Meta Learning-based Few-Shot Learning: Remote Sensing

Project report submitted in partial fulfillment
of the requirements for the degree of

Bachelor of Technology
in
Computer Science and Engineering

by

Priyanshu Gupta - 22ucs155
Shantanu Gupta - 22ucc094
Shashwat Agarwal - 22uec123

Under Guidance of
Mentor : Dr. Lal Upendra Pratap Singh

LNMIIT
The LNM Institute of
Information Technology

Department of Computer Science and Engineering
The LNM Institute of Information Technology, Jaipur

November 2025

The LNM Institute of Information Technology
Jaipur, India

CERTIFICATE

This is to certify that the project entitled “Meta Learning-based Few-Shot Learning: Remote Sensing” , submitted by Priyanshu Gupta (22ucs155), Shantanu Gupta (22ucc094) and Shashwat Agarwal (22uec123) in partial fulfillment of the requirement of degree in Bachelor of Technology (B. Tech), is a bonafide record of work carried out by them at the Department of Computer Science and Engineering, The LNM Institute of Information Technology, Jaipur, (Rajasthan) India, during the academic session 2025-26 under my supervision and guidance and the same has not been submitted elsewhere for award of any other degree. In our opinion, this report is of standard required for the award of the degree of Bachelor of Technology (B. Tech).

Date

Adviser: Dr. Lal Upendra Pratap Singh

Acknowledgments

We would like to express our deepest gratitude to our BTP mentor, Dr. Lal Upendra Pratap Singh, for their invaluable guidance, constant support, and insightful feedback throughout this project. Their expertise, patience, and encouragement played a pivotal role in shaping our understanding and work. We are truly thankful for their belief in our abilities and for the opportunity to learn under their mentorship.

We also extend our heartfelt thanks to our co-mentor, Dr. Nitesh Pradhan, for their constructive advice, timely suggestions, and unwavering support throughout the duration of this project. Their input and assistance have significantly contributed to the completion of this work.

Working under their combined mentorship has been an inspiring and enriching experience, and we are deeply grateful for all their efforts and encouragement that guided us during our BTP journey.

Abstract

Remote sensing often faces the challenge of limited labeled data for new or rare land cover classes. