

Rank-guided Diffusion for Noise Few-Shot Learning

Zelei Wu*, Kun Zhou*, Xulun Ye, Yifan Mei, Jie Hong, Jieyu Zhao

ICML 2026 · Few-Shot Learning · Robust Representation Learning

Noisy Support Sets

Low-Rank Structure

Guided Diffusion

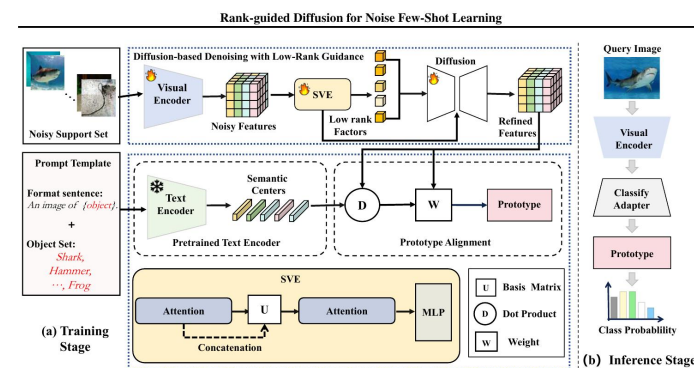
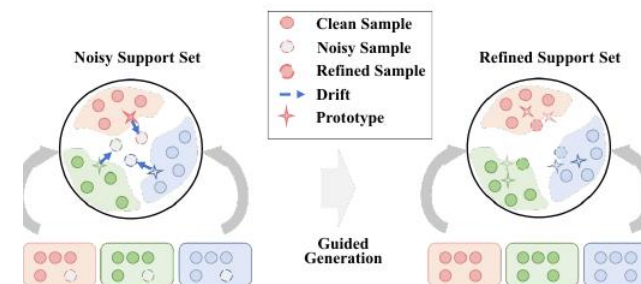
Robust Prototypes

Core message

Instead of discarding noisy few-shot samples, CRDProto detects rank anomalies and repairs corrupted support features with low-rank guided diffusion.

Why it matters: in 5-way 5-shot learning, even one mislabeled support image can shift the prototype and corrupt all query predictions.

Learning



Structure-aware correction from noisy support set to refined prototypes

Motivation: noisy support sets break few-shot learning

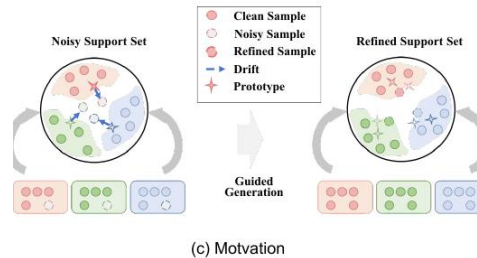
Few-shot models have very little redundancy: one noisy image can dominate a class prototype.

1 Support noise is inevitable

- Support sets may come from web data, weak supervision, or non-expert annotation.
- Few-shot episodes lack enough samples to average out corrupted labels.
- Existing methods often assume clean support sets or rely on large-scale statistics.

2 Prototype drift

Learning



A noisy support sample pulls the class prototype away from the true class center, shifting the decision boundary.

3 Structural prior

Key observation

Clean semantic features tend to concentrate in approximately low-rank, class-specific subspaces; noisy samples introduce off-subspace rank anomalies.

This motivates correction rather than removal: preserve semantic diversity while suppressing off-subspace corruption.

Empirical evidence: CLIP features are strongly low-rank

Low-rank geometry becomes a useful signal for detecting noisy support samples.

