



Semantic-Anchored, Class Variance-Optimized Clustering for Robust Semi-Supervised Few-Shot Learning

ICLR 2026

GRaM Workshop

Souvik Maji*, Rhythm Baghel*, Pratik Mazumder*

Indian Institute of Technology Jodhpur {b22cs089, b22cs042, pratikm}@iitj.ac.in

What Is Semi-Supervised FSL?

Goal: Learn to classify from very few labeled examples (K per class) by exploiting a large unlabelled pool \mathcal{U} .

Setup: Support $S = \{(x_i^s, y_i^s)\}_{i=1}^{N \times K}$, Query $Q = \{(x_i^q, y_i^q)\}_{i=1}^{N \times T}$, Unlabeled $U = \{x_i^u\}_{i=1}^{N \times u}$.

Prior Work Limitations

- **Geometric neglect:** Indiscriminate pseudo-labeling lets outliers corrupt the embedding manifold.
- **Rigid prototypes:** Single-centroid methods miss intra-class variability; break under distribution shift.
- **Semantic fragility:** Visual prototypes unanchored to class meaning; fragile with occluded or cluttered support.
- **Error amplification:** Pseudo-label accuracy $< 80\%$ compounds through the label-propagation graph.

Three-Stage Training Pipeline

- **Stage 1 — Pre-training (200 ep.):** Multi-task backbone training with classification + rotation prediction.
- **Stage 2 — Episodic Fine-tuning (200 ep.):** RandAugment + Embedding Propagation; CVOC (10 iters) and CST (1 iter) applied per episode.
- **Stage 3 — Semantic Network (100 ep.):** CLIP text encoder aligned with visual prototypes via 4-stage LLM description pipeline.

Benchmark Performance

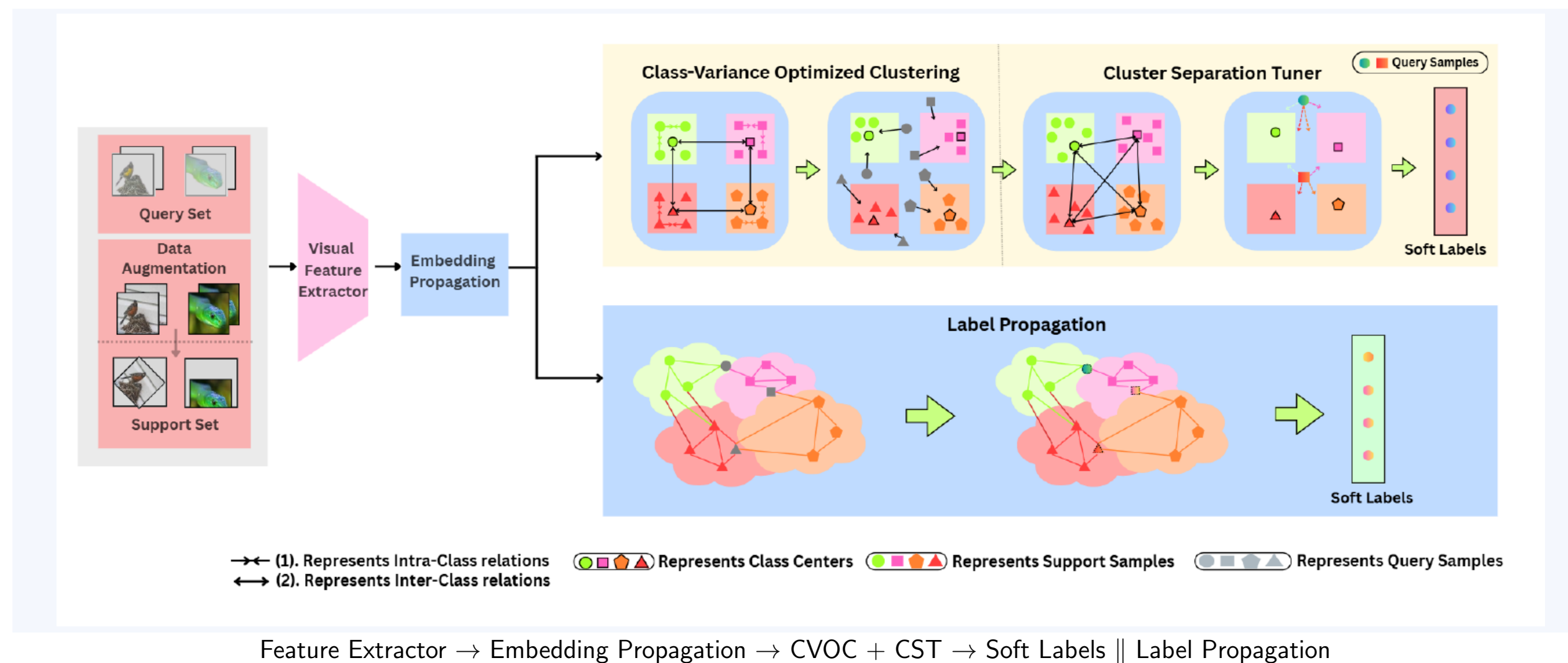
Dataset	1-shot	5-shot	Dataset	1-shot	5-shot
<i>miniImageNet</i> (R-12)			<i>tieredImageNet</i> (WRN)		
Cluster-FSL	77.81	85.55	Cluster-FSL	85.74	90.18
Ours	84.51	86.95	Ours	89.06	91.53
<i>miniImageNet</i> (WRN)			CUB-200 (R-12)		
Cluster-FSL	82.63	89.16	Cluster-FSL	87.36	92.17
Ours	85.94	91.03	Ours	88.63	93.37
<i>tieredImageNet</i> (R-12)			CUB-200 (WRN)		
Cluster-FSL	83.89	89.94	Cluster-FSL	91.80	95.07
Ours	85.55	91.13	Ours	93.15	96.25

