Simple Semantic-Aided Few-Shot Learning

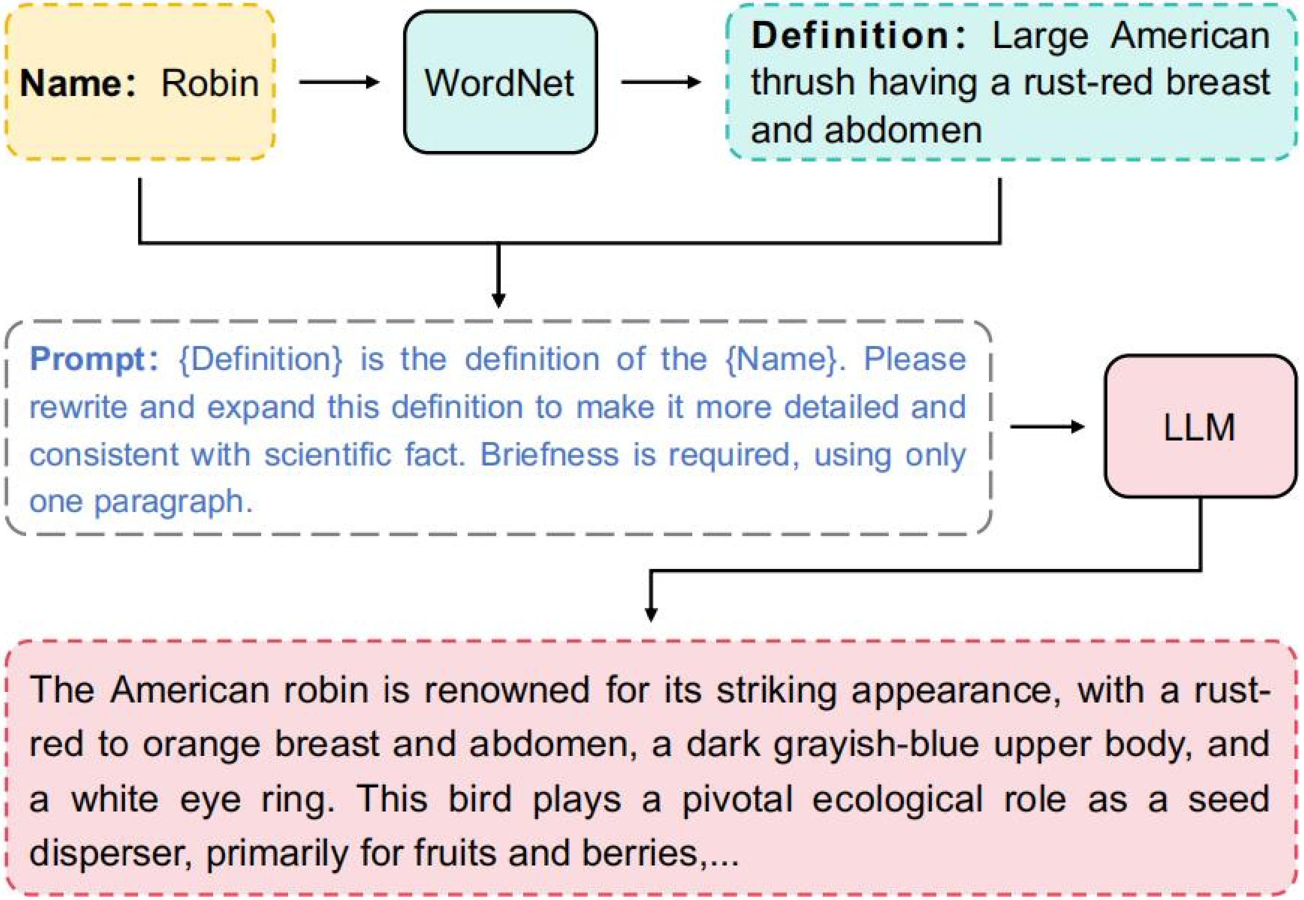
Hai Zhang[1*] Junzhe Xu[1,2*‡] Shanlin Jiang[3] Zhenan He[1†]

[1]College of Computer Science, Sichuan University
[2]Banma Network Technology, Alibaba Group
[3]Naveen Jindal School of Management, University of Texas at Dallas