Rank-guided Diffusion for Noise Few-Shot]

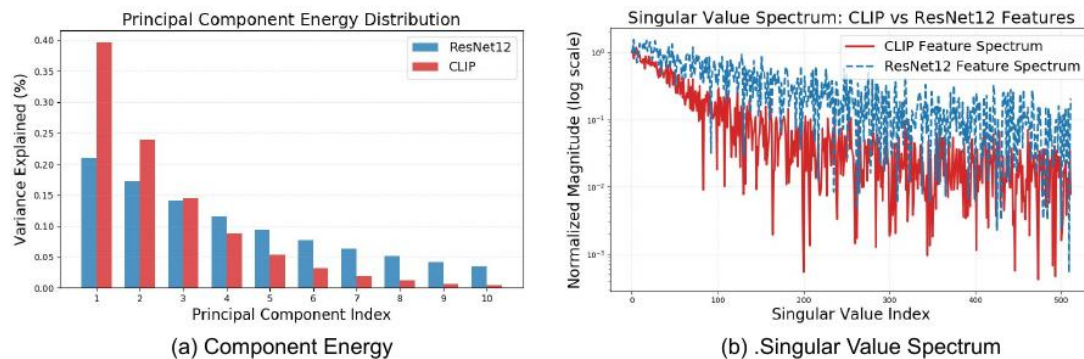


Figure 1 (a) Motivation: The solid colored circles represent clean samples in the sun

Principal energy + singular spectrum

Interpretation

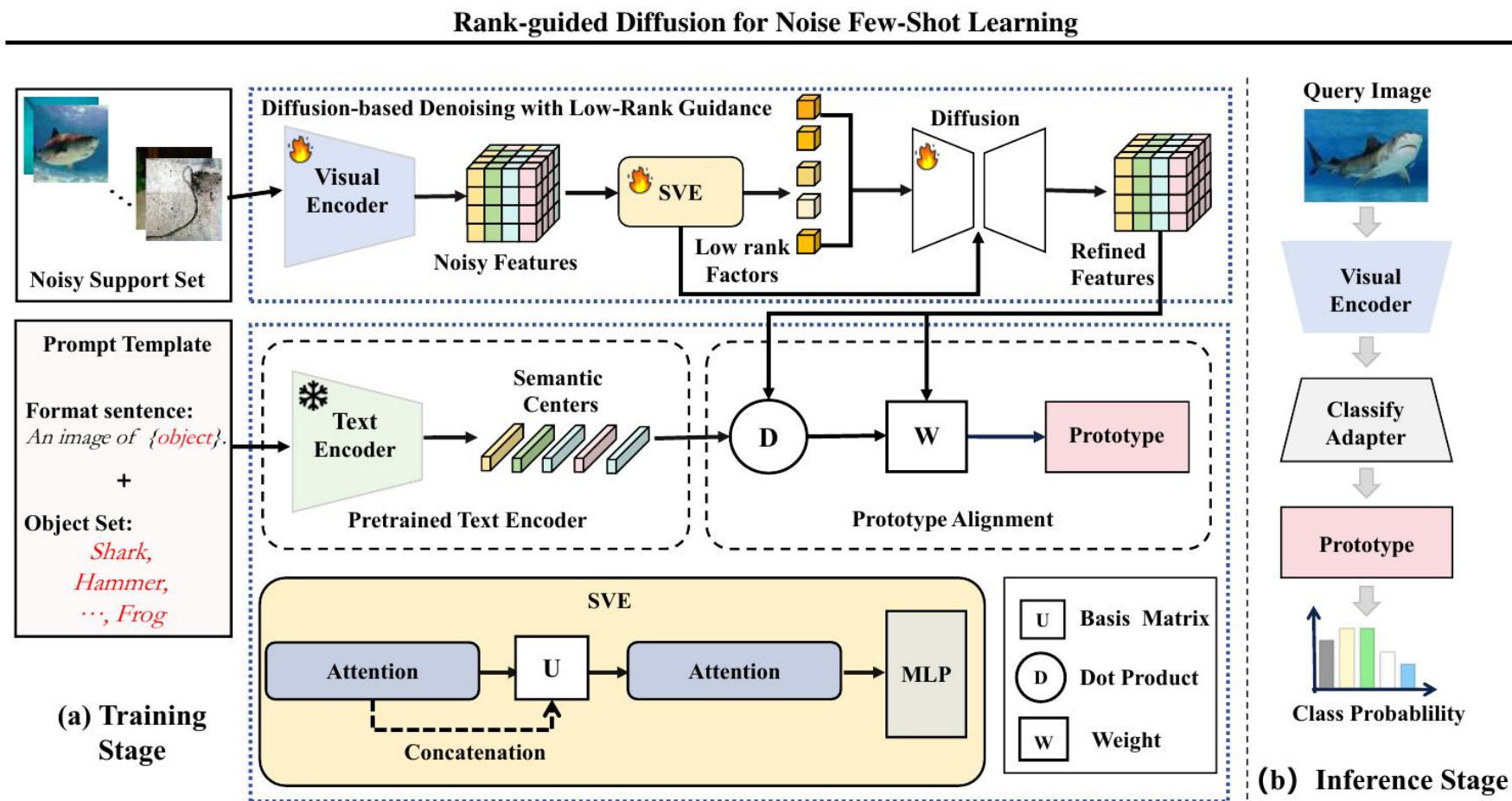
- Energy is concentrated in a few dominant components.
- A noisy feature adds energy in off-subspace directions.
- Projection residual reveals geometric inconsistency.
- Semantic similarity helps catch visually plausible but semantically wrong samples.

