RESEARCH ARTICLE

GeoFACT: Geospatial Few-Shot Adaptive and Compatible Training Framework for Remote Sensing Images Classification

Kawsar Farooq[1], **Muhammad Zarar**[1], **Yar Muhammad**[2,*], **Muhammad Hasnat**[3], **Yanze Li**[2] **and Xin Luo**[2]

[1] College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[2] School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[3] Department of Computer Science, City University of Science and Information Technology, Peshawar 24370, Pakistan

Abstract

Remote Sensing (RS) data streams necessitate classification models capable of assimilating novel information without the computational burden of retraining from scratch. Although Few-Shot Incremental Learning (FSIL) offers a promising paradigm for adaptation using limited samples, existing methodologies often falter due to the high inter-class similarity and complex background clutter inherent in aerial imagery. To address these challenges, we introduce a novel framework, Geospatial Few-Shot Adaptive Compatible Training, designed to mitigate catastrophic forgetting in dynamic environments. This method ensures forward compatibility by strategically allocating embedding space for future categories via virtual prototypes, while simultaneously maintaining backward compatibility through knowledge distillation and selective network freezing. The proposed framework was rigorously evaluated on the AID and NWPU-RESIS45 benchmark datasets to assess its ability to balance plasticity for new tasks with stability for previously learned classes. Experimental analysis using accuracy, loss, and F1-score metrics demonstrates the robustness of the approach, achieving a base-session accuracy of 96.62% on the AID dataset and 96.74% on the NWPU-RESISC45 dataset. These results confirm that the proposed model effectively addresses the limitations of current state-of-the-art techniques, offering a scalable solution for RS classification, Catastrophic Forgetting, Forward Compatibility, and Few-Shot Incremental Learning.

Keywords: incremental learning, few-shot learning, remote sensing, Geo-FACT, image processing.

1 Introduction

The classification of remote sensing (RS) imagery is a critical task with wide-ranging scientific and operational applications [1]. Methodological

Submitted: 26 January 2026
Accepted: 15 March 2026
Published: 19 April 2026

Vol. 1, **No.** 1, 2026.
10.62762/TAIC.2026.348838

***Corresponding author:**
✉ Yar Muhammad
yarkhan@buaa.edu.cn

Citation
Farooq, K., Zarar, M., Muhammad, Y., Hasnat, M., Li, Y., & Luo, X. (2026). GeoFACT: Geospatial Few-Shot Adaptive and Compatible Training Framework for Remote Sensing Images Classification. *ICCK Transactions on Applied Intelligence and Cybernetics*, 1(1), 53–66.

advancements have evolved from an early reliance on handcrafted features to unsupervised feature learning [2] and, more recently, deep learning-based approaches [3]. Despite their success, most contemporary methods operate under a static batch learning paradigm, assuming a fixed dataset and a predefined set of classes. This assumption is often invalid in real-world scenarios where RS data streams arrive sequentially with continually expanding volumes and categories. Retraining models from scratch under such dynamic conditions is computationally prohibitive. To address this, Incremental Learning (IL) [4–6] presents a promising alternative, enabling models to learn continuously from new data without the catastrophic forgetting of prior knowledge.

In contrast to classical machine learning, which requires the entire dataset, IL algorithms learn from data provided gradually over time [7]. This is particularly crucial for real-world applications where new classes emerge continuously [8], a scenario formalized as Class Incremental Learning (CIL) [12]. A central challenge in applying deep neural networks to CIL is catastrophic forgetting, in which a model's performance on previously learned tasks degrades significantly as it learns new ones. This contrasts with biological learning, which naturally integrates new skills while retaining old ones. Consequently, the field of Continual Learning (CL) [9], or lifelong learning [10], seeks to bridge the gap between biological and artificial systems. This challenge is exacerbated in Few-Shot Incremental Learning (FSIL) settings. FSIL is common in RS domains, where data collection and labeling are costly, requiring models to learn new tasks with only a few labeled examples while maintaining performance across all prior classes. Figure 1 illustrates the dual challenge inherent in this process: mitigating catastrophic forgetting while preventing overfitting to limited new data. While existing IL strategies—categorized into replay-based, regularization-based, and parameter-expansion-based methods that offer mechanisms for retaining knowledge [11]—often face limitations in the RS domain.

Despite this, replay and regularization-based approaches often lose plasticity as incremental sessions accumulate, while parameter-expansion strategies can introduce prohibitive memory and computation overhead on resource-constrained platforms. As shown in Figure 1(a), the initial learning phase involves a CNN backbone, parameterized by θ_{old}, trained on a base dataset $\mathcal{D}_1$ to discriminate between initial classes (e.g., Airport, Playground). Subsequently, the incremental update phase (Figure 1(b)) processes a new dataset $\mathcal{D}_2$ containing novel categories (e.g., Bridge, Port). Here, the classifier head is dynamically expanded, and parameters are updated to θ_{new} to facilitate the classification of both existing and newly added scenes. In remote sensing (RS), these limitations are amplified by strong non-stationarity: variations in sensors, spatial resolution, seasonality, and acquisition conditions induce substantial distribution shift, violating the i.i.d. assumption. Consequently, an effective incremental learning model for RS must accommodate evolving class semantics while controlling catastrophic forgetting and avoiding unbounded resource growth. Therefore, developing efficient IL models that adapt to evolving distributions without catastrophic forgetting or prohibitive resource growth is a vital research frontier.

To address these, we proposed a novel FSIL approach, Geo-FACT, the Geospatial Few-Shot Adaptive Compatible Training framework, to explicitly dual prospective representation planning with stable incremental updating. It generates virtual prototypes to reserve embedding sectors for unseen classes, ensuring forward compatibility. A dual-space regularization (feature + probability) preserves discriminative geometry and stabilizes predictions under distribution shift. When new classes arrive, a back-compatible update maps them into the reserved sectors, improving scalability and reducing interference. The key contributions of this work are:

- The proposal of the Geo-FACT framework, designed to integrate new RS classes from limited labeled data while preserving knowledge of prior classes.
- A parameter-efficient expansion mechanism that adapts the model to new data distributions—specifically tailored for RS complexities without the linear growth in parameters characteristic of existing expansion-based methods.
- A dedicated stability-plasticity strategy that simultaneously addresses catastrophic forgetting and few-shot overfitting via a novel attention distillation loss.
- Extensive experiments on established and novel RS benchmarks, demonstrating state-of-the-art

