A novel reliability aware patch selection based few shot learning for target recognition and open-set identification in ill posed SAR images

Anonymous WACV Algorithms Track submission

Paper ID *****

Abstract

Earth observation with synthetic aperture radar (SAR) imagery has recently seen an extensive use of deep learning for target recognition. However, SAR remains ill posed by scarce labeled data, severe class imbalance and the appearance of open-set targets. The proposed work address these challenges through a patch wise manifold learning framework that integrates a residual multi-scale contextual encoder with entropy guided patch selection producing structurally reliable features. On this discriminative features the model optimizes instance level contrastive learning, prototype based clustering and an imbalance aware classification loss yielding compact intraclass features and interclass separation. Open-set identification is achieved through a distance criterion that identifies queries with excessive prototype deviation enabling unknown target detection. Experiments on the MSTAR and SAMPLE benchmark dataset demonstrate state-of-the-art performance alongside substantially improved identification of previously unseen targets.

1. Introduction

A recent technical leap in the field of modern Earth observation has emerged as Synthetic Aperture Radar (SAR) images [12]. It is a transformation extending far beyond the optical images wherein active sensors transmit electromagnetic pulses and record the energy reflected from the Earth's surface. As such, these reflected backscatter signatures generate SAR imagery regardless of weather or daylight conditions [12, 18]. SAR images have become indispensable for applications ranging from environmental monitoring and disaster assessment to defense applications, maritime safety and urban planning [3, 12].

SAR images are shaped by interactions between the transmitted radar signal and surface geometry aided by the illumination angle. These influences enhance SAR images with rich structural features distinguishing them from their optical counterparts yet introducing challenges in interpretation [12, 18]. Speckle noise pervades SAR data and complicates segmentation and classification tasks [18]. Moreover, data heterogeneity caused by variations in incidence angle and surface structure leads to high intra-class diversity hindering the effectiveness of conventional deep learning models.

A significant challenge in SAR image analysis is the scarcity of accurately labeled datasets owing to the complexities inherent in SAR data (lack of specialized knowledge for annotation, privacy and security restrictions and the high cost of data acquisition) [1]. It results in lack of variability representation encountered in operational scenarios and subsequently limiting generalization [2]. This scarcity of labeled data has driven the adoption of semi-supervision, few shot learning and synthetic data generation strategies to supplement limited training samples [2].

Further, SAR-based machine learning pipelines are frequently grappled with the persistent problem of class imbalance presenting a nonuniform distribution of target categories [2, 15]. These imbalance means that deep learning models, when trained on such data, tend to overfit the majority classes. This develops representational bias for underrepresented minority categories compromising target recognition. Conventional mitigation strategies, such as oversampling, class-rebalancing heuristics and synthetic minority generation offer only partial relief [2, 15]. They often amplify noise and escalate intra-class variability, while failing to address the heterogeneous minority samples [2]. The problem of class imbalance is further magnified when SAR recognition models must function in operational settings with limited supervision [15].

Compounding this challenge is the fundamentally openset nature of operational SAR analysis. The datasets assume a closed set of object categories, but reality regularly encounters new, never-seen SAR classes [4, 6]. Models trained on closed-set images misclassify these unknown samples implicitly expecting that every input belongs to one of the pre-defined categories [4, 6]. This systematic misalignment creates genuine risk in critical domains, such as security monitoring or marine anomaly detection, where failing to recognize novelty can result in threats and operational errors [3, 12].

In response to these interconnected challenges, the SAR research community has seen the emergence of advanced deep learning techniques[18]. Among these methods contrastive learning improves few-shot generalization under scarce labeled data while intra-class diversity pulls together semantically similar SAR targets and pushes apart confusing negatives in the embedding space [11]. Prototype-based classification strengthens open-set and few-shot recognition by forming compact class centers that support distance discrimination between known and unseen targets [11]. Attention mechanisms mitigate clutter and visual variability by reweighting features so that the network focuses on salient and contextual structures [17]. Multi-scale feature extraction, meanwhile, has become essential as it contains both textural patterns and fine-grained spatial information[6, 17].

Despite their individual merits, these strategies have rarely been synthesized into a unified framework that can simultaneously confront class imbalance and deliver reliable open-set recognition. The proposed system combines a residual convolutional backbone capable of deep structural feature extraction with an adaptive multi-scale context head. It aims to enrich intermediate representations via a novel entropy guided patch selection and associated heterogeneous feature understanding. The structurally reliable patches from all categories are emphasized while redundant majority class responses are suppressed allowing minority class features to contribute proportionally to prototype construction. The loss function cumulates contrastive, cluster-wise, and imbalance-aware classification for organizing the feature space into compact and well separated clusters guided by discriminative class prototypes. A dependency matrix of the internal channels further enhance the protoptype of each class. A distance based open-set recognition mechanism does not rely on auxiliary support samples and leverages the learned embedding space. The framework reliably identifies samples that do not conform to any known category and flags them as potential unknowns. Underlining this work is the systematic extraction of feature representations that remain robust and generalize effectively even as SAR images remain ill posed (scarce labels, imbalance categories and open-set targets) while offering three key novelties; (a) A novel reliability aware patch learning scheme in SAR through an entropy guided patch selection to ensure that only discriminative, imbalance resilient and structurally stable regions contribute to feature embedding, (b) A prototype construction strategy on a dependency matrix manifold encoding channel wise relational structures into enhancing intra-class consistency and robustness, and (c) A novel unified episodic learning that integrates instance-level contrastive alignment, prototype-driven clustering and an imbalance aware classification to jointly address label scarcity and class imbalance with a distance based open-set identification.