$$e_i = \|f_i - U_k U_k^T f_i\|_2^2$$

Large residual → likely noise

CRDProto: structure-guided support-set repair

Original framework: low-rank extraction, rank-guided diffusion, and semantic prototype alignment.



Step 1: noise detection via differentiable low-rank structure

Detect anomalies by combining geometric residual and semantic agreement.

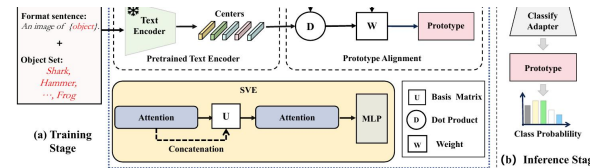
Class-wise feature matrix

For class c , construct feature matrix $X(c)$ from its K support features.

$$X(c) = L + S$$

L captures the dominant low-rank class structure; S captures sparse deviations caused by noise.

Differentiable SVE



Singular Value Embedding approximates the dominant subspace with smooth, differentiable operations, enabling end-to-end optimization.

Two complementary cues

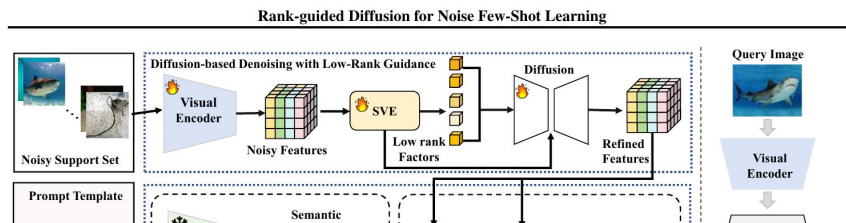
$$e_i = \|f_i - U_k U_k^T f_i\|_2^2$$

$$s_i = f_i^T y_c / (\|f_i\| \|y_c\|)$$

A sample is flagged when it violates both low-rank geometry and semantic consistency, avoiding over-reliance on a single cue.

Step 2: rank-guided diffusion repairs corrupted support features

The reverse process is biased toward class-specific low-rank subspaces.



$$R(z) = \|(I - P_U(c))z\|_2^2$$

The guidance term penalizes only the off-subspace component, so the sample moves back toward class geometry while preserving useful in-subspace variation.

Why diffusion instead of direct deletion?

