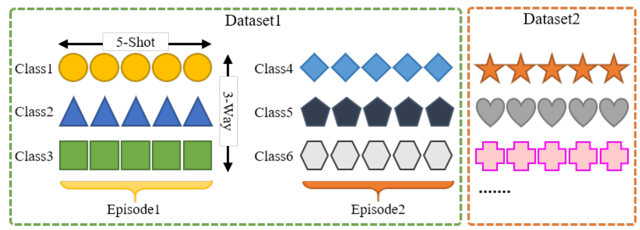
Graph Attention Prototype Network for Robust Few-Shot Classification
Supplemental Material

Tingyun Liu[1], Licheng Liu[1,*], Qibin Zhang[1], Qiying Feng[2], C. L. Philip Chen[3]
[1]Hunan University, [2]Guangzhou University, [3]South China University of Technology
{liutingyun, lichengliu, zhangqibin135}@hnu.edu.cn, qy@gzhu.edu.cn, Philip.Chen@ieee.org

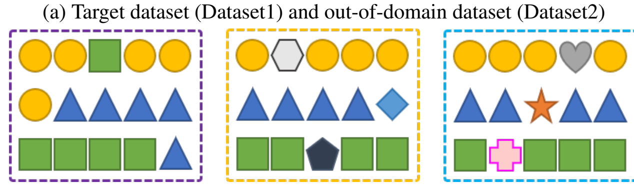
1. Noisy support set examples

Humans can typically separate noisy samples from clean ones through careful observation, largely due to their preexisting conceptual knowledge of the classes involved. In contrast, few-shot models facing noisy support sets must learn to distinguish the target classes without prior exposure to them, making the task far more challenging. To simulate realistic conditions, we introduce three distinct types of label noise into both support and query sets. A concrete example of a 3-way 5-shot task with the above three types of noise (at a noise ratio $R = 20\%$) helps illustrate these definitions, as shown in Fig. 1. Assume that the model is trained on Dataset1, which contains 6 classes with 5 images per class. In the first episode, Class1, Class2 and Class3 form the support set. In the second episode, Class4, Class5 and Class6 are selected. Taking Episode1 as the current episode, IE noise indicates mislabeled samples originate from the same dataset and episode but different classes, OOE noise denotes mislabeled samples come from the same dataset but different episodes, while OOD noise signifies mislabeled samples drawn from other datasets.

IE noise reflects intra-dataset class confusion (*e.g.*, radiologists misclassifying visually similar diseases). OOE noise models episode-wise annotation standard inconsistency (*e.g.*, varying thresholds for distinguishing "shrubland" and "grassland" in remote sensing recognition). OOD noise simulates external contamination, such as automatic labeling systems inadvertently merging images from other datasets into the target one. Real-world noise is often mixed and more complex. However, GAPNet suppresses unreliable samples via feature similarity rather than nominal labels, making it robust across all noise types.



(a) Target dataset (Dataset1) and out-of-domain dataset (Dataset2)



(b) IE noise (c) OOE noise (d) OOD noise

Figure 1. A 3-way 5-shot FSL task with three types of label noise. IE noise is generated by permuting samples within the same episode but across different classes. OOE noise originates from classes within the same dataset but outside the current few-shot episode. OOD noise is drawn from entirely different datasets.

2. Datasets

The effectiveness and robustness of GAPNet are evaluated through extensive experiments on four image benchmarks that are commonly used in few-shot classification tasks.

- CIFAR-FS [1] is a few-shot learning benchmark adapted from the CIFAR-100 dataset [4]. It includes 100 classes, with each class containing 600 color images of size 32×32 pixels. These classes are reorganized into three disjoint subsets: 64 base classes, 16 validation classes and 20 novel classes. This split is well-suited for evaluating the generalization performance of FSL models.
- miniImageNet [12] is derived from the ImageNet dataset [9], comprising 100 classes with 600 images per class. For dataset splitting, we follow the division proposed in [7], where 64 classes are used for training, 16 for validation, and 20 for testing, respectively.
- tieredImageNet [8] is a larger subset of ImageNet, designed to address the limitation of fine-grained class similarity in miniImageNet. It groups similar classes into higher-level classes, resulting in a total of 608 superclasses. Following the data partitioning approach introduced in [8], these superclasses are split into 351 for training, 97 for validation, and 60 for testing.