FRI-PM-428

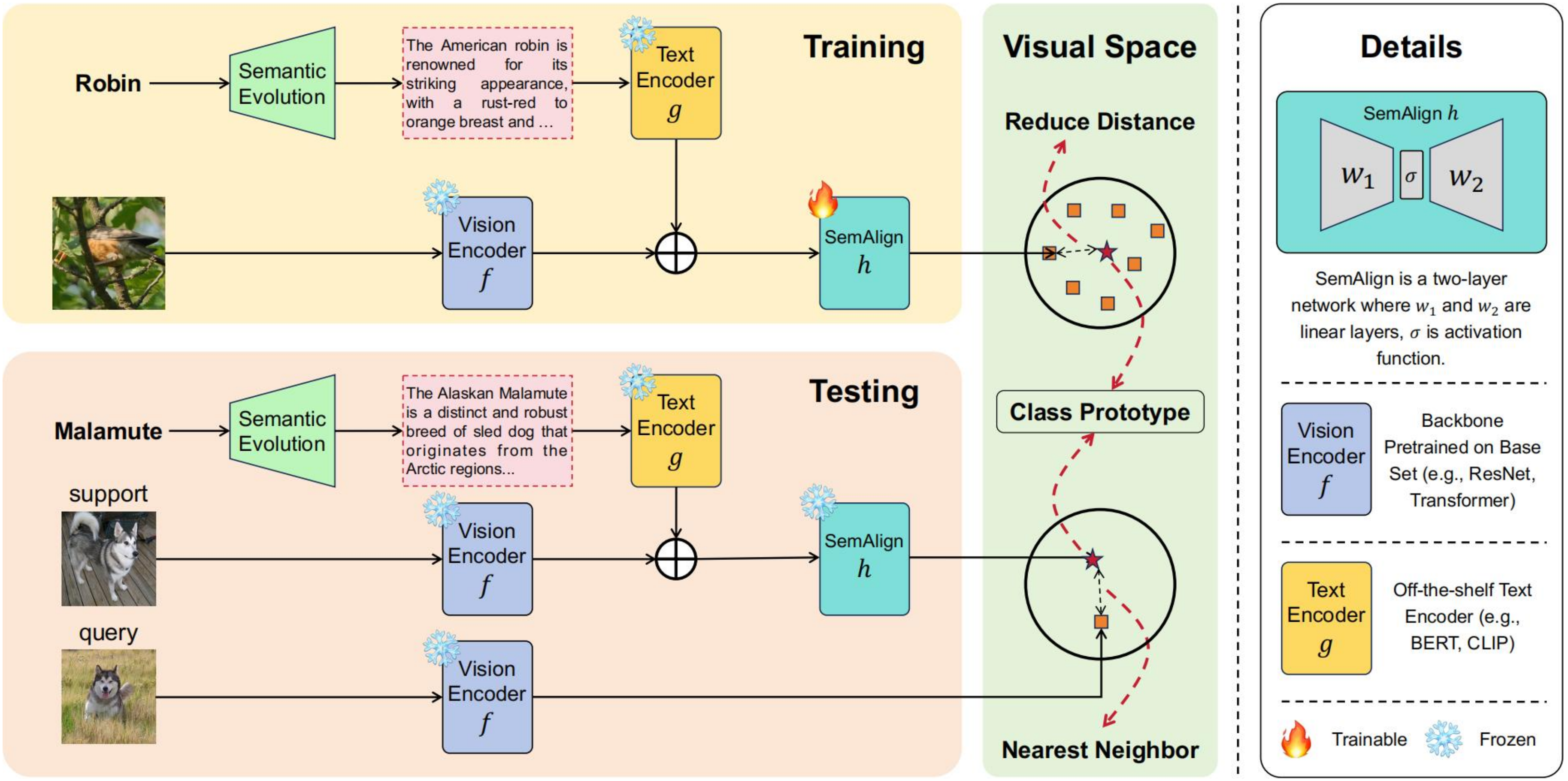
Quick Preview

We propose the **Sem**antic **Evo**lution module to generate high-quality category descriptions.



Quick Preview

We design a Simple **Sem**antic **Align**ment Network that translates high-quality semantic features into robust class prototypes.



Quick Preview

We achieved consistent improvements on two tasks across the four datasets.