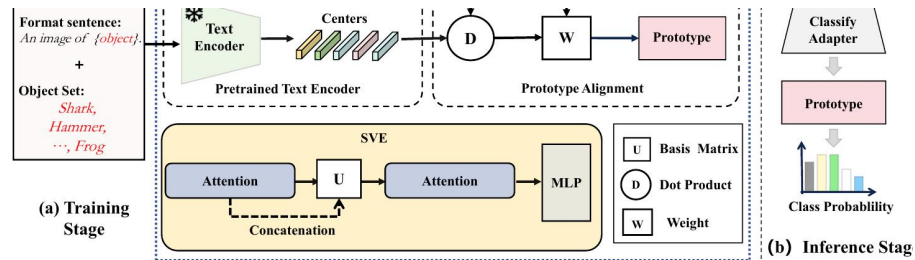
- 1 Deleting samples**
throws away valuable few-shot information.
- 2 Hard projection**
may collapse semantic diversity.
- 3 Rank-guided diffusion**
gradually reconstructs a clean, class-consistent substitute.

Result: refined support set with improved intra-class coherence

Step 3: semantic-weighted prototype estimation

Final classification balances refined visual evidence and semantic centers.

Prototype alignment



The prototype is not simply an arithmetic mean. Samples are weighted by agreement with both the refined class geometry and semantic prior.

Weighted MAP estimate

$$p_{\text{final}} = (\beta \sum_i w_i f_i + \gamma C_s) / (\beta \sum_i w_i + \gamma)$$

Weights down-weight residual outliers after diffusion:

$$w_i \propto \exp(-\alpha e_i) \cdot \exp(\delta s_i)$$

- β controls support-feature evidence.
- γ controls semantic-center prior strength.
- Large residual lowers sample weight; high semantic agreement raises it.

Experiments: robust few-shot learning under support noise

Evaluation focuses on severe label-swap and outlier noise under 5-way 5-shot settings.

Setup

- 5-way 5-shot few-shot classification.
- MiniImageNet and TieredImageNet main benchmarks.
- Additional evaluation on CUB-FS and CIFAR-FS.
- Noise ratios: 0%, 20%, 40%, 60%.
- Backbones: ResNet12, CLIP ViT-B/32, DINO ViT-S.

Performance under label-swap noise

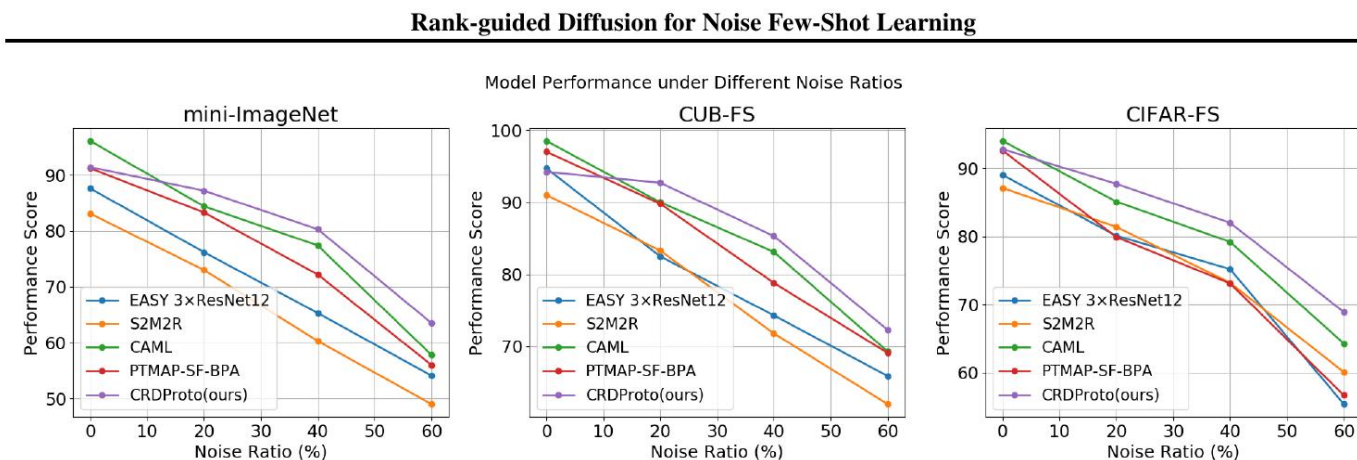


Figure 3. Experimental comparison charts of our 5-way 5-shot experiments on mini-ImageNet, CIFAR-FS, and CUB-FS under varying

CRDProto keeps the best degradation curve across datasets: accuracy remains higher as noise ratio increases.