2. Related work

Research in target recognition of synthetic aperture radar (SAR) imagery has grown rapidly in recent years owing to the advances in deep neural networks. Much of this progress, however, continues to rest on assumptions of balanced class distributions, fully closed label spaces and abundant training samples.

2.1. Scarcity of labeled SAR data

The deployment of deep learning in SAR imagery is constrained by the scarcity of labeled datasets owing to annotation issues. Public SAR benchmarks remain small compared to optical-image datasets. However reviews of SAR deep learning highlight that many early systems implicitly assume abundant and balanced labeled data. Few-shot learning has therefore become central aiming to learn transferable representations that adapt to new classes with only a few labeled examples [13]. Meta-learning approaches train across multiple episodes and tasks so that the model can quickly adapt to new categories under limited supervision[9, 16]. Graph and relation based few-shot methods, such as position aware graph neural networks, model structural relationships between support and query samples to improve classification in low data regimes [17]. Contrastive and prototype centered SAR frameworks leverage metric learning, multi-level contrastive cues and sample mining to pull same class together while pushing different classes apart [11, 14].

2.2. Class imbalance in SAR target recognition

The scarcity of annotations is further compounded by class imbalance, where common target classes dominate while rare but critical categories remain severely under-represented [1, 8]. Constraint- and optimization-based approaches under imbalanced categories explicitly reweight gradients or impose class-wise constraints to better preserve minority-class decision boundaries. They typically act only at the loss level and do not control which image regions drive prototype formation [8]. Data-centric strategies such as generative augmentation or entropy-based resampling attempt to rebalance sample distributions by synthesizing or selectively replicating minority examples while inadvertently amplifying noise or distort the underlying features [2]. Despite these advances, most existing methods focus primarily on representation transfer for novel classes and do not explicitly couple within-image patch level reliability assessment for imbalance-aware prototype construction. This allows unimportant features to overfit the reliable ones.

2.3. Open-Set recognition (OSR) in SAR

Operational SAR imaging face fundamentally open-set conditions (presence of unseen targets) due to environmental changes and anomalous events [12]. Conventional SAR models trained under closed-set assumptions misclassify these unseen classes into known categories. Threshold-free open-set learning networks address this by casting OSR as an additional class problem, thereby avoiding manually tuned rejection thresholds [6]. Capsule network models extract pose embeddings and use divergence techniques to separate known and unknown categories [4]. Approaches also emphasize the importance of multi-scale feature extraction to exploit both fine grained and contextual structure in SAR targets [6]. Recent work further integrates open-set recognition with incremental learning using structural similarity based feature extraction and decision models to progressively incorporate new targets without retraining from scratch [7]. However, most existing methods either rely on synthetic unknowns or heavy multitask supervision where unreliable patches and biased prototypes can blur the boundary between known and unknown classes. Unified method of tackling this fundamental problem posed by unknown classes emerging at test time in limited labeled scenario with unbalanced category distribution still lacks enough attention.

3. Methodology

The proposed framework enables discriminative and open-set aware few-shot SAR target recognition under class imbalance.

3.1. Problem Definition

Let the training dataset be denoted as $D_{base} = \{(x_i, y_i)\}_{i=1}^{M}$. Here, x_i denotes SAR image assigned to a class y_i from a set of known n categories $\mathcal{C}_{base} = \{c_1, c_2, \ldots, c_n\}$. In real SAR scenarios, the number of samples per category $|D(c)|$ is highly imbalanced. The degree of imbalance i_d is quantified as the ratio of the number of samples in the category with minimum number of samples to the category with maximum number of samples.

$$i_d = \frac{|D(c_{min})|}{|D(c_{max})|} \tag{1}$$

where, $|D(c_{min})|$ is the number of samples of the rarest class and $|D(c_{max})|$ is the number of the samples of the most-occurring class.

During testing, the model encounters two recognition environmental settings as explained below.

1) Closed set classification: A novel set D_{new} containing few labeled samples and sufficiently large number of unlabelled samples (representing labeled data sparsity). The labelled set forms the support set S_{new}, while the unlabelled sample belonging forms the query set Q_{new}.

$$D_{new} = S_{new} \cup Q_{new}, \tag{2}$$

The set of the classes in the novel set D_{new} is denoted as C_{new} , while ensuring $\mathcal{C}_{base} \cap \mathcal{C}_{new} = \varnothing$. The objective is to classify each query $x \in Q_{new}$ by leveraging S_{new} and the features of D_{base}.

2) Open set recognition: The classification of known classes should be accompanied by the identification of samples from completely unknown categories (without any labels). The label space can thus be defined as $\mathcal{C}_{new} \cup \{-1\}$, where the label "$-1$" denotes unknown classes. For a given test query sample x, the model is required to determine if it generalize effectively to the novel target categories or reliably identify unknown targets in open-set conditions.