	Method	Venue	Backbone	MiniImageNet		TieredImageNet	
				5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
Visual-Based	MatchNet [53]	NeurIPS'16	ResNet-12	65.64 ± 0.20	78.72 ± 0.15	68.50 ± 0.92	80.60 ± 0.71
	ProtoNet [46]	NeurIPS'17	ResNet-12	62.39 ± 0.21	80.53 ± 0.14	68.23 ± 0.23	84.03 ± 0.16
	MAML [13]	ICML'17	ResNet-12	49.24 ± 0.21	58.05 ± 0.10	67.92 ± 0.17	72.41 ± 0.20
	TADAM [37]	NeurIPS'18	ResNet-12	58.50 ± 0.30	76.70 ± 0.30	62.13 ± 0.31	81.92 ± 0.30
	CAN [20]	NeurIPS'19	ResNet-12	63.85 ± 0.48	79.44 ± 0.34	69.89 ± 0.51	84.23 ± 0.37
	CTM [29]	CVPR'19	ResNet-18	64.12 ± 0.82	80.51 ± 0.13	68.41 ± 0.39	84.28 ± 1.73
	RFS [49]	ECCV'20	ResNet-12	62.02 ± 0.63	79.64 ± 0.44	69.74 ± 0.72	84.41 ± 0.55
	FEAT [60]	CVPR'20	ResNet-12	66.78 ± 0.20	82.05 ± 0.14	70.80 ± 0.23	84.79 ± 0.16
	Meta-Baseline [6]	ICCV'21	ResNet-12	63.17 ± 0.23	79.26 ± 0.17	68.62 ± 0.27	83.29 ± 0.18
	SUN [11]	ECCV'22	ViT-S	67.80 ± 0.45	83.25 ± 0.30	72.99 ± 0.50	86.74 ± 0.33
	FewTURE [19]	NeurIPS'22	Swin-T	72.40 ± 0.78	86.38 ± 0.49	76.32 ± 0.87	**89.96 ± 0.55**
	FGFL [7]	ICCV'23	ResNet-12	69.14 ± 0.80	86.01 ± 0.62	73.21 ± 0.88	87.21 ± 0.61
	Meta-AdaM [45]	NeurIPS'23	ResNet-12	59.89 ± 0.49	77.92 ± 0.43	65.31 ± 0.48	85.24 ± 0.35
Semantic-Based	KTN [38]	ICCV'19	Conv-128	64.42 ± 0.72	74.16 ± 0.56	74.16 ± 0.56	-
	AM3 [57]	NeurIPS'19	ResNet-12	65.30 ± 0.49	78.10 ± 0.36	69.08 ± 0.47	82.58 ± 0.31
	TRAML[28]	CVPR'20	ResNet-12	67.10 ± 0.52	79.54 ± 0.60	-	-
	AM3-BERT [59]	ICMR'21	ResNet-12	68.42 ± 0.51	81.29 ± 0.59	77.03 ± 0.85	87.20 ± 0.70
	SVAE-Proto [58]	CVPR'22	ResNet-12	74.84 ± 0.23	83.28 ± 0.40	76.98 ± 0.65	85.77 ± 0.50
	SP-CLIP [5]	CVPR'23	Visformer-T	72.31 ± 0.40	83.42 ± 0.30	78.03 ± 0.46	88.55 ± 0.32
	SemFew	Ours	ResNet-12	77.63 ± 0.63	83.04 ± 0.48	78.96 ± 0.80	85.88 ± 0.62
	SemFew-Trans	Ours	Swin-T	**78.94 ± 0.66**	**86.49 ± 0.50**	**82.37 ± 0.77**	**89.89 ± 0.52**

Table 1. Results (%) on MiniImageNet and TieredImageNet. The ± shows 95% confidence intervals. The best results are shown in **bold**.

Method	Venue	Backbone	CIFAR-FS		FC100	
			5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
ProtoNet [46]	NeurIPS'17	ResNet-12	72.20 ± 0.70	83.50 ± 0.50	41.54 ± 0.76	57.08 ± 0.76
TADAM [37]	NeurIPS'18	ResNet-12	-	-	40.10 ± 0.40	56.10 ± 0.40
MetaOptNet [27]	CVPR'19	ResNet-12	72.80 ± 0.70	84.30 ± 0.50	47.20 ± 0.60	55.50 ± 0.60
MABAS [23]	ECCV'20	ResNet-12	73.51 ± 0.92	85.65 ± 0.65	42.31 ± 0.75	58.16 ± 0.78
RFS [49]	ECCV'20	ResNet-12	71.50 ± 0.80	86.00 ± 0.50	42.60 ± 0.70	59.10 ± 0.60
SUN [11]	ECCV'22	ViT-S	78.37 ± 0.46	88.84 ± 0.32	-	-
FewTURE [19]	NeurIPS'22	Swin-T	77.76 ± 0.81	88.90 ± 0.59	47.68 ± 0.78	63.81 ± 0.75
Meta-AdaM [45]	NeurIPS'23	ResNet-12	-	-	41.12 ± 0.49	56.14 ± 0.49
SP-CLIP [5]	CVPR'23	Visformer-T	82.18 ± 0.40	88.24 ± 0.32	48.53 ± 0.38	61.55 ± 0.41
SemFew	Ours	ResNet-12	83.65 ± 0.70	87.66 ± 0.60	**54.36 ± 0.71**	62.79 ± 0.74
SemFew-Trans	Ours	Swin-T	**84.34 ± 0.67**	**89.11 ± 0.54**	54.27 ± 0.77	**65.02 ± 0.72**

Table 2. Results (%) on CIFAR-FS and FC100. The ± shows 95% confidence intervals. The best results are shown in **bold**.