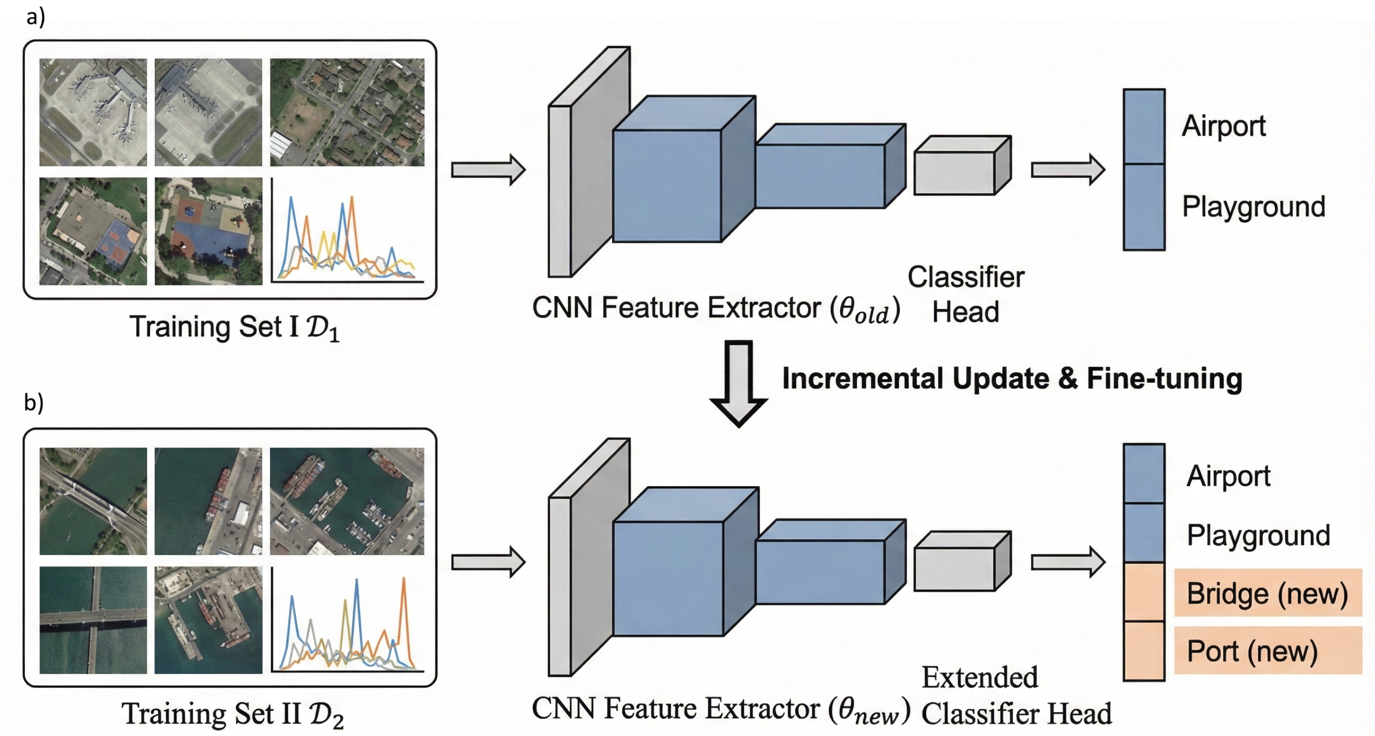


Figure 1. Schematic overview of the proposed incremental learning framework for remote sensing scene classification. (a) The CNN backbone is trained on the base dataset, and (b) the incremental update phase, where the classifier head is dynamically expanded to include new classes, enabling the sequential learning of new scene concepts without retraining.

(SOTA) performance in FSIL scenarios and superior efficiency compared to existing parameter-expansion approaches.

2 Related Work

The challenge of continuous learning from dynamic data streams has attracted significant research interest. This section reviews the development of these methodologies, organized into two distinct categories: **Class Incremental Learning** (**CIL**) and **Few-Shot Incremental Learning** (**FSIL**). Section 2.1 discusses CIL, where models must acquire new knowledge without erasing previously learned information. Subsequently, section 2.2 examines FSIL, exploring how recent works have extended incremental capabilities to data-scarce scenarios.

2.1 Class Incremental Learning

CIL addresses the critical stability-plasticity dilemma, requiring models to assimilate new categories (plasticity) while mitigating the catastrophic forgetting of previously acquired knowledge [32]. In RS scene classification, methodologies are predominantly categorized into regularization-based, parameter-isolation, and replay-based strategies [13]. Regularization approaches aim to preserve knowledge by constraining the loss function; for instance, Elastic Weight Consolidation (EWC) uses the Fisher Information Matrix (FIM) to penalize changes to parameters critical for prior tasks [14]. Similarly, Learning without Forgetting (LwF) employs knowledge distillation to maintain consistency between the outputs of the old and updated models [15, 31]. Recent advancements, such as LWF-Plus and PodNet [16], have mitigated these constraints through spatially aware distillation. However, these methods often falter in RS applications because significant distributional shifts and complex background clutter render rigid regularization insufficient for distinguishing fine-grained land cover classes.

Moreover, parameter-isolation methods, such as Progressive Neural Networks (PNN) [17], dynamically expand the network architecture to accommodate new tasks. Meanwhile, replay or rehearsal methods have emerged as a dominant paradigm in RS, exemplified by Incremental Classifier and Representation Learning (iCaRL) [18], which maintains a limited buffer of exemplars for rehearsing prior concepts. Advanced variations, such as ER-ACE [19], address class bias, whereas RS-specific implementations, such as RS-CIL and

Table 1. Comparison of representative CIL and FSIL methods related to Geo-FACT.

Cat.	Method	Replay Buffer	Param. Exp.	Backbone Update	Prototype/Topology Modeling	Main Limitation in RS–FSIL Context
CIL	EWC [14]	×	×	✓	Importance-based weight regularization	Rigid; struggles with geo-domain shifts and cluttered RS scenes.
CIL	LwF/PodNet [16, 31]	×	×	✓	Logit / feature distillation	Distillation alone is insufficient for fine-grained RS classes.
CIL	PNN [17]	×	✓	✓	Task-specific columns (no explicit prototypes)	Parameter growth per task; poor scalability for long RS streams.
CIL	iCaRL [18]	✓	×	✓	Nearest-mean exemplars as prototypes	Requires storage, overfits to small buffers; misses intra-class diversity.
CIL	ER-ACE [19]	✓	×	✓	Replay-based classifier correction	Bounded by buffer size and high rehearsal costs in large RS datasets.
FSIL	TOPIC [28]	×	×	✓	Neural gas to preserve feature topology	Retrospective: fixes topology, limiting plasticity for novel RS structures.
FSIL	CEC [21]	×	×	✓	Decoupled classifier with evolving centroids	No explicit mechanism for RS-specific distribution shifts or drift.
FSIL	NC-FSCIL [22]	×	×	✓	ETF-based prototypes (neural collapse)	Assumes stable geometry; sensitive to RS background clutter.
FSIL	FACT [20]	×	×	✓	**Virtual prototypes** for forward compatibility	Not tailored to non-stationary RS distributions or complex scenes.
FSIL	**GeoFACT**	×	×	✓	**Virtual prototypes + embedding reservation**	Specifically handles RS feature drift without replay or parameter growth.

CGR, utilize generative replay to synthesize complex textures. Although rehearsal generally outperforms regularization, it faces severe limitations: strict storage constraints restrict buffer size, and models tend to overfit to the sparse set of stored exemplars, failing to capture the rich intra-class diversity of complex scenes.

2.2 Few-Shot Incremental Learning

FSIL extends the CIL framework to realistic scenarios where new classes arrive with scarce annotated data. This regime presents a dual challenge: the model must resist forgetting old base classes while avoiding severe overfitting to the few novel examples. Early attempts focused on preserving the structure of the feature space. The Topology-Preserving Knowledge Incremental Learning (TOPIC) framework employs a neural gas network to anchor the topology of the feature manifold [28]. While effective for stability, TOPIC is fundamentally retrospective; it constrains the model to fit a predefined topology, limiting plasticity when encountering RS scenes with novel semantic structures.