3.2. Overview

As illustrated in Fig. 1, (i) meta-training is performed on the imbalanced base set D_{base} using episodic sampling of support and query data. Each SAR image and (a) its augmented views are processed through (b) a residual encoder. This is followed by (c) patchification of the features and (d) an entropy-based filtering stage that retains only reliable regions for feature embedding. In addition to that (e) global features are also extracted to maintain global semantics and are (f) aggregated with the patchified reliable feature embeddings. The (g) class embeddings through (h) a dependence matrix manifold finally forms (i) the class prototypes. A novel multi-level loss (instance contrastive, prototype clustering, imbalance-aware classification) jointly optimize (j) distance based discriminative representation learning . During (ii) meta-testing, the trained encoder and prototype construction pipeline are kept fixed. Query samples (from the scarcely labelled samples) are classified via distance-guided prototype matching, and an unknown identification threshold enables accurate identification of unknown targets in open-set SAR scenarios.

3.3. Meta-training

The meta-training stage is designed for episodic learning that mimic few-shot classification with imbalanced and open-set conditions. Each episode is constructed from the base dataset D_{base} by sampling a small set of classes and partitioning their examples into support and query subsets. Within each episode, every support and query image is first processed by the residual contextual encoder aided by patch-wise decomposition, entropy-based reliability filtering and multi-scale fusion of global and local descriptors. The resulting feature vector provides the basis for learning of instance-level contrastive alignment, manifold prototype regularization and imbalance aware classification. A

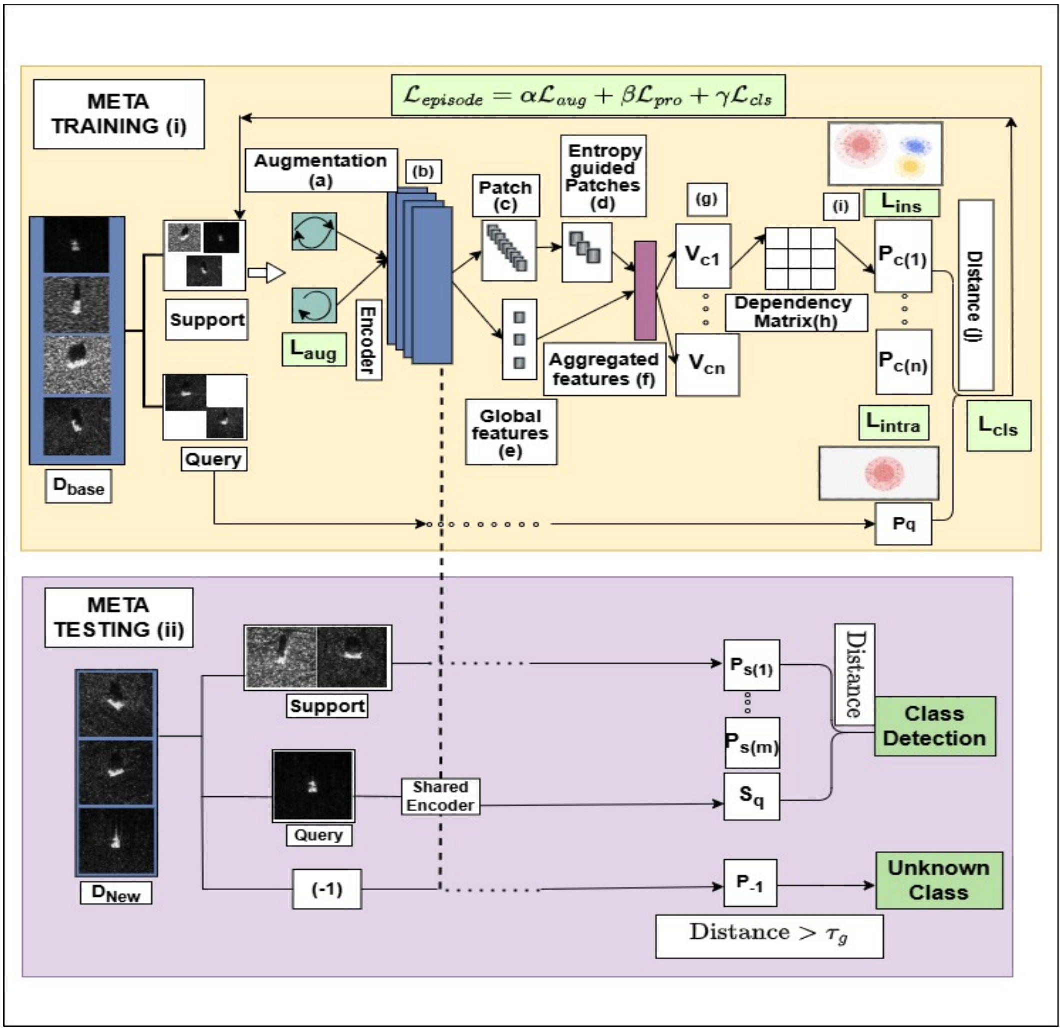

Figure 1. Overall architecture of the proposed framework.

prototype driven clustering loss then pulls query descriptors toward the prototype of their support class and pushes them away from other class prototypes. Joint optimization of these objectives guides the encoder and the prototyping module toward feature representations that are simultaneously few-shot transferable and resilient to imbalance.