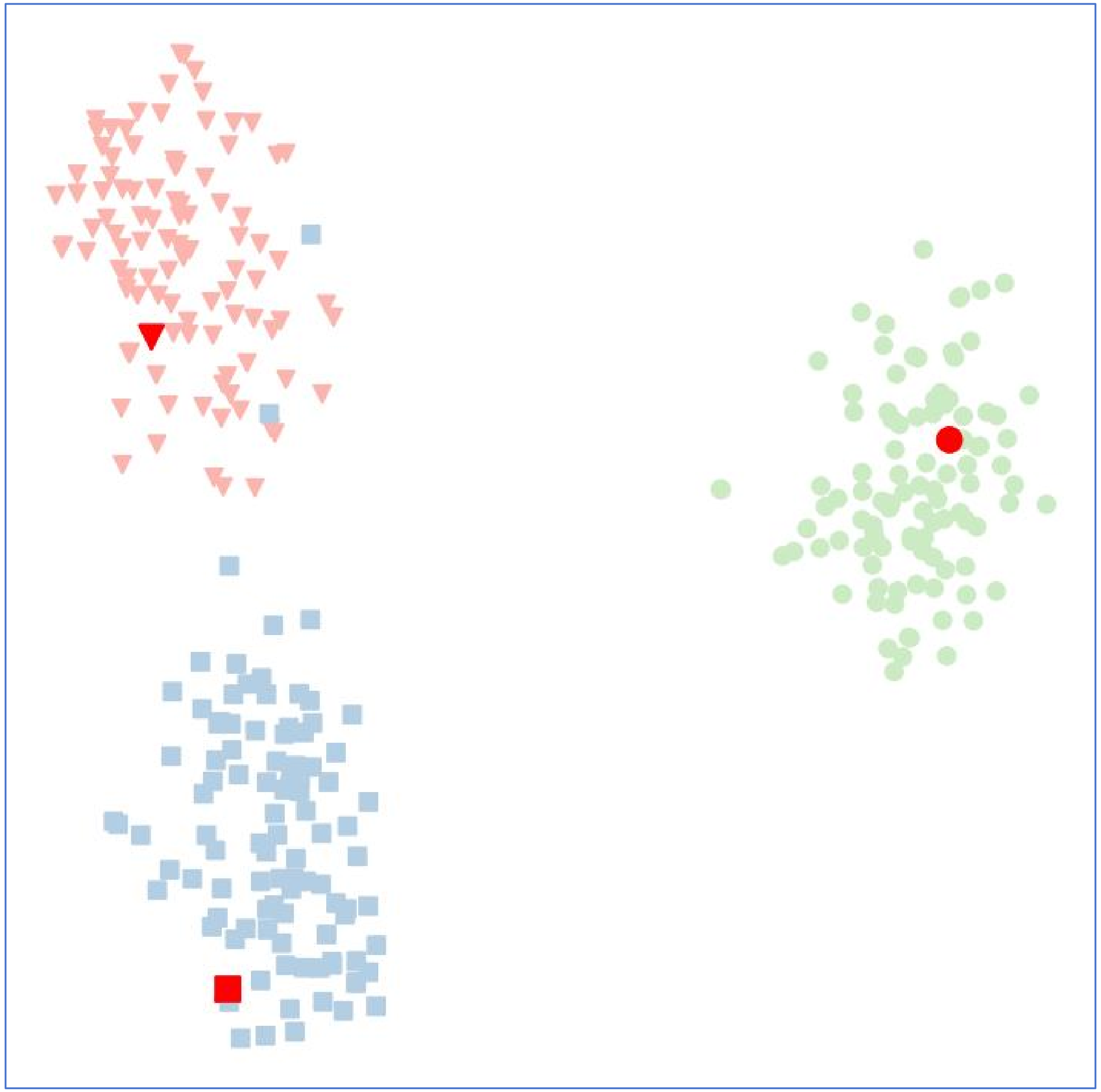
Motivation

Support Samples

■ ▲ ●

Query Samples

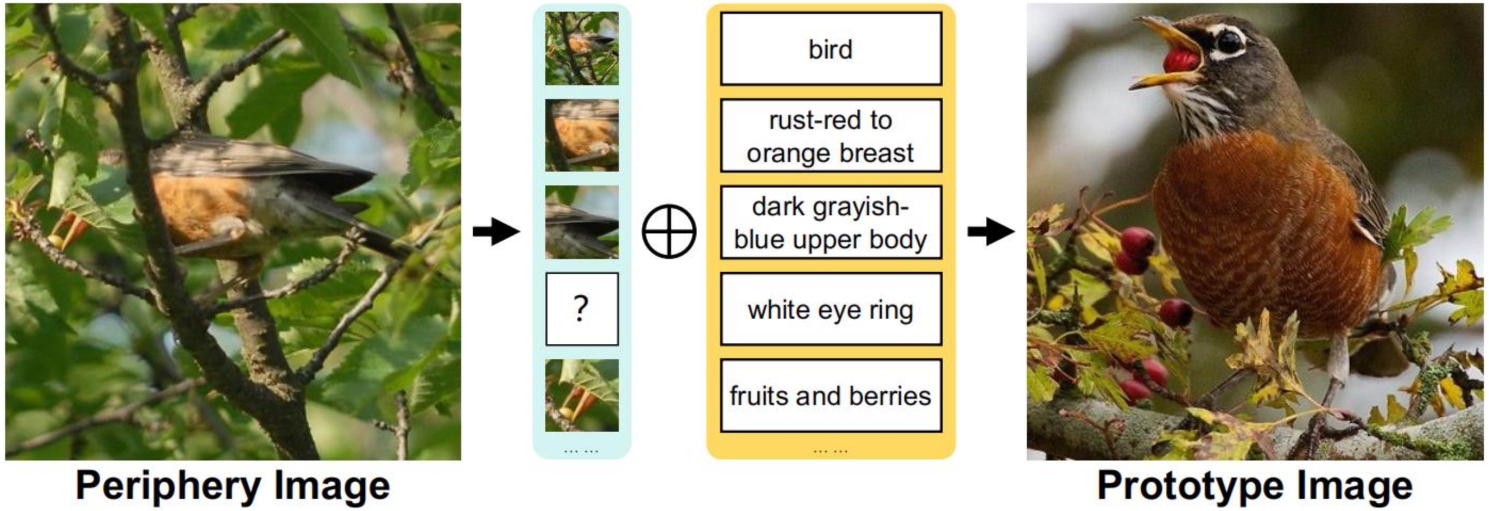