Alternatively, methods like Continually Evolving Classifiers (CEC) and NC-FSCIL decouple the feature extractor from the classifier [21, 22]. NC-FSCIL leverages the geometry of neural collapse to pre-assign class prototypes to an equiangular tight frame, providing a robust mathematical foundation for expansion. In RS, methods such as RSPrompt [25] utilize visual prompt learning to adapt frozen representations without altering the backbone. However, reliance on a static backbone is a major bottleneck in RS, where the feature extractor often fails to generalize to unseen, spectrally distinct land-cover types without adaptation. Recent research has shifted toward prospective learning, exemplified by the Forward-Compatibility Training (FACT) framework [20]. FACT explicitly allocates "virtual prototypes" in the embedding space to reserve capacity for unseen classes. Our work identifies the need to adapt these prospective paradigms to the complex, non-stationary distributions of RS data. Table 1 summarizes the main differences between representative CIL/FSIL methods and our GeoFACT framework in terms of replay, parameter growth, prototype modeling, and suitability for RS FSIL.

3 Materials and Method

3.1 Datasets

To evaluate the proposed GeoFACT framework, we utilized two widely recognized benchmarks in RS, the Aerial Image Dataset (AID) [23] and the NWPU-RESIS45 [24] datasets.

AID: The AID is a large-scale collection comprising 10,000 aerial images sourced from Google Earth. The dataset spans 30 distinct scene classes, as illustrated in Figure 2(a). It exhibits class imbalance, with the number of samples per category ranging from 220 to 420. Each image has a fixed resolution of 600×600 pixels. Notably, the images were acquired from diverse geographic locations and at various times, ensuring significant intra-class diversity and introducing heterogeneous imaging conditions.

NWPU-RESIS45: This dataset is a more extensive benchmark containing 31,500 images distributed across 45 scene classes. Unlike AID, NWPU-RESIS45 is balanced, with exactly 700 images per class, as shown in Figure 2(b). The images are standardized to 256×256 pixels, with spatial resolutions varying from approximately 30m to 0.2m. This broad resolution spectrum captures both large-scale natural features (e.g., islands, lakes, mountains) and fine-grained urban structures.

In particular, these datasets are curated and largely static, with limited temporal evolution and no explicit modelling. Consequently, we evaluate the targets of few-shot incremental class learning primarily under benchmark conditions to extend Geo-FACT to fully streaming multi-temporal settings.

3.2 Preliminaries

The learning process is divided into an initial base session followed by a sequence of a few-shot incremental sessions. The main notation used in this paper is summarized in Table 2. In the base session, the model is trained on a large-scale dataset, denoted as:

$$\mathcal{D}^0 = \{(x_i, y_i)\}_{i=1}^{N_0} \tag{1}$$

where $x_i \in \mathbb{R}^D$ represents a training instance and $y_i \in \mathcal{Y}_0$ is its corresponding label from the base class set $\mathcal{Y}_0$. The model is evaluated on a separate testing set:

$$\mathcal{D}^0_{test} = \{(x_j, y_j)\}_{j=1}^{M_0} \tag{2}$$

Table 2. Summary of main notation used in the Geo-FACT framework.

Symbol	Description
D_0	Base-session training set of RS images and labels
D_b	Few-shot incremental dataset at session b
$\mathcal{Y}_0$	Set of base classes in the initial session
$\mathcal{Y}_{\text{old}}$	Set of all previously learned classes
N, K	N-way K-shot configuration parameters
x, y	Input RS image and its ground-truth label
$\varphi(\cdot)$	Feature embedding / backbone mapping
$f(x)$	Classification model outputs (logits/scores)
W	Classifier weight / prototype matrix
w_y	Prototype (weight vector) of class y
P_v	Matrix of virtual prototypes
K_{virt}	Number of virtual classes / prototypes
τ	Softmax temperature scaling parameter
$p(y\|x)$	Predicted probability of class y for input x
L_{CE}	Cross-entropy classification loss
L_{KD}	Knowledge distillation loss
L_{feat}	Feature-level regularization loss
L_{virt}	Loss over real and virtual classes
L_{total}	Final objective (combined loss terms)

The objective in this phase is to minimize the empirical risk by learning a function $f(x)$ that maps inputs to labels:

$$\min_f \sum_{(x_j,y_j)\in\mathcal{D}^0_{test}} \mathcal{L}\big(f(x_j), y_j\big) \tag{3}$$

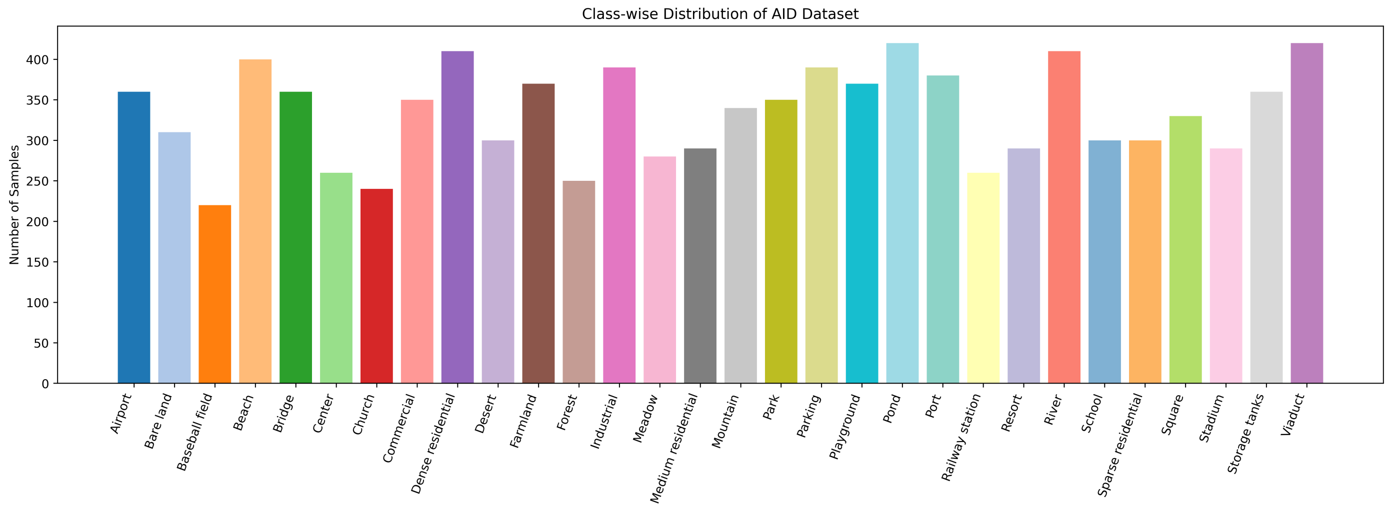
where $\mathcal{L}(\cdot)$ denotes the loss function quantifying the discrepancy between the predicted and ground-truth labels. The model $f(x)$ is decomposed into two components: a feature embedding module $\phi(\cdot) : \mathbb{R}^D \to \mathbb{R}^d$ and a linear classifier $W \in \mathbb{R}^{d\times|\mathcal{Y}_0|}$. The prediction is given by:

