, we use ResNet12 as backbone and adopt SGD for optimizer with a weight decay of 5e-4 and momentum of 0.9. The learning rate is initialized to 0.1 and adapted with a cosine learning rate scheduler. During the training, we also implement the strategy of S2M2R [18] to build a two-branch model for self-supervision, as done in baseline [15]. To testify our design, we follow the 5way-1shot and 5way-5shot paradigms to randomly generate 2000 episodes from test sets with 15 query samples for each class and report the mean accuracy with 95% confidence interval.

3.2 Comparison with SOTA

Table 1 compares the results on 5-way 1-shot and 5-way 5-shot benchmarks of our methods on *mini*ImageNet and cifar-fs with other state-of-the-art few-shot learning methods. Our baseline model shows a close level with [15] since we keep its basic training structure. After including pseudo generation, we obtain 1.85% $\sim$ 2.83% improvements on 1-shot task by using RM, and 5.9% $\sim$ 6.33% improvements on 1-shot task by using GM. For the *mini*ImageNet dataset, our methods achieve the best performance on 1-shot task, outperforming the current best one by 1.72%, and the second best performance on 5-shot task. For cifar-fs dataset, our methods achieve the best by using GM and the second best by using RM on both tasks. Table 2 compares our method with the others on FC-100. It shows that both GM and RM pseudo generation can realize certain improvements on the baseline model, and still achieve first class by surpassing [15] by 1.3%

Table 2. Comparison with SOTA works on FC-100.

Method	Backbone	FC-100	
		1-shot	5-shot
DeepEMD [24]	ResNet-12	46.60 ± 0.26	63.22 ± 0.71
TADAM [23]	-	40.10 ± 0.40	56.10 ± 0.40
MetaOptNet [26]	ResNet-12	47.20 ± 0.60	62.50 ± 0.60
RFS [27]	ResNet-12	42.60 ± 0.70	59.10 ± 0.60
Inv-equ [28]	ResNet-12	47.76 ± 0.77	65.30 ± 0.76
R2-D2 [29]	ResNet-12	44.75 ± 0.43	59.94 ± 0.41
EASY [15]	ResNet-12	47.94 ± 0.19	64.14 ± 0.19
FewTRUE [30]	-	47.68 ± 0.78	63.81 ± 0.75
Ours-baseline	ResNet-12	47.06 ± 0.19	63.63 ± 0.19
Ours-RM	ResNet-12	47.98 ± 0.20	63.88 ± 0.19
Ours-GM	ResNet-12	49.06 ± 0.19	64.03 ± 0.19

Note: Average 5-way 1-shot and 5-way 5-shot accuracy (%) with 95% confidence intervals. The best and the second performance is highlighted in red and blue.

on 1-shot task, but lags behind the best model on 5-shot task.

In total, compared with the well-behaved methods [15, 30, 31], our proposed method is more competitive with convincing results, which indicates that it can produce more representative features by matching features and word embeddings, along with pseudo generation to show it effectiveness. Such an advantage is more obvious as the number of support data decreases.

3.3 Ablation Study

Influence of Modules. Our method can be divided into the use of proposed constraints and ways of pseudo generation. To explore their influence on the final outcome, we further conduct experiments with or without these designs. Table 3 illustrates that Relation constraint is more effective than Abosulte constraint, which indicates that rather than directly aligning the features to corresponding word embeddings, constructing the relations as in word embedding space is more helpful. But using both constraints can achieve further improvement. On the other hand, pseudo samples produced by GM outperform those by RM, which can infer that the class relation between base and novel classes in word embedding may differ in that in visual space, although we have matched those relations among base classes. This conclusion can also be testified when using two structures at the same time, the performance degrades compared using GM only. We infer that this is because the base class features we use to generate pseudo samples cannot cover the whole information of novel class, and it is unnecessary to take all base classes into consideration since it may not be helpful. In total, base on current dataset information, directly generating with FC layers can achieve a better outcome.

Table 3. Abalation study of modules on *mini*ImageNet.

Baseline	$\mathcal{L}_{AC}$	$\mathcal{L}_{RC}$	RM	GM	1-shot
✓					70.56 ± 0.20
✓	✓		✓		71.87 ± 0.19
✓	✓			✓	74.86 ± 0.18
✓		✓	✓		72.03 ± 0.19
✓		✓		✓	76.16 ± 0.18
✓	✓	✓	✓		72.41 ± 0.19
✓	✓	✓		✓	$\mathbf{76.46 \pm 0.18}$
✓	✓	✓	✓	✓	75.76 ± 0.18

Note: Average 5-way 1-shot accuracy (%) with 95% confidence intervals.