3.3.1. Residual contextual feature encoding

The first phase of the proposed framework is the efficient feature extraction from the base set D_{base}. Each episode begins with considering a subset of the base class as a support set and query set. Given an image x from the support set, a geometric augmentation $t(\cdot)$ produces a variant $x' = t(x)$ that preserves semantic structure while maintaining stability under SAR imaging variations. Both x and x' are processed through a shared residual convolutional network parameterized by θ, yielding deep spatial feature maps

$$f_\theta(x), \quad f_\theta(x') \in \mathbb{R}^{H\times W\times C}. \tag{3}$$

In an attempt to capture discriminative structures inherent to each SAR image, the feature map is partitioned into $G \times G$ non-overlapping spatial regions.

$$f_\theta(x) = \{p_i\}_{i=1}^{G^2}, \tag{4}$$

where each $p_i \in \mathbb{R}^{h\times w\times C}$ denotes a localized patch. These patches are converted to patch-wise vectors through a transformation function $\phi(\cdot)$ as shown below.

$$\mathbf{z}_i = \phi(p_i) \in \mathbb{R}^d. \tag{5}$$

Each vector $\mathbf{z}_i \in \mathbb{R}^d$ is first normalized into an activation $\mathbf{a}_i(k)$ using a softmax function.

$$\mathbf{a}_i(k) = \frac{\exp\big(\mathbf{z}_i(k)\big)}{\sum_{k=1}^{d}\exp\big(\mathbf{z}_i(u)\big)}, \quad k = 1, 2, \ldots, d, \tag{6}$$

where d is number of feature channels in each patch p_i. The uncertainty of patch p_i is quantified using Shannon entropy as shown below.

$$E_i = -\sum_{k=1}^{d} \mathbf{a}_i(k)\, \log \mathbf{a}_i(k), \tag{7}$$

A binary mask is applied to suppress unreliable patches such that

$$m_i = \begin{cases} 1, & E_i < \tau, \\ 0, & E_i \geq \tau, \end{cases} \tag{8}$$

This yields the filtered descriptor.

$$\tilde{\mathbf{z}}_i = m_i\, \mathbf{z}_i. \tag{9}$$

Only the set of low-uncertainty descriptors $\tilde{\mathcal{Z}} = \{\tilde{\mathbf{z}}_i\}_{i=1}^{G^2}$ is retained, allowing the model to focus on structurally meaningful regions. While individual patches provide localized detail, the overall target structure is captured through a global descriptor obtained by spatial average pooling as shown below.

$$\mathbf{z}_g = \text{GlobalAvgPool}(f_\theta(x)) \in \mathbb{R}^d. \tag{10}$$

The multi-scale representation is then constructed by combining the global feature with the set of localized embeddings,

$$\mathbf{z}(x) = \psi\big(\mathbf{z}_g, \{\tilde{\mathbf{z}}_i\}_{i=1}^{G^2}\big), \tag{11}$$

where $\psi(\cdot)$ denotes the fusion of global semantics and spatially selective features.

The encoder is trained using an instance-level contrastive (cross-entropy) loss using cosine similarity that aligns each sample with its augmented counterpart.

$$\mathcal{L}_{aug} = \text{CE}\big(\text{cossim}\big(\mathbf{z}(x), \mathbf{z}(x')\big), Y\big) \tag{12}$$

Here, Y specifies positive pairings between each sample and its augmented view.

3.3.2. Prototype embedding

Following the obtaining of the aggregated descriptor $\mathbf{z}(x)$, the support samples belonging to the same class are first aligned in a shared space. A transformation $\eta(\cdot)$ is applied to generate embeddings as follows.

$$\mathbf{v}(x) = \eta(\mathbf{z}(x)) \in \mathbb{R}^d. \tag{13}$$

For each class c_i in the support set, the center of that class is estimated. This yields a prototype defined by the following equation.

$$\mathbf{v}_{c_i} = \frac{1}{|S_c|} \sum_{x \in S_{c_i}} \mathbf{v}(x). \tag{14}$$

where, S_{c_i} is the set of images belonging to class c_i. Instead of relying solely on the prototype vector $\mathbf{v}_{c_i}$, the internal relationships between feature channels are further modelled. Let $\mathbf{v}_{c_i} = [v_{c_i,1}, v_{c_i,2}, \ldots, v_{c_i,d}]$ denote the d-dimensional class prototype. All possible channel pairs (i,j) are compared through a similarity operator $\kappa(\cdot,\cdot)$ to generate $m_{ij} = \kappa(v_{c_i,i}, v_{c_i,j})$ and yield a symmetric dependency matrix.

$$\mathbf{M}_{c_i} = [m_{ij}] \in \mathbb{R}^{d\times d}, \qquad i,j = 1,\ldots,d. \tag{15}$$

Here, m_{ij} is the activation between channels i and j. The diagonal entries correspond to self-dependencies and are therefore less informative for discriminative purposes. Only the unique channel interactions are retained in each class prototype p_{c_i} by extracting its upper triangular portion.

$$p_{c_i} = Upper(\mathbf{M}_{c_i}), \tag{16}$$

This process produces a compact structural reference p_c for each class. During inference, a query image is processed similarly into $\mathbf{p}_q$, and recognition is performed by comparing its structural pattern with the stored prototype descriptors.