$$f(x) = W^T\phi(x) \tag{4}$$

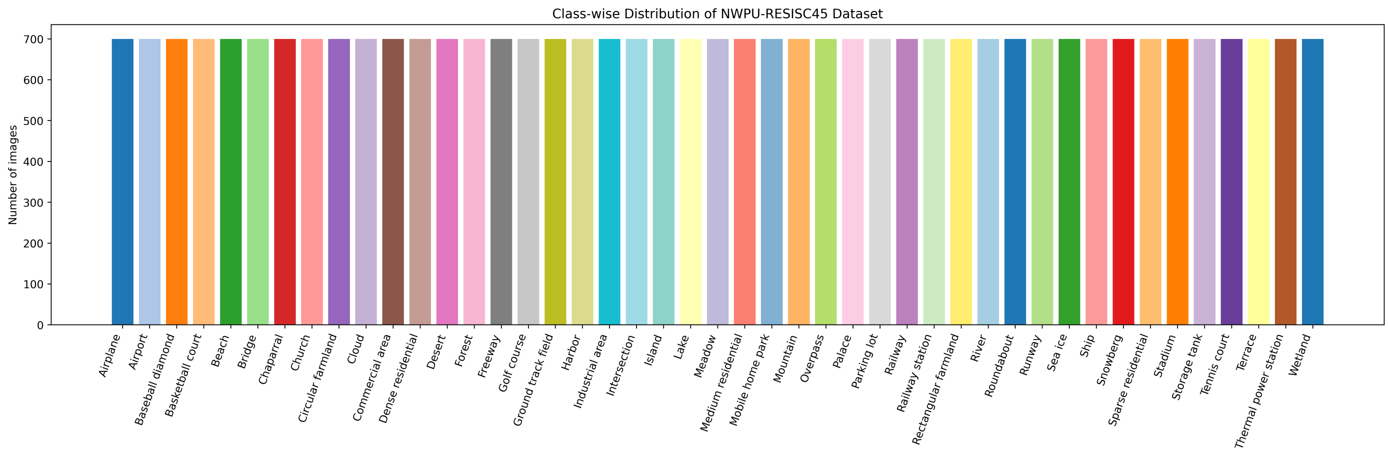
where $W = [w_1, w_2, \ldots, w_{|\mathcal{Y}_0|}]$ contains the weight vectors for each class.

In the subsequent incremental phase, new classes arrive sequentially in small batches, forming a series of datasets $\{\mathcal{D}^1, \ldots, \mathcal{D}^B\}$. Each dataset $\mathcal{D}^b$ is formulated as an N-way K-shot task, with N new classes and K labeled examples per class. The goal is to learn the new classes in $\mathcal{D}^b$ while maintaining performance on all previously learned classes $\mathcal{Y}_0 \cup \cdots \cup \mathcal{Y}_{b-1}$. This requires minimizing the cumulative empirical risk across all encountered test sets:

$$\sum_{(x_j,y_j)\in\bigcup_{k=0}^{b}\mathcal{D}^k_{test}} \mathcal{L}\big(f(x_j), y_j\big) \tag{5}$$



(a)



(b)

Figure 2. Class-wise distribution of the dataset.

3.2.1 *Knowledge Distillation*

To ensure backward compatibility in FSIL, the model is enforcing consistency between the current model and the model from the previous session, thereby mitigating catastrophic forgetting:

$$L_{total} = (1-\lambda)L_{CE}(x, y) + \lambda L_{KD}(x) \tag{6}$$

The distillation loss L_{KD} is defined as:

$$L_{KD}(x) = -\sum_{k=1}^{|\mathcal{Y}_{old}|} \sigma_k\big(\bar{W}^T\bar{\phi}(x)\big)\log\sigma_k\big(W^T\phi(x)\big) \tag{7}$$

where $\mathcal{Y}_{old}$ represents the set of previously learned classes, and $\sigma_k(\cdot)$ denotes the softmax probability for the k-th class. The terms $\bar{W}$ and $\bar{\phi}$ refer to the frozen classifier and embedding parameters from the previous session, serving as the teacher model.

3.2.2 *Prototype-Based Learning*

Standard knowledge distillation can lead to overfitting when applied to scarce data. Consequently, FSL methods often adopt prototype-based approaches to enhance stability. For instance, ProtoNet [17] utilize the embedding function $\phi(\cdot)$ to compute a class prototype P_k by averaging the feature vectors of all support instances belonging to class k:

$$P_k = \frac{1}{N_k}\sum_{v=1}^{|\mathcal{D}^b|}\mathbb{I}(y_v = k)\phi(x_v) \tag{8}$$

where $\mathbb{I}(\cdot)$ is the indicator function, the prototype P_k serves as the classifier weight w_k to keep the embedding fixed across progressive stages of compatibility between older and newer models is maintained. These approaches primarily address backward compatibility—ensuring the model retains past knowledge. They largely neglect *forward*

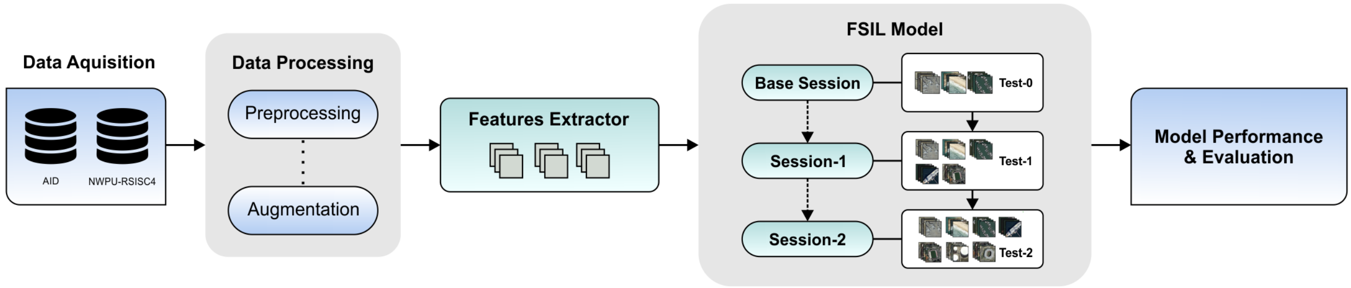


Figure 3. Overall workflow of the proposed methodology RS image classification.

compatibility, failing to prepare the model during the base session for future updates. Our approach addresses this limitation by explicitly optimizing the future extendability.

3.2.3 *Image Processing and Augmentation*

To bolster model robustness and generalization across the AID and NWPU-RESISC45 datasets, we employ a multi-stage augmentation pipeline. Images are first resized to 256×256 pixels and randomly cropped to 224×224 pixels. We further apply random horizontal flips and the Auto-augment policy to simulate complex environmental variations. These transformations ensure that the feature extractor learns invariant representations, which are critical for the stability of downstream incremental updates.

3.3 Proposed Geo-FACT Framework

We propose the Geo-FACT framework, designed for FSIL in RS scene classification. The existing traditional incremental learning that often suffers from catastrophic forgetting or rigid embedding spaces, Geo-FACT proactively reserves capacity for future, unseen classes during the base training phase. The overall system workflow is presented in Figure 3 and the proposed architecture is illustrated in Figure 4.