■ ▲ ●



Motivation

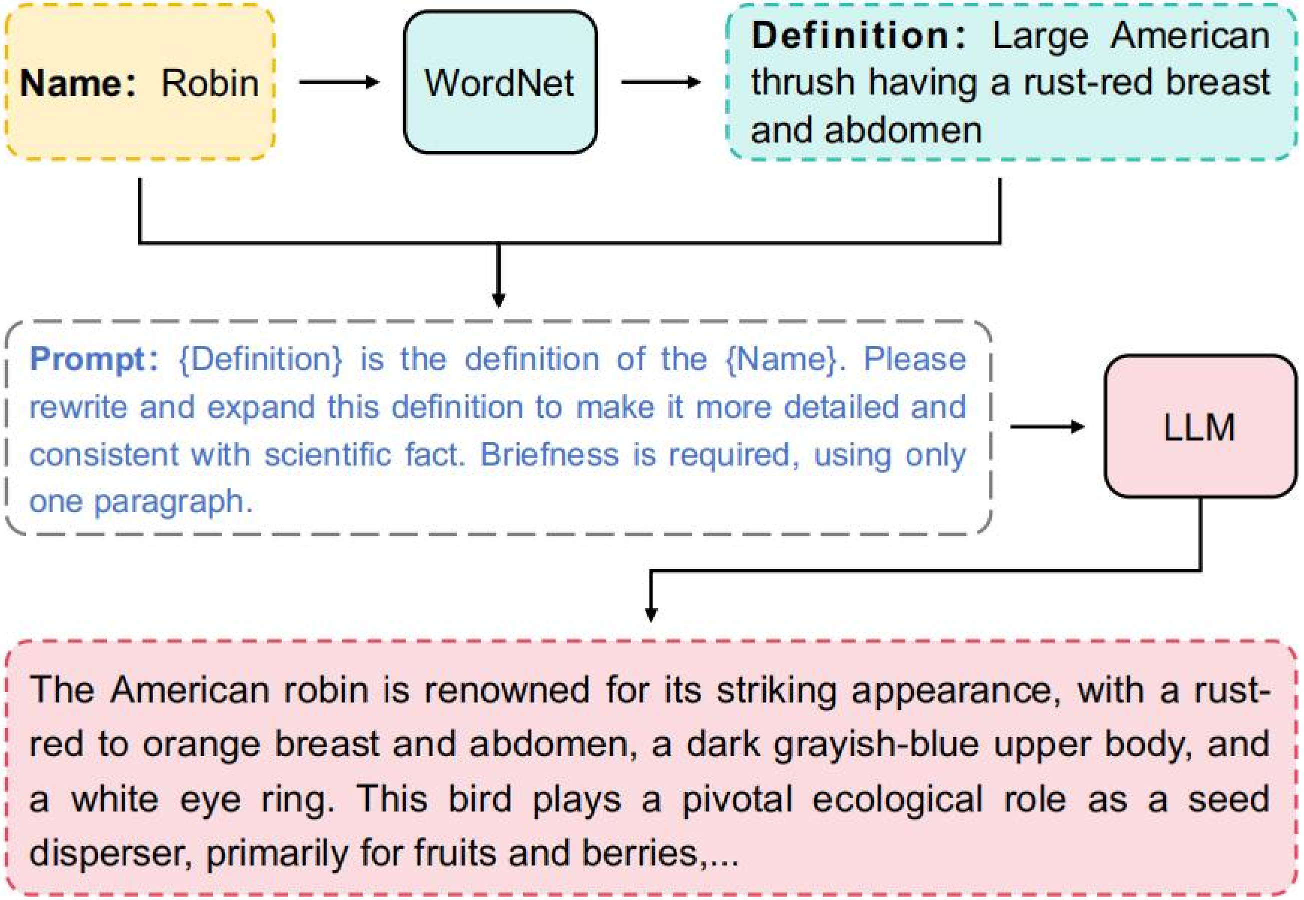
The text contains rich visual descriptions.