3.3.3. Episodic structural learning

During meta-training, episodic learning is employed to ensure that the learned features generalize effectively to novel classes with only a few labeled examples. In each training episode, a support set is used to construct class prototypes in the structural space as described previously, while query samples are used to optimize loss.

$\mathbf{p}_q$ denote the structural descriptor of a query example x_q, with y_q as its ground-truth class label. To encourage the query sample to be placed close to the prototype of its correct class p_{y_q}, a prototype-driven cluster discrimination loss is introduced to promote intra-class compactness.

$$\mathcal{L}_{\text{intra}} = -\log \frac{\exp\left(-d(\mathbf{p}_q, \mathbf{p}_{y_q})\right)}{\sum_{c_i \in \mathcal{C}_{\text{epi}}} \exp\left(-d(\mathbf{p}_q, \mathbf{p}_{c_i})\right)}, \tag{17}$$

where $d(\cdot,\cdot)$ denotes a distance metric and $\mathcal{C}_{\text{epi}}$ represents the set of classes in the current episode.

At the instance level, each query is further compared with the embeddings of all support samples ensured through a loss function formulated as follows.

$$\mathcal{L}_{\text{ins}} = \sum_{x_s \in S_{c_i}} \ell\big(d(\mathbf{p}_q, \mathbf{p}_{c_i}), \mathbb{I}(y_q = y_s)\big), \tag{18}$$

where y_s is the label of that particular class and $\ell(\cdot)$ pulls together same class pairs while pushing away cross class pairs. The overall episodic optimization is expressed as the following equation.

$$\mathcal{L}_{\text{pro}} = \eta\, \mathcal{L}_{\text{intra}} + \lambda\, \mathcal{L}_{\text{ins}}, \tag{19}$$

To alleviate class imbalance a classification constraint is introduced as

$$\text{Logits}_{c_i} = -d(\mathbf{p}_q, \mathbf{p}_{c_i}), \tag{20}$$

The imbalance-aware cross-entropy (CE) loss is then formulated as follows.

$$\mathcal{L}_{\text{cls}} = \text{CE}(\text{Logits}, y_q, w_{c_i}). \tag{21}$$

where, w_{c_i} is a class-frequency dependent weight that assigns higher importance to rare categories. Finally, the meta-training optimization function can be expressed as follows.

$$\mathcal{L}_{\text{episode}} = \alpha\, \mathcal{L}_{\text{aug}} + \beta\, \mathcal{L}_{\text{pro}} + \gamma\, \mathcal{L}_{\text{cls}}, \tag{22}$$

where $\alpha, \beta,$ andγ modulates the contribution of the classification constraint. This episodic objective ensures that the structural prototype space remains well-separated across classes, while also resilient to imbalanced sample availability in SAR target datasets.

3.4. Meta testing and distance based open set recognition

During meta-testing, the model is evaluated in an open-set setting where two types of classes may appear; novel classes that provide only a few labeled support samples and unknown classes that provide no support samples at all.

Given a query SAR image q, its structural embedding $\mathbf{s}(q)$ is computed using the same feature encoder and manifold established during meta-training. For classes $s_i \in C_{new}$ with known prototypes of the support set $\mathcal{S}_{new}$ denoted as p_{s_i} (done as in meta-testing with frozen encoder), the recognition of the class c^* of the query q follows a nearest-prototype decision rule.

$$c^* = \arg \min_{s_i \in \mathcal{C}_{new}} d(\mathbf{s}(q), \mathbf{p}_{s_i})\,, \tag{23}$$

The open-set scenario demands the ability to reject samples that do not belong to any known category. To achieve this, the minimum prototype distance is compared against a heuristically measured acceptance threshold τ_{global}

$$\hat{y}(q) = \begin{cases} c^*, & \text{if } \min_{s_i} d(\mathbf{s}(q), \mathbf{p}_{s_i}) \leq \tau_{global}, \\ -1, & \text{otherwise.} \end{cases}$$

Here, the label -1 denotes a prediction into the "unknown" class, representing targets that were not observed during training.

4. Results and Discussion

A comprehensive experimental evaluation of the proposed framework has been made on the benchmark MSTAR and SAMPLE datasets. At the onset an ablation study has been conducted to validate the necessity of each architectural component. It is followed by a quantitative comparisons of the proposed work against state-of-the-art few-shot learning benchmarks. Finally, the proposed method's robustness has been validated under open-set condition.