3.3.1 *Base Module with Feature Extraction*

The core of Geo-FACT lies in its ability to treat the embedding space as a dynamic resource. The dataset is split into sequential sessions $\mathcal{S} = \{\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_n\}$, and the model must learn new categories without collapsing the representations of preceding ones. As shown in Figure 4, input images are processed through a CNN backbone to generate a fixed-dimensional embedding $\varphi(x) \in \mathbb{R}^d$.

To maintain geometric stability during incremental updates, we employ a fixed embedding strategy. Once the base parameters are optimized, the feature extractor is frozen to prevent feature drift. These features are then projected via fully connected layers into an embedding space $\varphi(x) \in \mathbb{R}^d$. The probability of an instance x belonging to a known class y is as follows:

$$p(y \mid x) = \frac{\exp(\tau \cdot \cos\langle w_y, \varphi(x)\rangle)}{\sum_{j \in Y} \exp(\tau \cdot \cos\langle w_j, \varphi(x)\rangle)} \tag{9}$$

where w_y represents the learnable class prototype.

3.3.2 *Virtual Prototype Environment*

The core innovation of our framework is the prototype environment, which regularizes the model in two distinct spaces: *Feature Space* and *Probability Space*, as shown in Figure 4. This module ensures the pre-allocation of embedding space for future classes using virtual prototypes. The model initially learns $|Y_0|$ classes but must eventually accommodate $Y_0 + (N \times B)$ classes. We address this by substituting the traditional linear classifier with a cosine similarity classifier.

Feature Space via Manifold Mixup and Metric Learning: To anticipate future classes, we synthesize virtual classes via manifold mixup. To use interpolation between features of known classes $h(x_i)$ and $h(x_j)$, we generate synthetic representations z that simulate the distribution of unseen data:

$$z = \lambda h(x_i) + (1 - \lambda) h(x_j), \quad \lambda \sim \text{Beta}(\alpha, \alpha) \tag{10}$$

As an instance, the base session contains three RS classes $\{\text{Airport}, \text{Playground}, \text{Forest}\}$. We sample x_i from Airport and x_j from Playground of a virtual prototype $\lambda \sim \text{Beta}(\alpha, \alpha)$:

$$(\text{e.g.}, \lambda = 0.3), z = 0.3\, h(x_i) + 0.7\, h(x_j) \tag{11}$$

The feature-space objective keeps z separated from seen-class regions (reserved sector), while the probability-space regularizer discourages confident assignment of z to any seen class; future novel classes are then aligned to these reserved sectors during incremental updates. In the feature space, we apply

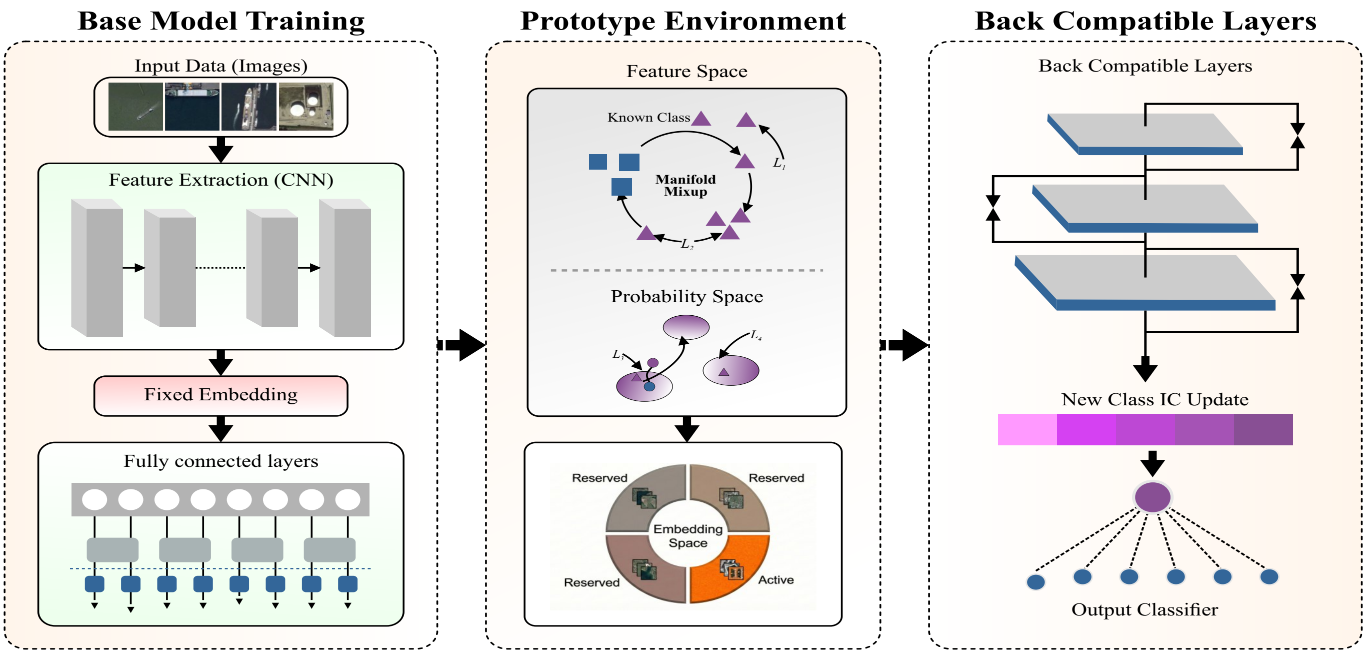


Figure 4. Proposed Geo-FACT framework, highlighting the use of virtual prototypes and manifold mixup for forward-compatible learning.

dual metric constraints explicitly labeled as L_1 and L_2 in Figure 4. Intra-class compactness (L_1) ensures that the virtual classes (L_4) remain distinct from those of the last reservation (L_2). The generated virtual instances z maintain a margin from known classes, effectively reserving voids in the manifold. The combined feature loss is formulated as:

$$\mathcal{L}feat = \underbrace{\mathcal{L}CE(f_v(x), y)} L_1 + \gamma \underbrace{\mathcal{L}margin(\text{Mask}(f_v(z)), \hat{y})}_{L_2} \tag{12}$$

where $\hat{y}$ is the assigned virtual class label, effectively pushing virtual features into the "Reserved" sectors shown in the Embedding Space diagram.

Probability Space via Distribution Regularization: To regulate the output distributions in the probability space to calibrate confidence. As depicted in Figure 4, this involves two specific losses: L_3 known as class confidence, maximizes the probability of the correct class for real data. L_4 virtual class uncertainty enforces a specific distribution shape for virtual classes to prevent overconfidence. The feature distribution is a bimodal Gaussian mixture to capture both the confident core of known classes and the transitional nature of virtual classes as:

$$p(\phi(x) \mid w_i, p_v) = \eta\mathcal{N}(\phi(x) \mid w_i, \Sigma) + (1-\eta)\mathcal{N}(\phi(x) \mid p_v, \Sigma) \tag{13}$$

This stochastic modeling ensures that the probability distributions for virtual classes (L_4) remain distinct from known classes (L_3).