The American robin is renowned for its striking appearance, with a **rust-red to orange breast** and abdomen, a **dark grayish-blue upper body**, and a **white eye ring**. This **bird** plays a pivotal ecological role as a seed disperser, primarily for **fruits and berries**,...



Semantic Evolution (SemEvo)

we propose an automatic step-by-step Semantic Evolution process to acquire detailed and accurate semantics.



The framework of SemFew

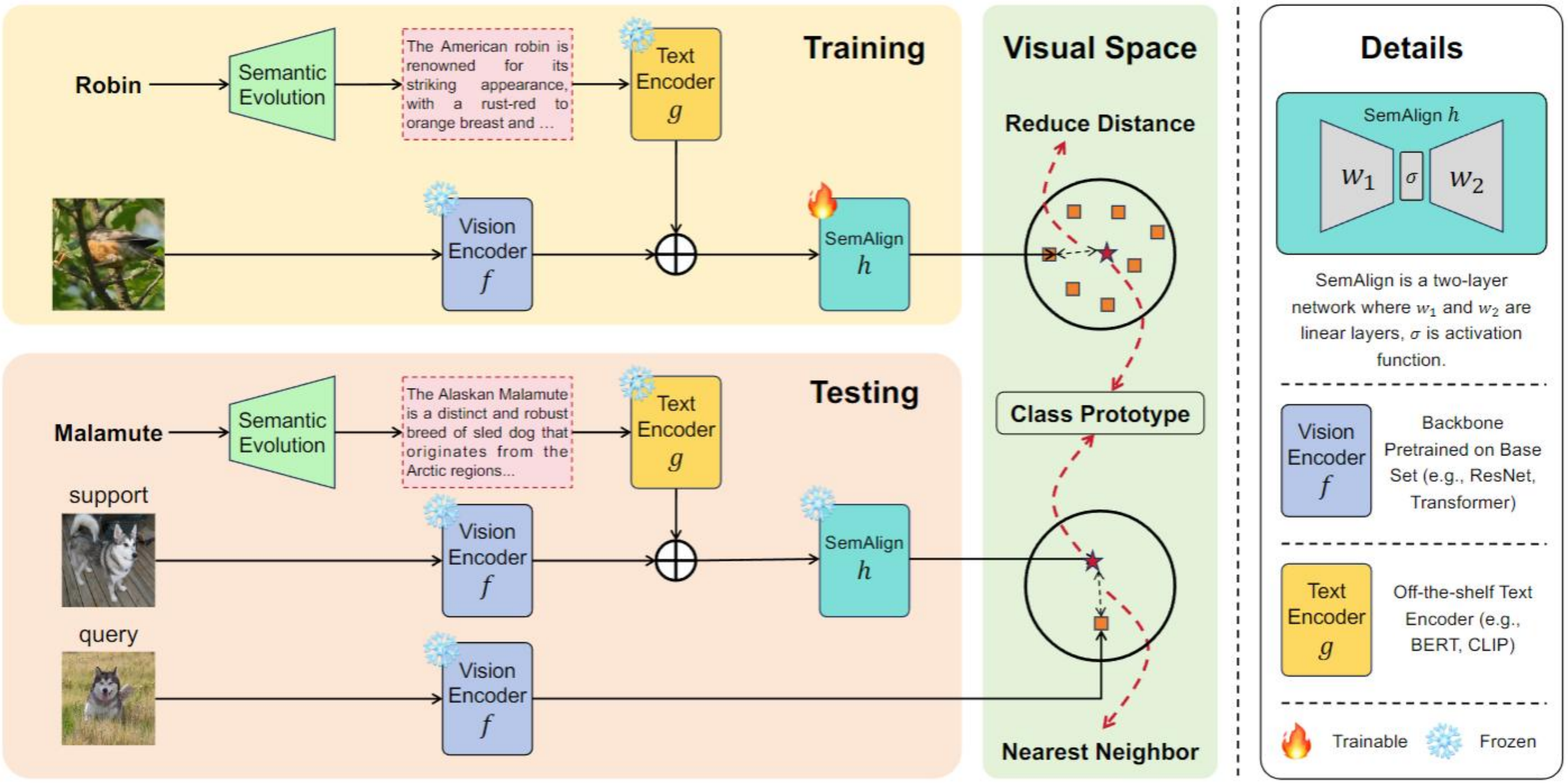


Figure 3. The framework of our proposed SemFew. During the training stage, images and paraphrased semantics are encoded and fed into SemAlign h, with the objective of reducing the distance between the output of h and the class prototype in the visual space. During the testing stage, images in the support set are transformed into class prototypes by h, and query images are classified by identifying the nearest prototype. The symbol $\oplus$ denotes a concatenation operation.