Influence of hyperparameters. To explore the influence of hyperparameters on loss functions. we conduct experiment on $beta$ and $gamma$ separately by freezing one on 0.5 while testing the other. The comparison in Table 4 shows that the performance reaches the best when two hyperparameters are both set to 0.5. In detail, the accuracy in 1-shot case varies more rapidly when changing $gamma$ than $beta$, which implies that the model is more sensitive to the learning of relation graph than direct feature representation.

Table 4. Abalation study of hyperparameters on *mini*ImageNet.

β	γ	1-shot
0.5	0.3	76.17 ± 0.20
0.5	0.4	76.37 ± 0.19
0.5	0.5	76.90 ± 0.18
0.5	0.6	76.36 ± 0.19
0.5	0.7	76.25 ± 0.18
0.3	0.5	76.62 ± 0.19
0.4	0.5	76.72 ± 0.18
0.5	0.5	76.90 ± 0.18
0.6	0.5	76.63 ± 0.19
0.7	0.5	76.54 ± 0.18

Note: Average 5-way 1-shot accuracy (%) with 95% confidence intervals.

3.4 Computational complexity Study

To test the computation complexity, we conduct the experiment on baseline model and our designed model on GFLOPS and amount of parameters. The outcome is shown in Table 5, it shows that compared to baseline model, the majority of extra computation cost lies on additional mapping structure while using GM, which is not significant. For using RM, the extra computation cost is $\mathcal{O}(MN)$ by calculating similarities between novel class and base class matrix, which is far below the record scale. In total, our extra designed structures do not consume much resources.

Table 5. Experiment on GFLOPs and parameter amount on baseline and our-desgined model.

Method	GFLOPs	Param.
baseline	3.52304	12.66M
Our-RM	3.52324	12.66M
Our-GM	4.12425	13.26M

3.5 Visualization

To make a direct demonstration, we conduct t-SNE plots in inference stage for clear comparison. The visualization of two test episodes are shown in Figure 2. In comparison, the class centers in our method using