3.3.3 *Back-Compatible Update and Embedding Reservation*

In an incremental inference stage, embedding space is divided into active sectors (current classes) and reserved sectors (pre-allocated via virtual prototypes). When new classes arrive, we employ back-compatible layers (shown as residual blocks with specific updates). These layers perform the new class IC update by refining the features to map strictly into the previously reserved sectors without interfering with the active sectors.

To achieve robust inference during this update, we utilize a progressive mixing function that balances the base knowledge with the new increments:

$$\text{Mix}\eta = \frac{\exp(w_i^T(\phi(x) + p_v))}{\sum q = 1^{|Y_a|} \exp(w_q^T(\phi(x) + p_v))} \tag{14}$$

This mechanism ensures the output classifier can distinguish between old and new classes by leveraging the geometric representation established during base training.

However, our proposed Geo-FACT framework introduces a novel paradigm for proactive latent-space management in IL. Departing from reactive methods that adapt to new data at the risk of catastrophic forgetting, Geo-FACT employs a forward-compatible reservation strategy. To incorporate manifold-mixed virtual prototypes and a bimodal stochastic model during initial training, it dynamically structures the embedding space to anticipate unseen classes. This foresight enables the model to pre-allocate decision boundaries for future categories and to expand knowledge seamlessly without retraining or parameter changes.

4 Experiment and Results

4.1 Implementation setup

The proposed framework was evaluated using a dual-platform environment to ensure experimental stability. The local computations were performed on an HP EliteBook 840 G5 workstation equipped with an Intel Core i5-7200U CPU (2.50 GHz), 16 GB RAM, and a 512 GB SSD. Large-scale training is offloaded to a remote laboratory server running Ubuntu 20.04.6 LTS, powered by a 13th-generation Intel Core i7-13700 CPU (32 GB of RAM) and accelerated by an Intel A780 and an NVIDIA 2503 GPU.

The model was implemented in PyTorch using a ResNet-18 backbone, selected for its robust feature-extraction capabilities in the RS scene classification task. Table 3 outlines the key variables used to train the proposed framework. The Geo-FACT algorithm 1 showed a lower accuracy on the TS-0 dataset than on the NWPU-RESISC45 dataset, but the decrease was less pronounced with an initial learning rate of 0.1. To evaluate the impact of mitigating the negative effects of forgetting by expanding the experiments, three distinct batch sizes were used: 64, 128, and 256.

Table 3. Configuration of parameters for Geo-FACT.

Parameters	Value
Batch size	64, 128, 256
Starting Learning Rate	0.1
Optimizer	Stochastic Gradient Descent (SGD)
Epochs	200

Algorithm 1 outlines the forward-compatible training procedure, by initializing the feature extractor ϕ_0 and constructing an extended classifier weight matrix $W \in \mathbb{R}^{d\times(K_{base}+K_{virt})}$. This is formed by concatenating the weights of the existing base classes (W_{base}) with a set of randomly initialized prototypes (P_v) representing K_{virt} virtual classes. In the training phase, the algorithm iterates through mini-batches of the base dataset $\mathcal{D}^0$. The input features are extracted via ϕ_0, and logits are computed using the extended classifier W. The model parameters are updated via gradient descent to minimize a total objective function $\mathcal{L}_{total}$, which aggregates a virtual classification loss $\mathcal{L}_{virt}$ calculated over the union of all classes and an auxiliary regularization term $\lambda\mathcal{L}_f$.

Algorithm 1 Forward Compatible Training for FSIL

Input: Base dataset $\mathcal{D}^0 = \{(x_i, y_i)\}$, number of virtual classes K_{virt}
Output: Feature extractor ϕ_θ, Weights $W = [W_{base}, P_v]$.

Initialize parameters θ, W_{base}, and P_v randomly
$W \leftarrow [W_{base}, P_v] \in \mathbb{R}^{d\times(K_{base}+K_{virt})}$ ▷ Construct extended classifier
for ($epoch = 1$ **to** N) **do**
  for (each mini-batch $\mathcal{B} \sim \mathcal{D}^0$) **do**
    $Z = \phi_\theta(x)$ for $x \in \mathcal{B}$ ▷ Extract features
    $O = W^T Z$ ▷ Compute logits
    $\mathcal{L}_{virt} = -\frac{1}{|\mathcal{B}|}\sum_{(x,y)\in\mathcal{B}} \log \frac{\exp(w_y^T\phi_\theta(x))}{\sum_{j\in C_{all}}\exp(w_j^T\phi_\theta(x))}$ ▷ Calculate Virtual Loss
    $\mathcal{L}_{total} = \mathcal{L}_{virt} + \lambda\mathcal{L}_f$ ▷ Total loss
    $\theta \leftarrow \theta - \eta\nabla_\theta\mathcal{L}_{total}$ ▷ Update θ
    $W \leftarrow W - \eta\nabla_W\mathcal{L}_{total}$ ▷ Update W
  End for
End for
Return Updated ϕ_θ, W

4.2 Main results

We present a comprehensive evaluation of the proposed Geo-FACT framework on two widely recognized remote sensing image classification benchmarks, NWPU-RESISC45 and AID. The experiments were designed to assess the model's performance under FSIL constraints, specifically its ability to mitigate catastrophic forgetting while adapting to novel classes.

4.2.1 Result on NWPU-RESIS45 Benchmark

Evaluation on the NWPU-RESIS45 dataset further validates the scalability of our approach. As shown in Table 4, the proposed method exhibits graceful degradation in accuracy across all batch sizes, compared with standard incremental learning techniques. Notably, the Geo-FACT framework successfully distinguishes between highly similar satellite categories, such as dense residential and medium residential, owing to the refined cross-entropy loss and an optimized batch-processing strategy.

Table 4. IL performance on NWPU-RESISC45 and AID benchmarks across batch sizes.

Dataset	Session	Batch 64	Batch 128	Batch 256
NWPU-RESISC45	TS-0	95.50%	96.23%	**96.74%**
	TS-1	83.97%	84.00%	**86.51%**
	TS-2	76.66%	77.40%	**79.96%**
	TS-3	73.03%	73.22%	**76.33%**
	TS-4	70.60%	69.67%	**73.23%**
AID	TS-0	93.96%	96.30%	**96.62%**
	TS-1	85.53%	**85.93%**	85.73%
	TS-2	74.57%	79.42%	**79.55%**
	TS-3	67.13%	**71.79%**	70.30%
	TS-4	61.92%	**65.89%**	64.71%

Table 5. Ablation study results on AID and NWPU-RESISC45 datasets.

Variant	AID		NWPU-RESISC45	
	Avg. Inc. (%)	TS-4 (%)	Avg. Inc. (%)	TS-4 (%)
Full Geo-FACT	**79.30**	**64.71**	**82.51**	**73.23**
w/o L_{virt}	68.42	51.63	71.18	63.36
w/o L_2	73.15	59.50	77.43	69.04
w/o Prob. Reg.	75.86	61.28	78.94	70.55
Baseline (Class. only)	65.94	48.37	68.87	60.91

4.2.2 Result on AID Benchmark:

The performance of the Geo-FACT framework on the AID benchmark is summarized in Table 4. Our model achieved a base-session (TS-0) accuracy of 96.62% on the AID dataset, indicating a strong initial embedding space. Throughout the incremental stages, the model maintained high classification outcomes, significantly outperforming in the final session, TS-4. This suggests that the use of virtual prototypes effectively stabilizes the representation of aerial imagery even with limited training samples. The AID dataset showed a consistent pattern of reducing accuracy as the number of training sessions increased. It began with lower accuracy on TS-0 than on the NWPU-RESISC45 dataset, but the decrease was less pronounced by TS-4.