Experimental results

- Few-Shot Image Classification

	Method	Venue	Backbone	MiniImageNet		TieredImageNet	
				5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
Visual-Based	MatchNet [53]	NeurIPS'16	ResNet-12	65.64 ± 0.20	78.72 ± 0.15	68.50 ± 0.92	80.60 ± 0.71
	ProtoNet [46]	NeurIPS'17	ResNet-12	62.39 ± 0.21	80.53 ± 0.14	68.23 ± 0.23	84.03 ± 0.16
	MAML [13]	ICML'17	ResNet-12	49.24 ± 0.21	58.05 ± 0.10	67.92 ± 0.17	72.41 ± 0.20
	TADAM [37]	NeurIPS'18	ResNet-12	58.50 ± 0.30	76.70 ± 0.30	62.13 ± 0.31	81.92 ± 0.30
	CAN [20]	NeurIPS'19	ResNet-12	63.85 ± 0.48	79.44 ± 0.34	69.89 ± 0.51	84.23 ± 0.37
	CTM [29]	CVPR'19	ResNet-18	64.12 ± 0.82	80.51 ± 0.13	68.41 ± 0.39	84.28 ± 1.73
	RFS [49]	ECCV'20	ResNet-12	62.02 ± 0.63	79.64 ± 0.44	69.74 ± 0.72	84.41 ± 0.55
	FEAT [60]	CVPR'20	ResNet-12	66.78 ± 0.20	82.05 ± 0.14	70.80 ± 0.23	84.79 ± 0.16
	Meta-Baseline [6]	ICCV'21	ResNet-12	63.17 ± 0.23	79.26 ± 0.17	68.62 ± 0.27	83.29 ± 0.18
	SUN [11]	ECCV'22	ViT-S	67.80 ± 0.45	83.25 ± 0.30	72.99 ± 0.50	86.74 ± 0.33
	FewTURE [19]	NeurIPS'22	Swin-T	72.40 ± 0.78	86.38 ± 0.49	76.32 ± 0.87	**89.96 ± 0.55**
	FGFL [7]	ICCV'23	ResNet-12	69.14 ± 0.80	86.01 ± 0.62	73.21 ± 0.88	87.21 ± 0.61
	Meta-AdaM [45]	NeurIPS'23	ResNet-12	59.89 ± 0.49	77.92 ± 0.43	65.31 ± 0.48	85.24 ± 0.35
Semantic-Based	KTN [38]	ICCV'19	Conv-128	64.42 ± 0.72	74.16 ± 0.56	74.16 ± 0.56	-
	AM3 [57]	NeurIPS'19	ResNet-12	65.30 ± 0.49	78.10 ± 0.36	69.08 ± 0.47	82.58 ± 0.31
	TRAML[28]	CVPR'20	ResNet-12	67.10 ± 0.52	79.54 ± 0.60	-	-
	AM3-BERT [59]	ICMR'21	ResNet-12	68.42 ± 0.51	81.29 ± 0.59	77.03 ± 0.85	87.20 ± 0.70
	SVAE-Proto [58]	CVPR'22	ResNet-12	74.84 ± 0.23	83.28 ± 0.40	76.98 ± 0.65	85.77 ± 0.50
	SP-CLIP [5]	CVPR'23	Visformer-T	72.31 ± 0.40	83.42 ± 0.30	78.03 ± 0.46	88.55 ± 0.32
	SemFew	Ours	ResNet-12	77.63 ± 0.63	83.04 ± 0.48	78.96 ± 0.80	85.88 ± 0.62
	SemFew-Trans	Ours	Swin-T	**78.94 ± 0.66**	**86.49 ± 0.50**	**82.37 ± 0.77**	**89.89 ± 0.52**

Table 1. Results (%) on MiniImageNet and TieredImageNet. The ± shows 95% confidence intervals. The best results are shown in **bold**.