Conclusions & Limitations

- **CVOC + CST** boosts pseudo-label accuracy by **+6.7%** over SOTA (1-shot).
- **Semantic Anchoring** via 4-stage LLM pipeline stabilises fragile prototypes in low-shot regimes.
- **Limitation:** Reduced efficacy for visually or semantically ambiguous classes.

Code: github.com/majisouvik26/SA-CVOC

Fine-Tuning Phase



Methodology

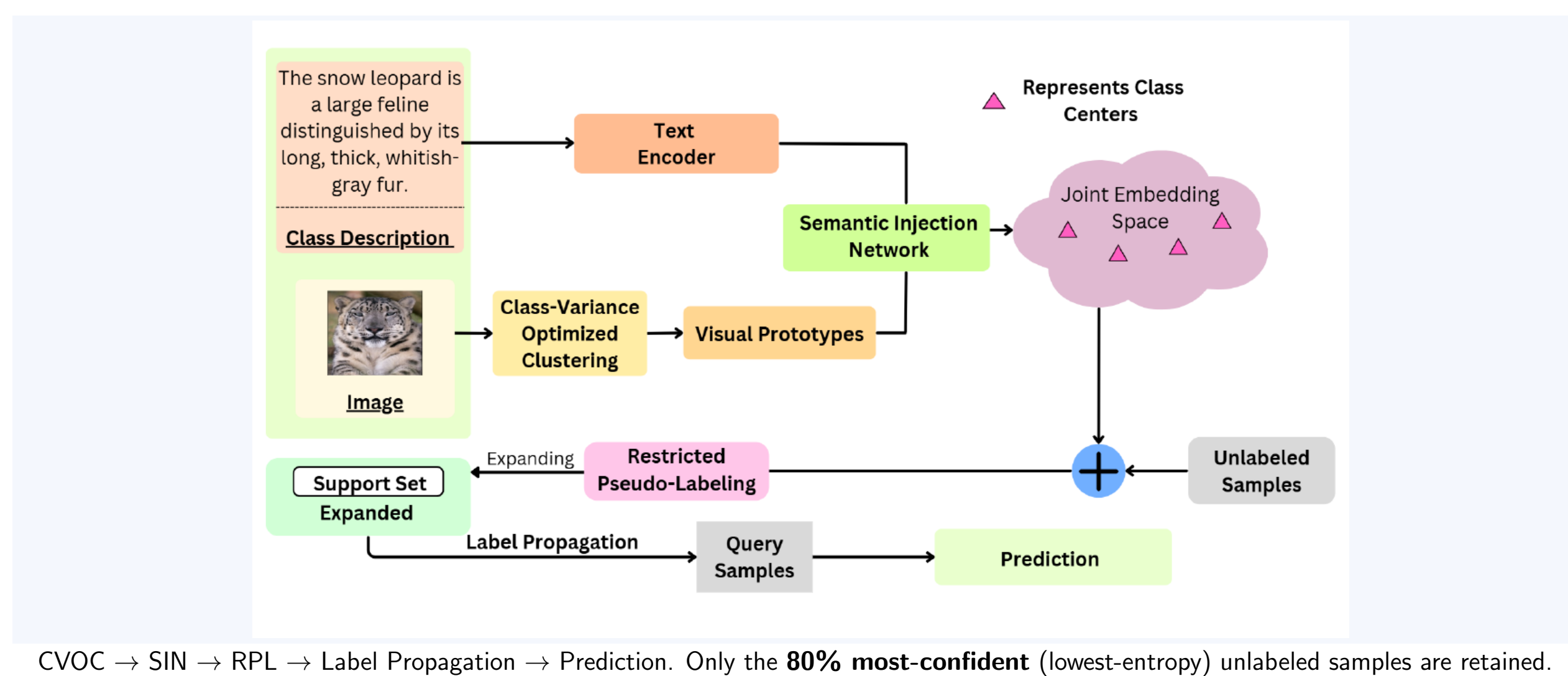
A. Class-Variance Optimized Clustering (CVOC)

Samples assigned to classes by minimising a combined distance that balances reconstruction residual, intra-class compactness, and inter-class repulsion, fusing local geometry with global separability. Runs **10 iterations** per episode.

B. Cluster Separation Tuner (CST)

Firefly-inspired refinement: prototypes scored by brightness (distance + proximity penalty); dimmer prototypes migrate toward brighter neighbours.

Evaluation Strategy: Restricted Pseudo-Labeling (RPL)



Semantic Injection Network (SIN)

Visual prototypes are nudged toward CLIP text embeddings. Class descriptions generated by a **4-stage LLM chain** (Writer → Filter → Visualizer → Finalizer, GPT-4o-mini).