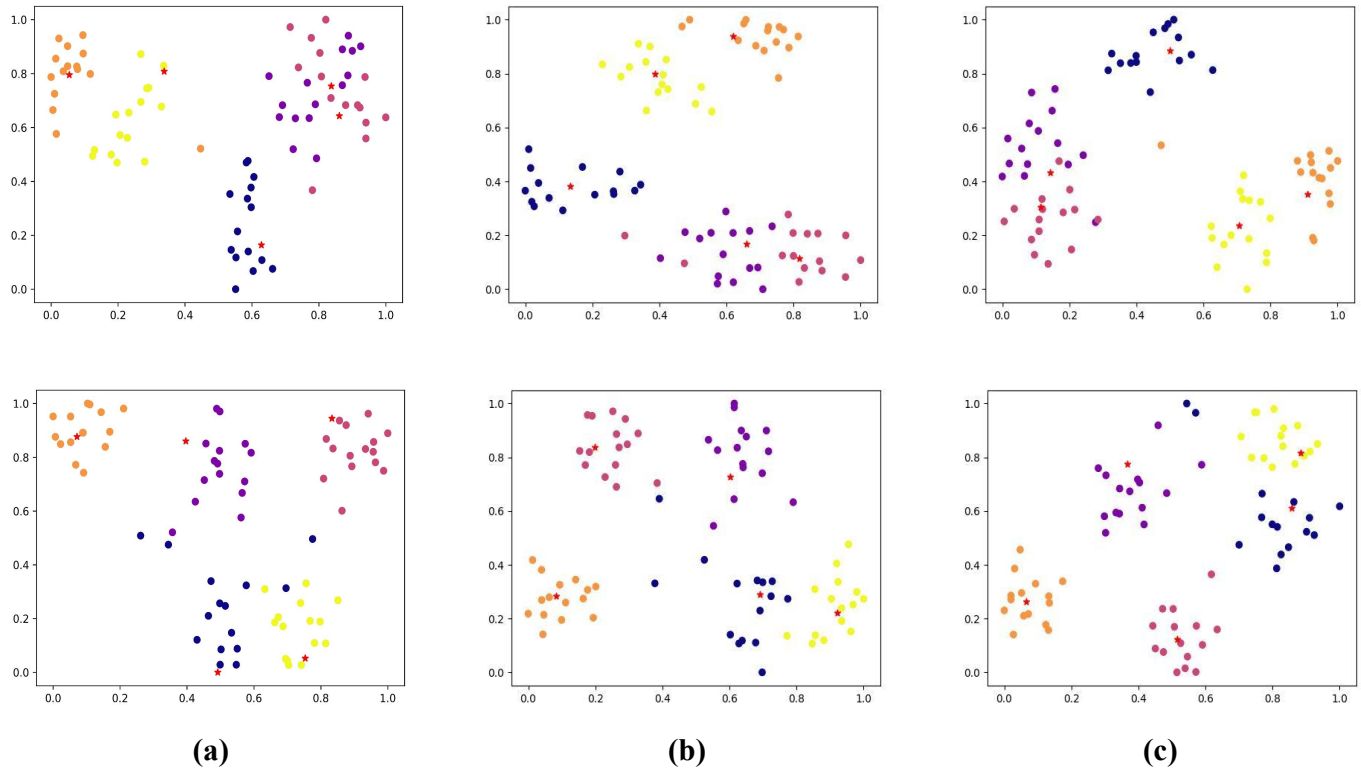


Figure 2. Visualization by t-SNE of test episodes in inference stage on *mini*ImageNet: (a) Baseline model. (b) Ours-RM. (c) Ours-GM. In the figures the dot of different colors represent the query samples and the red stars represent class centers.

RM/GM modules are refined to the ideal positions, which are closer to the real centers of clusters, proving the effection of our method.

3.6 Discussion

Although our method achieves good performance in few-shot datasets, there stills exists limitations. At first, our design is not an end-to-end structure, which needs extra training in GM case or novel class relations calculations in RM case. But designing end-to-end ones like GAN may be unstable because the volume of few-shot datasets is small. Secondly, the model lacks information when novel superclasses appears in inference stage, the trained relation graph is not well useful. In further study, these problems can be solved by using Large language models or supporting the model with larger datasets.

4 Conclusion

This paper proposes a few-shot learning method that utilizes word embeddings to solve the lack of data, which is a better fit for the real-world classification case. We introduce Absolute Constraint($\mathcal{L}_{AC}$) for direct feature assignment and Relation Constraint($\mathcal{L}_{RC}$) for class relation transfer to visual domain, which establishes a semantic-related structure for good reasoning. For clear advantage, we design Generative Module(GM) and Relation Module(RM) to construct pseudo support samples for the novel set to guide the recognition. Extensive experiments on multiple datasets and ablation study have proved our baseline model has achieved large improvement compared with baseline model. It provides an inspiration that good feature relation is helpful for few-shot learning.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by National Natural Science Foundation of China under Grant U20B2075.

Conflicts of Interest

Zijing Liu and Chenggang Wang are employees of the No.10th Research Institute, China Electronics Technology Group Corporation, Chengdu 610036, China.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Kim, J., Oh, T.-H., Lee, S., Pan, F., & Kweon, I. S. (2019). Variational prototyping-encoder: One-shot learning with prototypical images. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 9462–9470).

[2] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H. S., & Hospedales, T. M. (2018). Learning to compare: Relation network for few-shot learning. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1199–1208).

[3] Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one shot learning. *Advances in neural information processing systems*, 29.

[4] Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30.

[5] Huang, J., Chen, F., Wang, K., Lin, L., & Zhang, D. (2022). Enhancing prototypical few-shot learning by leveraging the local-level strategy. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1660–1664). IEEE. [CrossRef]

[6] Chen, Y., Liu, Z., Xu, H., Darrell, T., & Wang, X. (2021). Meta-baseline: Exploring simple meta-learning for few-shot learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9062-9071).

[7] Schwartz, E., Karlinsky, L., Feris, R., Giryes, R., & Bronstein, A. M. (2022). Baby steps towards few-shot learning with multiple semantics. *Pattern Recognition Letters*, 160, 142–147. [CrossRef]

[8] Li, A., Huang, W., Lan, X., Feng, J., Li, Z., & Wang, L. (2020). Boosting few-shot learning with adaptive margin loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12576-12584).

[9] Xing, C., Rostamzadeh, N., Oreshkin, B., & O Pinheiro, P. O. (2019). Adaptive cross-modal few-shot learning. *Advances in neural information processing systems*, 32.