4.1. Datasets

The proposed framework is evaluated using two SAR benchmark MSTAR [10] and SAMPLE [5] datasets. MSTAR consists of ten military vehicle categories imaged using X-band radar. Seven of the categories (BMP2, BTR70, T72, BTR60, D7, T62, and ZIL131) have been assigned as base classes, while the remaining three (2S1, BRDM2 and ZSU234) are set as new classes. To maintain imbalance the training samples are varied from a range of 150-520 samples per class as shown in Fig. 2. For the open set scenario, ZSU234 has been selected as the unseen class. The SAMPLE dataset comprises of ten types of vehicle targets derived from CAD models. Seven classes has been designated as the base set (2S1, BMP2, M2, BTR70, M60, T72, and ZSU234). The remaining three categories (M548 M1, and M5) constitute the novel set while maintaining imbalance in the base set as shown in Fig. 2. M1 accounts to be the unseen class in the open set scenario.

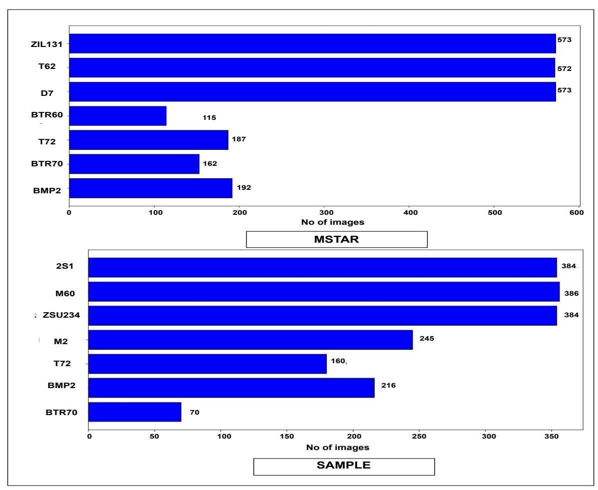

Figure 2. Class imbalance in the datasets

.

4.2. Experimental Settings

The proposed model is trained under a meta-learning paradigm using 3-way K-shot classification tasks, with $K \in \{1, 3, 5, 7\}$. During meta-training, each episode contains 3 classes, 5 support images per class, and 15 query images in total. Novel-class support sets follow the same configuration during meta-testing, whereas unknown classes contribute only unlabeled query chips, reflecting realistic open-set requirements. A class imbalance ratio of approximately $1:5$ is maintained among the base classes during training.

A residual convolutional encoder with approximately 3.2M trainable parameters is used for feature extraction, producing $d = 256$-dimensional patch embeddings. Patchification divides each feature map into $G = 4$ regions. A softmax entropy threshold, experimentally fixed at $\tau = 0.65$, filters unreliable patches. While for open-set rejection, the global threshold $\tau_{\text{global}} = \mu_d + 3\sigma_d$ is computed using the mean μ_d and standard deviation σ_d of the support prototype distance distribution. Training is performed using Adam optimizer with learning rate 1×10^{-4} and weight decay 5×10^{-4}. Each model is trained for 1000 episodes with batch size 1 episode. This experimental configuration replicates real SAR target recognition where data is imbalanced, annotated samples are scarce, and unseen targets must be reliably rejected.

4.3. Ablation Studies

To assess the independent contribution of each module within the proposed framework, ablation experiments under 3-way 1-shot and 3-way 5-shot configurations have been conducted. The results, summarized in Table 1, demonstrate that the absence of any core module leads to a measurable decline in classification accuracy.

Table 1. Ablation study performance under few-shot settings.

Absent module	3-way 1-shot (%)		3-way 5-shot (%)	
	MSTAR	SAMPLE	MSTAR	SAMPLE
Patchification	81.53	80.42	86.28	84.50
Entropy Filtering	84.39	82.56	88.25	86.54
Cluster Consistency Loss	86.61	84.59	90.16	88.67
Dependency Matrix	87.25	85.32	91.08	89.92
Proposed	**91.12**	**90.58**	**95.92**	**94.26**

The improved performance observed in the combined framework confirms the the functional necessity of each components for generating the discriminative feature representations necessary for the classification.

To provide a qualitative assessment of the learned representation space, t-SNE visualizations are generated for the structural embeddings under balanced, imbalanced, and open-set conditions. As shown in Fig. 3, the proposed work forms compact and well-separated clusters for known target categories strongly even for the imbalanced training set. For the unknown target, a separate sample cluster forms, however, with more outliers as compared to the previous settings. This confirms that the model effectively enhances both intra-class compactness and inter-class separability, supporting reliable open-set rejection.

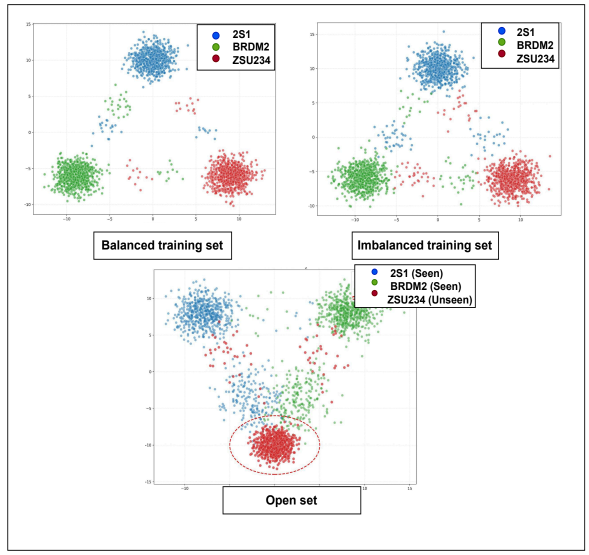

Figure 3. t-SNE visualization of structural embeddings learned under balanced, imbalanced, and open-set scenarios.