4.3 Ablation Study

Component Effects Analysis. To validate the contributions of each component to the overall Geo-FACT, we conduct an ablation study on both benchmarks (AID and NWPU-RESISC45). The effectiveness of the full Geo-FACT model against four variants: w/o loss over virtual classes L_{virt}, which removes virtual prototype reservation; w/o L_2, which excludes feature-space margin constraints; w/o Prob. Reg. (L_3/L_4), which omits probability-space uncertainty; and a baseline, utilizing only standard cross-entropy with a cosine classifier. We report the average incremental accuracy (the average of top-1 accuracies across all incremental sessions) and the final session accuracy (TS-4) for both datasets, as shown in Table 5.

Efficiency Analysis: To evaluate the practical deployability of GeoFACT, we report a quantitative efficiency comparison against representative incremental learning baselines. Table 6 summarizes training time per session, replay-buffer memory,

Table 6. Quantitative efficiency comparison of incremental learning methods. All estimates are based on a ResNet-18 backbone with an input size of 224×224, trained over 5 incremental sessions on a remote sensing benchmark.

Method	**Training Time (per session)**	**Memory Usage (Buffer)**	**Parameter Growth (Total)**	**Inference Latency**
iCaRL [18]	≈120 min; full retraining on combined old + new data	≈250 MB; stores 20 exemplars per class at 224×224 px	11.2M parameters; backbone + linear classifier	≈25 ms; forward pass plus nearest-neighbor search
PNN [17]	≈90 min; trains a new column from scratch per session	≈45 MB; stores previous column weights (no raw data)	56.0M parameters; linear growth with one new column per task	≈60 ms; forward pass through all 5 columns
GeoFACT (Ours)	≈20 min; backbone frozen, fine-tunes lightweight adapters	0 MB; no exemplars, uses virtual prototypes	11.4M parameters; minimal growth (≈0.2M) for residual adapters	≈12 ms; single backbone pass, latency essentially constant

Table 7. Comparative analysis of the proposed Geo-FACT model against State-of-the-Art methods.

Benchmark	**Ref.**	**Method**	**S-0**	**S-1**	**S-2**	**S-3**	**S-4**
Comparison on General Benchmarks (Cub200)	[18]	iCaRL	68.47	52.65	48.61	44.16	36.62
	[26]	EEIL	68.68	53.63	47.91	44.20	36.30
	[27]	Rebalancing	68.68	57.12	44.21	28.78	26.71
	[28]	TOPIC	68.68	62.49	54.81	49.99	45.25
	[29]	Decoupled-Cosine	75.52	66.46	66.46	61.20	60.86
	[30]	Decoupled-DeepEMD	75.35	70.69	66.68	62.34	59.76
Comparison on Remote Sensing (RS) Benchmarks	[31]	LwF	89.40	65.90	51.30	44.00	38.40
	[32]	CIL	89.40	63.50	50.40	47.30	40.90
Ours	**Geo-FACT**	**AID**	**96.62**	**85.73**	**79.55**	**70.30**	**64.71**
	Geo-FACT	**NWPU-RESIS45**	**96.74**	**86.51**	**79.96**	**76.33**	**73.23**

parameter growth, and inference latency under a matched ResNet-18 backbone (224 × 224 input) over five incremental sessions.

The replay-based iCaRL [18] incurs substantial memory overhead due to exemplar storage and adds inference cost from nearest-neighbor retrieval. In contrast, parameter-isolation methods such as PNN [17] eliminate raw-data storage but introduce linear parameter growth and increased inference latency because all columns must be evaluated at test time. GeoFACT achieves a more favorable trade-off: it requires no replay buffer, introduces only minimal parameter growth from lightweight adapters, and maintains near-constant inference latency with a single backbone pass. These results support the claim that GeoFACT is both memory-efficient and scalable for long-session remote sensing FSIL.

4.4 Comparative Analysis with SOTA Methods

To evaluate the efficacy of the proposed Geo-FACT framework, we conducted a comprehensive comparative analysis against prominent SOTA methods, ranging from general incremental learning techniques [18, 26–30] to specialized remote sensing approaches [31, 32]. Table 7 presents the quantitative performance across sequential learning sessions (S-0 to S-4) to show the outcome of each strategy, giving Geo-FACTs about their suitability in situations that demand strong adjustment to unseen data without compromising earlier learned data. Table 7 illustrates a comparative analysis of the proposed approach with the SOTA methods used in the literature.

The computational concentration of Geo-FACT backbone architecture remains fixed, and only the classifier head expands with new classes, leading to a predictable parameter increase proportional to the number of added categories. In addition, the virtual prototype environment is generated on-the-fly from existing embeddings and does not require storing large replay buffers, which limits memory overhead. We report parameter counts and per-session training time, along with an accuracy comparison, to quantify scalability under resource constraints.

5 Conclusion

In this research study, we proposed a novel Geo-FACT framework to address the persistent challenge of catastrophic forgetting in RS image classification. Unlike existing SOTA techniques, our approach establishes a scalable and robust direction for IL by explicitly preparing the model for future, unseen data distributions. The core contribution of this work lies in the strategic allocation of the embedding space. We introduced the concept of virtual prototypes to reserve latent space for future classes during the initial training session, thereby enabling the system to acquire new knowledge without overwriting existing representations. We integrated manifold mixup to generate virtual instances that anticipate future class distributions. The proposed Geo-FACT efficacy was validated through extensive experiments on the AID and NWPU-RESIS45 datasets. The experimental results demonstrate that our method significantly outperforms baseline techniques in sequential learning scenarios. While performance fluctuates as the number of sessions increases, the incremental reduction in data size and the introduction of new classes are accompanied by superior stability and retention of prior knowledge, as the proposed Geo-FACT method exhibits. Consequently, future research will focus on developing unified frameworks that seamlessly transition between FSIL and large-scale class-incremental learning paradigms.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Cheng, G., Xie, X., Han, J., Guo, L., & Xia, G. S. (2020). Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13*, 3735-3756. [CrossRef]

[2] Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE geoscience and remote sensing magazine, 5*(4), 8-36. [CrossRef]

[3] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing, 152*, 166-177. [CrossRef]

[4] Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural networks, 113*, 54-71. [CrossRef]

[5] Gepperth, A., & Hammer, B. (2016). Incremental learning algorithms and applications. In *European symposium on artificial neural networks* (*ESANN*).

[6] De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., ... & Tuytelaars, T. (2021). A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence, 44*(7), 3366-3385. [CrossRef]

[7] Lopez-Paz, D., & Ranzato, M. A. (2017). Gradient episodic memory for continual learning. *Advances in neural information processing systems, 30*.