Method	Venue	Backbone	CIFAR-FS		FC100	
			5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
ProtoNet [46]	NeurIPS'17	ResNet-12	72.20 ± 0.70	83.50 ± 0.50	41.54 ± 0.76	57.08 ± 0.76
TADAM [37]	NeurIPS'18	ResNet-12	-	-	40.10 ± 0.40	56.10 ± 0.40
MetaOptNet [27]	CVPR'19	ResNet-12	72.80 ± 0.70	84.30 ± 0.50	47.20 ± 0.60	55.50 ± 0.60
MABAS [23]	ECCV'20	ResNet-12	73.51 ± 0.92	85.65 ± 0.65	42.31 ± 0.75	58.16 ± 0.78
RFS [49]	ECCV'20	ResNet-12	71.50 ± 0.80	86.00 ± 0.50	42.60 ± 0.70	59.10 ± 0.60
SUN [11]	ECCV'22	ViT-S	78.37 ± 0.46	88.84 ± 0.32	-	-
FewTURE [19]	NeurIPS'22	Swin-T	77.76 ± 0.81	88.90 ± 0.59	47.68 ± 0.78	63.81 ± 0.75
Meta-AdaM [45]	NeurIPS'23	ResNet-12	-	-	41.12 ± 0.49	56.14 ± 0.49
SP-CLIP [5]	CVPR'23	Visformer-T	82.18 ± 0.40	88.24 ± 0.32	48.53 ± 0.38	61.55 ± 0.41
SemFew	Ours	ResNet-12	83.65 ± 0.70	87.66 ± 0.60	**54.36 ± 0.71**	62.79 ± 0.74
SemFew-Trans	Ours	Swin-T	**84.34 ± 0.67**	**89.11 ± 0.54**	54.27 ± 0.77	**65.02 ± 0.72**

Table 2. Results (%) on CIFAR-FS and FC100. The ± shows 95% confidence intervals. The best results are shown in **bold**.

Experimental results

- Cross-Domain Few-Shot Learning

Method	Venue	CUB		Places	
		1-shot	5-shot	1-shot	5-shot
GNN [43]	ICLR'18	45.69	62.25	53.10	70.84
S2M2 [32]	WACV'20	48.24	70.44	-	-
FT [51]	ICLR'20	47.47	66.98	55.77	73.94
ATA [55]	IJCAI'21	45.00	66.22	53.57	75.48
AFA [21]	ECCV'22	46.86	68.25	54.04	76.21
StyleAdv [14]	CVPR'23	48.49	68.72	58.58	**77.73**
LDP-net [64]	CVPR'23	49.82	70.39	53.82	72.90
SemFew-Name	Ours	57.58	72.26	63.22	74.54
SemFew	Ours	**59.07**	**72.47**	**64.01**	74.70

Table 3. Average results (%) on cross-domain scenarios. SemFew-Name denotes that semantics are class names.

Experimental results

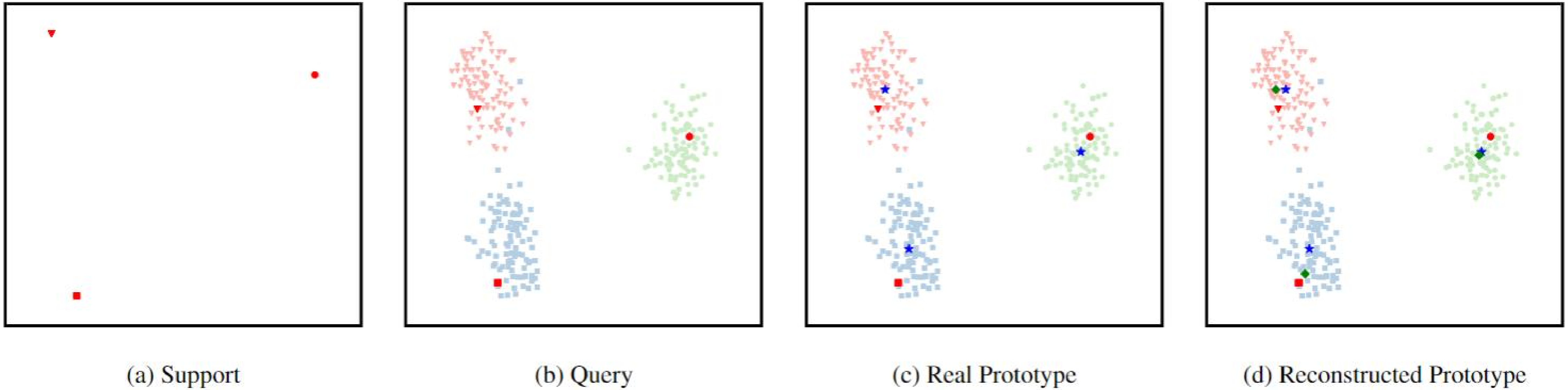


Figure 5. Visualization results on the MiniImageNet dataset. Different colors or shapes represent different classes. The $\star$ represents the class prototypes, and the $\diamond$ denotes the prototypes reconstructed by our method.

Summary

- We propose a method for automatically collecting high-quality semantics and applying them in few-shot learning.
- We design a simple and efficient way to translate high-quality semantics and visual features into prototypes, without any intricate semantic understanding modules.
- Our approach achieves competitive performance across six benchmarks in FSL research, underscoring that a basic network can obtain excellent performance when supported by high-quality semantics.