[10] Xu, J., & Le, H. (2022). Generating representative samples for few-shot classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9003-9013).

[11] Chen, Z., Fu, Y., Zhang, Y., Jiang, Y.-G., Xue, X., & Sigal, L. (2019). Multi-level semantic feature augmentation for one-shot learning. *IEEE Transactions on Image Processing*, 28(9), 4594–4605. [CrossRef]

[12] Peng, Z., Li, Z., Zhang, J., Li, Y., Qi, G. J., & Tang, J. (2019). Few-shot image recognition with knowledge transfer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 441-449).

[13] Tokmakov, P., Wang, Y. X., & Hebert, M. (2019). Learning compositional representations for few-shot recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6372-6381).

[14] Zhang, B., Li, X., Ye, Y., Huang, Z., & Zhang, L. (2021). Prototype completion with primitive knowledge for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3754-3762).

[15] Bendou, Y., Hu, Y., Lafargue, R., Lioi, G., Pasdeloup, B., Pateux, S., & Gripon, V. (2022). Easy—ensemble augmented-shot-y-shaped learning: State-of-the-art few-shot classification with simple components. *Journal of Imaging*, 8(7), 179. [CrossRef]

[16] Finn, C., Abbeel, P., & Levine, S. (2017, July). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126-1135). PMLR.

[17] Wang, Y., Chao, W. L., Weinberger, K. Q., & Van Der Maaten, L. (2019). Simpleshot: Revisiting nearest-neighbor classification for few-shot learning. *arXiv preprint arXiv:1911.04623*.

[18] Mangla, P., Kumari, N., Sinha, A., Singh, M., Krishnamurthy, B., & Balasubramanian, V. N. (2020). Charting the right manifold: Manifold mixup for few-shot learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 2218-2227).

[19] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

[20] Rong, X. (2014). word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*.

[21] Bertinetto, L., Henriques, J. F., Torr, P. H., & Vedaldi, A. (2018). Meta-learning with differentiable closed-form solvers. *arXiv preprint arXiv:1805.08136*.

[22] Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images. Technical Report, University of Toronto.

[23] Oreshkin, B., Rodríguez López, P., & Lacoste, A. (2018). Tadam: Task dependent adaptive metric for improved few-shot learning. *Advances in neural information processing systems*, 31.

[24] Zhang, C., Cai, Y., Lin, G., & Shen, C. (2020). DeepEMD: Few-shot image classification with differentiable earth mover's distance and structured classifiers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12203-12213).

[25] Simon, C., Koniusz, P., Nock, R., & Harandi, M. (2020). Adaptive subspaces for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4136-4145).

[26] Lee, K., Maji, S., Ravichandran, A., & Soatto, S. (2019). Meta-learning with differentiable convex optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10657-10665).

[27] Tian, Y., Wang, Y., Krishnan, D., Tenenbaum, J.

B., & Isola, P. (2020). Rethinking few-shot image classification: a good embedding is all you need?. In *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16 (pp. 266-282). Springer International Publishing.

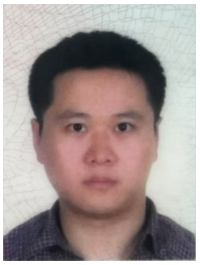
[28] Rizve, M. N., Khan, S., Khan, F. S., & Shah, M. (2021). Exploring complementary strengths of invariant and equivariant representations for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10836-10846).

[29] Liu, J., Chao, F., & Lin, C. M. (2020). Task augmentation by rotating for meta-learning. *arXiv preprint arXiv:2003.00804*.

[30] Hiller, M., Ma, R., Harandi, M., & Drummond, T. (2022). Rethinking generalization in few-shot classification. *Advances in neural information processing systems*, 35, 3582-3595.

[31] He, Y., Liang, W., Zhao, D., Zhou, H. Y., Ge, W., Yu, Y., & Zhang, W. (2022). Attribute surrogates learning and spectral tokens pooling in transformers for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9119-9129).

Zijing Liu received the M.E. degree in information and communication engineering from University of Electronic Science and Technology of China, Chengdu 611731, China, in 2024. (Email: 853000906@qq.com)



Chenggang Wang received the D.E. degree in signal and information processing from Institude of Optics and Electronics, Chinese Academy of Sciences, Cheng 610209, China, in 2007. (Email: cgwang@126.com)