*Corresponding author: Licheng Liu

- CUB-200-2011 [13] serves as a fine-grained few-shot classification benchmark, which is widely used to evaluate the ability of FSL models to extract discriminative features. It consists of 11788 images from 200 bird classes, with each class containing 30–60 images. Following [3], this dataset is split into 100 training classes, 50 validation classes, and 50 testing classes.

3. Model complexity

Model complexity of GAPNet against competitive FSL methods is evaluated on MiniImageNet in Tab. 1, with metrics including MFLOPs (FLOPs scaled by 1e-6), parameters (Params), and inference time. All results are measured across 200 5-way 5-shot 15-query tasks. As shown in Tab. 1, the trainable parameters of GAPNet are not fewer than those of most competitive models, a result of the construction of complex relation graph and the attention mechanisms in feature extraction and prototype generation procedures. Its MFLOPs are also higher than many baseline methods, primarily due to the extra computations required by its GABL, PLGC, and ANRPG modules. Notably, GAPNet is not the most parameter-dense or FLOP-heavy model available. Unsurprisingly, the higher FLOPs and larger parameter of GAPNet count lead to longer training time, yet surprisingly, its inference time remains comparable to other complex models, making it well-suited for practical few-shot classification scenarios despite its enhanced functional modules.

Table 1. Comparisons of model complexity

Methods	MFLOPs	Params	inference time(s)
RapNets [5]	9956.46	131493	22.29
ProtoNet [10]	9956.07	113088	16.64
MatchingNet [12]	10355.83	779840	24.91
RelationNet [11]	24853.77	236881	17.68
RNNP [6]	9965.04	164353	16.04
HGNN [15]	9964.75	121664	19.61
BiFRN [14]	10050.15	150528	16.87
APPN [2]	10081.43	470772	30.02
GAPNet	10078.45	915355	35.39

4. Edge Pruning

During training, a stochastic edge pruning strategy is adopted to balance important edge preservation and structural diversity, that is, edges with low α are discarded with a probability proportional to their importance, retaining only reliable connections. Set the edge pruning rate to ρ, the base edge retaining rate is $\sigma_{\text{base}} = 1 - \rho$, the probability of retaining edge between node i and j is defined as

$$\sigma = \sigma_{\text{base}} + \delta(\frac{\alpha_{ij}}{\max(\alpha) + \epsilon} - 1) \tag{1}$$

where α_{ij} is the attention score of edge $V(i, j)$, ϵ is a small constant for numerical stability, and $\delta = \rho/2$ controls the dynamic adjustment range of edge retaining probability, balancing the influence of attention scores against randomness during stochastic edge pruning. This strategy progressively reinforces reliable connections while suppressing potentially noisy edges, creating a self-cleaning graph structure over training iterations.

The edge pruning rate ρ controls how many low-importance edges are pruned in the PLGC module. To verify its impact on model accuracy, ρ is varied from 0 to 1 in steps of 0.1. with the comparison results shown in Fig. 2. As observed, the model accuracy increases slightly when ρ ranges from 0 to 0.2. At $\rho = 0$, the accuracy is 60.26%, indicating that the graph contains redundant or noisy edges that disrupt effective sample relation modeling. As ρ increases to 0.2, the accuracy peaks at 61.62%, as moderate pruning filters low-importance invalid edges while preserving core effective connections, thereby optimizing the structural quality of the relation graph. However, once ρ exceeds 0.2, the precision decreases steadily. Higher pruning ratios not only remove noise edges but also mistakenly cut valid ones. As ρ approaches 1, nearly all edges are pruned, completely breaking the relationships between samples. This prevents the model from learning intra-class or inter-class patterns, dropping accuracy to 58.31%. In summary, $\rho = 0.2$ is the optimal value, it balances noise filtering and relation preservation, while excessively large or small ρ undermines performance.

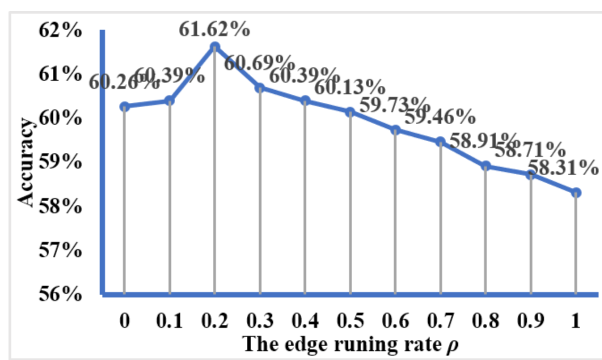


Figure 2. Effect of the edge pruning rate ρ on miniImageNet with 20% IE noise.