[8] Li, Z., & Hoiem, D. (2017). Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence, 40*(12), 2935-2947. [CrossRef]

[9] Zhou, D. W., Sun, H. L., Ning, J., Ye, H. J., & Zhan, D. C. (2024, August). Continual learning with pre-trained models: a survey. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence* (pp. 8363-8371). [CrossRef]

[10] Thrun, S. (1998). Lifelong learning algorithms. In *Learning to learn* (pp. 181-209). Boston, MA: Springer US. [CrossRef]

[11] Masana, M., Liu, X., Twardowski, B., Menta, M., Bagdanov, A. D., & Van De Weijer, J. (2022). Class-incremental learning: survey and performance evaluation on image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(5), 5513-5533. [CrossRef]

[12] Ye, Z., Zhang, Y., Zhang, J., Li, W., & Bai, L. (2024). A multiscale incremental learning network for remote sensing scene classification. *IEEE transactions on geoscience and remote sensing, 62*, 1-15. [CrossRef]

[13] Song, J., Jia, H., & Xu, F. (2023, July). Class-incremental learning for remote sensing images based on

knowledge distillation. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium* (pp. 5026-5028). IEEE. [CrossRef]

[14] Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences, 114*(13), 3521-3526. [CrossRef]

[15] Xu, M., Zhao, Y., Liang, Y., & Ma, X. (2022). Hyperspectral image classification based on class-incremental learning with knowledge distillation. *Remote Sensing, 14*(11), 2556. [CrossRef]

[16] Douillard, A., Cord, M., Ollion, C., Robert, T., & Valle, E. (2020, August). Podnet: Pooled outputs distillation for small-tasks incremental learning. In *European conference on computer vision* (pp. 86-102). Cham: Springer International Publishing. [CrossRef]

[17] Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., ... & Hadsell, R. (2016). Progressive neural networks. *arXiv preprint arXiv:1606.04671*.

[18] Rebuffi, S. A., Kolesnikov, A., Sperl, G., & Lampert, C. H. (2017, July). iCaRL: Incremental Classifier and Representation Learning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5533-5542). IEEE. [CrossRef]

[19] Bonicelli, L., Boschini, M., Porrello, A., Spampinato, C., & Calderara, S. (2022). On the effectiveness of lipschitz-driven rehearsal in continual learning. *Advances in Neural Information Processing Systems, 35*, 31886-31901.

[20] Zhou, D. W., Wang, F. Y., Ye, H. J., Ma, L., Pu, S., & Zhan, D. C. (2022, June). Forward Compatible Few-Shot Class-Incremental Learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 9036-9046). IEEE. [CrossRef]

[21] Zhang, C., Song, N., Lin, G., Zheng, Y., Pan, P., & Xu, Y. (2021, June). Few-Shot Incremental Learning with Continually Evolved Classifiers. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 12450-12459). IEEE. [CrossRef]

[22] Yang, Y., Yuan, H., Li, X., Lin, Z., Torr, P., & Tao, D. (2023). Neural collapse inspired feature-classifier alignment for few-shot class incremental learning. *arXiv preprint arXiv:2302.03004*.

[23] Xia, G. S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., ... & Lu, X. (2017). AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing, 55*(7), 3965-3981. [CrossRef]

[24] Cheng, G., Han, J., & Lu, X. (2017). Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE, 105*(10), 1865-1883. [CrossRef]

[25] Bahng, H., Jahanian, A., Sankaranarayanan, S., & Isola, P. (2022). Exploring visual prompts for adapting large-scale models. *arXiv preprint arXiv:2203.17274*.

[26] Castro, F. M., Marín-Jiménez, M. J., Guil, N., Schmid, C., & Alahari, K. (2018, September). End-to-End Incremental Learning. In *European Conference on Computer Vision* (pp. 241-257). Cham: Springer International Publishing. [CrossRef]

[27] Hou, S., Pan, X., Loy, C. C., Wang, Z., & Lin, D. (2019, June). Learning a Unified Classifier Incrementally via Rebalancing. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 831-839). IEEE. [CrossRef]

[28] Tao, X., Hong, X., Chang, X., Dong, S., Wei, X., & Gong, Y. (2020, June). Few-Shot Class-Incremental Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 12180-12189). IEEE. [CrossRef]

[29] Zhou, F., Wang, P., Zhang, L., Wei, W., & Zhang, Y. (2023, June). Revisiting Prototypical Network for Cross Domain Few-Shot Learning. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 20061-20070). IEEE. [CrossRef]

[30] Zhang, C., Cai, Y., Lin, G., & Shen, C. (2022). Deepemd: Differentiable earth mover's distance for few-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(5), 5632-5648. [CrossRef]

[31] Finn, C., Abbeel, P., & Levine, S. (2017, July). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126-1135). PMLR.

[32] Yu, C. C., Chen, T. Y., Hsu, C. W., & Cheng, H. Y. (2024). Incremental scene classification using dual knowledge distillation and classifier discrepancy on natural and remote sensing images. *Electronics, 13*(3), 583. [CrossRef]

Kawsar Farooq received the B.S. and M.S. degrees in electronics and information engineering from Beihang University, China, in 2021 and 2024, respectively. He is currently pursuing his Ph.D. in computer science at Tianjin University, China. His research interests include knowledge graphs, pattern recognition, knowledge-based reasoning, multimodal LLMs, and image processing. (Email: kawsarfarooq@tju.edu.cn)

Muhammad Zarar received an MS degree in software engineering from Wuhan University, China, in 2024. He is currently pursuing a PhD in software engineering at Tianjin University, China. His research interests are focused on cutting-edge areas of computer vision, multimodal LLMs, knowledge base reasoning, and computational informatics. (Email: zararkhan@tju.edu.cn)

Yar Muhammad received the B.S. degree in Computer Science from the University of Malakand (UoM), Khyber Pakhtunkhwa, Pakistan. He was awarded a gold medal in his B.S. degree. He received the M.S. degree in Computer Science from Abdul Wali Khan University Mardan (AWKUM), Khyber Pakhtunkhwa, Pakistan. He has published multiple research articles in numerous reputable journals and serves as a reviewer for more than 20 journals and transactions. Currently, he is pursuing his Ph.D. degree in Computer Science and Technology at Beihang University, Beijing, China. His research interests include machine learning, deep learning, natural language processing, the Internet of Things, and computational intelligence. (Email: yarkhan@buaa.edu.cn)

Muhammad Hasnat is currently pursuing the Bachelor of Science degree in Software Engineering at City University of Science & Information Technology (CUSIT), Pakistan. His research interests include machine learning, image processing, and computational intelligence. (Email: mh.hasnat85@gmail.com)

Yanze Li received the B.E. degree in Computer Science and Technology from Nankai University, Tianjin, China. He won an academic excellence scholarship in his B.E. degree. Currently, he is pursuing his M.E. degree in Computer Technology at Beihang University, Beijing, China. His research interests include machine learning, deep learning, multi-modal large models, and security of large language models. (Email: yanze_li@buaa.edu.cn)

Xin Luo received the B.E. degree in Computer Science and Technology from South China University of Technology, Guangzhou, China. She won a first-class scholarship in her B.E. degree. Currently, she is pursuing her M.E. degree in Computer Technology at Beihang University, Beijing, China. Her research interests include educational data mining, machine learning, and deep learning. (Email: ZY2406305@buaa.edu.cn)