.

4.4. Comparison with state-of-the-art approaches

The proposed work has been compared with five state-of-the-art few-shot classification approaches. In each of the work imbalance has been introduced in the training with an imbalance degree of 1:5. The classification accuracy on the has been computed as the following equation

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of query samples}} \times 100\%. \tag{24}$$

The quantitative classification accuracies have been presented in Table 2.

Table 2. Benchmark comparison on imbalanced MSTAR and SAMPLE few-shot classification.

Method	Acc. (%) for 3-Way K-Shot Classification							
	MSTAR				SAMPLE			
	K=1	K=3	K=5	K=7	K=1	K=3	K=5	K=7
ConvT[13]	77.76	79.78	81.05	82.43	76.12	80.82	82.15	82.56
Meta-FSCIL[9]	81.09	83.34	83.82	84.54	78.05	81.15	82.83	86.79
Position-aware [17]	81.76	86.15	87.56	89.87	79.70	83.07	85.78	89.50
MHD-ProtoNet[14]	88.67	90.56	91.87	92.06	88.09	88.53	90.59	90.72
MCL-DMM[11]	90.79	92.17	93.05	93.97	89.73	91.55	91.82	92.75
Proposed	**91.12**	**94.84**	**95.92**	**96.24**	**90.58**	**92.95**	**94.26**	**94.89**

The results in Table 2 show that methods based on conventional convolutional backbones (ConvT [13]) struggle to maintain high recognition accuracy when faced with both data scarcity and class-frequency skew. Their performance remains consistently lower by $\approx 12-15\%$ across all shot settings and on both datasets as compared to the proposed method. The proposed method surpasses the generic meta-learning and graph-based methods, Meta-FSCIL[9] and Position-aware [17] by $\approx 6-10\%$. Prototype based baselines such as MHD-ProtoNet[14] and MCL-DMM[11] achieve stronger results, yet they still exhibit substantially lower recognition accuracy of ($\approx 2-4\%$) with the proposed method when introduced to imbalanced and limited support samples. Under such imbalanced scenarios, the compared methods suffer prototype deformation since the structural information becomes strongly biased by unimportant responses mainly in the imbalanced class. In contrast, the proposed method introduces an entropy-based patch filtering that discards uncertain local responses before prototype construction. Consequently, it achieves the highest accuracy in every configuration across both datasets. The proposed work achieves an accuracy of 91% for MSTAR and 90.58% for SAMPLE in the 1-shot case, while consistently maintaining superior accuracy for $K = 3, 5$, and 7.

These results demonstrate that combining discriminative structural embedding with reliability assessment enables the proposed work to extract more distinctive features than existing methods. This makes it significantly more resilient to class imbalance and scarce occurrences.

4.5. Open-Set Recognition Evaluation

The proposed framework's capability was next assesed for open-set recognition (OSR), which measure it's ability to accurately classify known targets while simultaneously identifying unknown, unobserved targets encountered during testing while maintaining the imbalance ratio. In each of the dataset, one class has been made unknown. The known class as well as unknown class accuracy is measured in a 3 way 3 shot setting. Table 3 presents the OSR results on the MSTAR and SAMPLE dataset.

Table 3. Open-set recognition performance comparison on MSTAR and SAMPLE.

Method	Known Class Acc. (%)		Unknown Class Acc. (%)	
	MSTAR	SAMPLE	MSTAR	SAMPLE
Incremental[7]	90.32	88.76	73.71	69.93
Threshold-free[6]	89.10	87.45	70.60	68.97
Capsule + KLD[4]	88.87	89.12	71.89	69.28
Proposed	**91.42**	**90.01**	**79.81**	**74.56**

The state-of-the-art OSR methods exhibit comparitively weaker identification capability due to their feature distributions placing both discriminative and unimportant features in the embedding space. The proposed work leverages its entropy-based filtering to discard unreliable or clutter regions that typically confuse similarity metrics. The dependence matrix prototype further captures channel wise relations unique to each known target category. As a result, unknown class embeddings naturally remain distant from all known prototypes enabling confident rejection. Consequently, the proposed method delivers reliable open-set behavior achieving 79.81% unknown class accuracy on MSTAR and 74.56% on SAMPLE This demonstrates an improvement of $6-11\%$ over the state-of-art OSR approaches. It further demonstrates superior known class classification accuracy by surpassing the state-of-the art OSR approaches by $\approx 2-3\%$. However, the known class accuracy is seen to decrease by 3% in MSTAR and 2% in SAMPLE in the proposed method with the introduction of open-set targets as compared to all known targets (Table 2).