Main results: strong accuracy under both noise types

CRDProto remains robust even when 40–60% of support samples are corrupted.

Label-swap noise

90.61

MiniImageNet / 0%

91.13

TieredImageNet / 0%

80.02

MiniImageNet / 40%

81.39

TieredImageNet / 40%

63.77

MiniImageNet / 60%

65.43

TieredImageNet / 60%

CRDProto + ViT-B/32 (CLIP)

Outlier noise

87.86

MiniImageNet / 0%

88.38

TieredImageNet / 0%

77.27

MiniImageNet / 40%

78.64

TieredImageNet / 40%

64.02

MiniImageNet / 60%

65.68

TieredImageNet / 60%

CRDProto + ViT-B/32 (CLIP)

Ablation confirms each component

Method	MiniImageNet (5w5s)	TieredImageNet (5w5s)	Noise Robust. (40% noise)	Feat. Consist. (Var. ↓)
ProtoNet	68.18 ± 0.16	71.42 ± 0.18	57.07 ± 0.20	4.5 ± 0.43
CLIP Detection Only	75.35 ± 0.15	75.61 ± 0.17	63.27 ± 0.18	3.41 ± 0.22
Diffusion (w/o Low-Rank)	70.84 ± 0.14	69.92 ± 0.16	62.05 ± 0.17	4.28 ± 0.23
Diffusion (w/ Low-Rank)	80.02 ± 0.20	81.39 ± 0.22	73.11 ± 0.19	3.29 ± 0.21
Mean Prototype (Unweighted)	78.33 ± 0.18	80.15 ± 0.20	70.22 ± 0.21	2.92 ± 0.22
Semantic-Weighted Prototype	90.61 ± 0.19	91.13 ± 0.21	81.39 ± 0.21	2.25 ± 0.31

Table 3. Ablation study under the 5-way 5-shot setting. Each component is incrementally added from the ProtoNet baseline. Low-rank regularization and the proposed semantic-weighted prototype notably enhance accuracy, noise robustness, and feature consistency on MiniImageNet and TieredImageNet.

Semantic-weighted prototype reaches 90.61 / 91.13 and reduces variance to 2.25.

Takeaways: low-rank geometry is a practical robustness prior

CRDProto turns noisy support features into a refined, structurally consistent support set.

Feature-space visualization

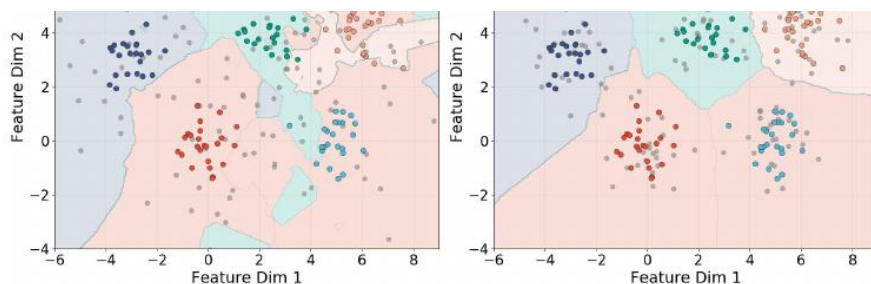


Figure 4. t-SNE visualization of feature distributions with and without low-rank guidance. (a) Without low-rank guidance ($\eta = 0$), the generated features show scattered distribution and poor structural consistency. (b) With low-rank guidance ($\eta > 0$), the features form compact, well-separated clusters that better preserve the intrinsic geometric structure of the clean support set.

Low-rank guidance yields compact and better-separated clusters.

What the audience should remember

1

Differentiable SVE

captures class-specific low-rank structure for noise detection.

2

Rank-guided diffusion

repairs corrupted support features instead of removing them.

3

Semantic-weighted prototypes

significantly improve robustness under severe support noise.

Code: github.com/wuzelei123/CRDProto