5. Visualizations of learned graphs and edge attention

We have visualized edge-attention distribution histogram with KDE curves and graph connectivity under 20% IE noise on miniImagenet. As shown in Fig. 3, EGAT assigns higher attention to clean and intra-class edges, though some noise-induced edges remains. GAPNet maintains dense intra-class edges (solid lines) and sparse reliable inter-class edges (dashed lines), demonstrating that PLGC effectively aggregates intra-class samples despite noise. These results indicate that pseudo-labels stably guiding for robust graph construction under label noise.

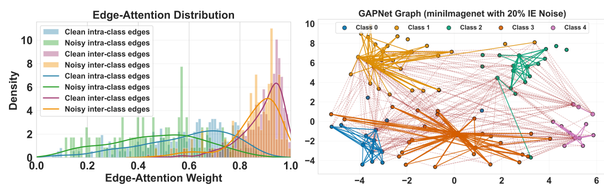


Figure 3. Visualizations of edge attention and learned graph.

6. Visualization of weighting mechanisms

Our weighting scheme (Eq.(9)) uses an exponential function to maintain high weights near class centers and rapidly reduce the weights of distant noisy samples. The scaling factor κ controls steepness of weight decay. Larger κ leads to stronger noise suppression but may also downweight clean samples and vice versa. Euclidean distance is adopted for its ability to capture local structure of feature space, accurately reflecting sample-center distances in sparse feature distributions to enable reliable weighting. The visualizations (Fig. 4) demonstrate that our weighting mechanism effectively suppress the contribution of noisy samples and leads to more accurate prototype generation.

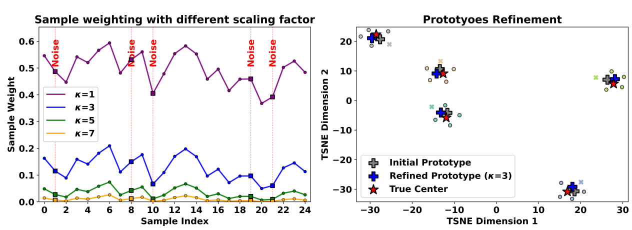


Figure 4. Visualizations of weighting and prototype refinement.

References

[1] Luca Bertinetto, Joao F. Henriques, Philip Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In *International Conference on Learning Representations*, 2019. 1

[2] Jiaqi Chen, Shizhuo Deng, Da Teng, Dongyue Chen, Tong Jia, and Hao Wang. APPN: An attention-based pseudo-label propagation network for few-shot learning with noisy labels. *Neurocomputing*, 602:128212, 2024. 2

[3] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *International Conference on Learning Representations*, 2019. 2

[4] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. 1

[5] Jiang Lu, Sheng Jin, Jian Liang, and Changshui Zhang. Robust few-shot learning for user-provided data. *IEEE Transactions on Neural Networks and Learning Systems*, 32(4): 1433–1447, 2021. 2

[6] Pratik Mazumder, Pravendra Singh, and Vinay P. Namboodiri. RNNP: A robust few-shot learning approach. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2663–2672, 2021. 2

[7] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International Conference on Learning Representations*, 2017. 1

[8] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, Hugo Larochelle, and Richard S. Zemel. Meta-learning for semi-supervised few-shot classification, 2018. 1

[9] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115 (3):211–252, 2015. 1

[10] Jake Snell, Kevin Swersky, and Richard S Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4077–4087, 2017. 2

[11] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018. 2

[12] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, koray kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, pages 3637–3645. Curran Associates, Inc., 2016. 1, 2

[13] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 2

[14] Jijie Wu, Dongliang Chang, Aneeshan Sain, Xiaoxu Li, Zhanyu Ma, Jie Cao, Jun Guo, and Yi-Zhe Song. Bidirectional feature reconstruction network for fine-grained few-shot image classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2821–2829, 2023. 2

[15] Tianyuan Yu, Sen He, Yi-Zhe Song, and Tao Xiang. Hybrid graph neural networks for few-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3179–3187, 2022. 2