5. Conclusion

The proposed method presents a novel few-shot SAR target recognition and open-set identification designed to operate effectively under class imbalance. By integrating residual contextual encoding, entropy-guided patch selection and dependence aware prototype learning, the proposed method generates discriminative and reliable representations even with limited imbalanced labeled samples. Extensive experiments conducted on the MSTAR and SAMPLE datasets demonstrate the proposed method's superiority in few-shot classification while maintaining substantially high rejection rate for open-set targets as compared to state-of-the-art approaches. Notably, the proposed approach preserves prototype integrity for rare target categories through the integration of entropy based reliable patch learning while also considering the internal dependence of each class's prototype channels. In future work, it is intended to extend the study to support multi-modal learning and to explore self-supervised pretraining for enhancing resilience against domain shifts.

References

[1] P. Chen et al. Class imbalance in the automatic interpretation of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2025. 1, 2

[2] N. Firoozy et al. Tackling sar imagery ship classification imbalance via deep convolutional generative adversarial network. *International Journal of Remote Sensing*, 2021. 1, 2

[3] Xudong Huang, Biao Zhang, William Perrie, Yingcheng Lu, and Chen Wang. A novel deep learning method for marine oil spill detection from satellite synthetic aperture radar imagery. *Marine Pollution Bulletin*, 179:113666, 2022. 1, 2

[4] Chunyun Jiang, Huiqiang Zhang, Ronghui Zhan, Wenyu Shu, and Jun Zhang. Open-set recognition model for SAR target based on capsule network with the KLD. *Remote Sensing*, 16(17):3141, 2024. 1, 3, 8

[5] Benjamin Lewis, Theresa Scarnati, Elizabeth Sudkamp, John Nehrbass, Stephen Rosencrantz, and Edmund Zelnio. A sar dataset for atr development: the synthetic and measured paired labeled experiment (sample). In *Algorithms for Synthetic Aperture Radar Imagery XXVI*, pages 39–54. SPIE, 2019. 6

[6] Yue Li, Haohao Ren, Xuelian Yu, Chengfa Zhang, Lin Zou, and Yun Zhou. Threshold-free open-set learning network for SAR automatic target recognition. *IEEE Sensors Journal*, 24 (5):6700–6710, 2024. 1, 2, 3, 8

[7] Xiaojie Ma, Kefeng Ji, Sijia Feng, Linbin Zhang, Boli Xiong, and Gangyao Kuang. Open set recognition with incremental learning for sar target classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023. 3, 8

[8] Yanjing Ma, Xing Zhang, Jifang Pei, Weibo Huo, Yin Zhang, Yuling Huang, and Jianyu Yang. A novel SAR target recognition approach under imbalanced categories: Constraint and optimization. In *IGARSS 2024 – 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 9644–9647, 2024. 2

[9] Hang Ran, Weijun Li, Lusi Li, Songsong Tian, Xin Ning, and Prayag Tiwari. Learning optimal inter-class margin adaptively for few-shot class-incremental learning via neural collapse-based meta-learning. *Information Processing & Management*, 61(3):103664, 2024. 2, 7, 8

[10] Timothy D Ross, Steven W Worrell, Vincent J Velten, John C Mossing, and Michael Lee Bryant. Standard sar atr evaluation experiments using the mstar public release data set. In *Algorithms for synthetic aperture radar imagery V*, pages 566–573. SPIE, 1998. 6

[11] Haoyue Tan, Zhenxi Zhang, Xiaoran Shi, Xinyao Yang, Yu Li, Xueru Bai, and Feng Zhou. Few-shot SAR ATR via multilevel contrastive learning and dependence matrix-based measurement. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:8175–8188, 2025. 2, 7, 8

[12] Arsenios Tsokas, Maciej Rysz, Panos M Pardalos, and Kathleen Dipple. Sar data applications in earth observation: An overview. *Expert Systems with Applications*, 205:117342, 2022. 1, 2, 3

[13] Chenwei Wang, Yulin Huang, Xiaoyu Liu, Jifang Pei, Yin Zhang, and Jianyu Yang. Global in local: A convolutional transformer for sar atr fsl. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. 2, 7

[14] Marii Zayani, Abdelmalek Toumi, and Ali Khalfallah. Mhd-protonet: Margin-aware hard example mining for sar few-shot learning via dual-loss optimization. *Algorithms*, 18(8): 519, 2025. 2, 7, 8

[15] Ying Zhang, Zhiyong Lei, Hui Yu, and Long Zhuang. Imbalanced high-resolution sar ship recognition method based on a lightweight cnn. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021. 1

[16] Li-Jun Zhao, Zhen-Duo Chen, Zi-Chao Zhang, Xin Luo, and Xin-Shun Xu. Bias mitigating few-shot class-incremental learning. *arXiv preprint*, 2024. 2

[17] Jia Zheng, Ming Li, Peng Zhang, Yan Wu, and Hongmeng Chen. Position-aware graph neural network for few-shot sar target classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:8028–8042, 2024. 2, 7, 8

[18] Xiao Xiang Zhu, Sina Montazeri, Mohsin Ali, Yuansheng Hua, Yuanyuan Wang, Lichao Mou, Yilei Shi, Feng Xu, and Richard Bamler. Deep learning meets sar: Concepts, models, pitfalls, and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, 9(4):143–172, 2021. 1, 2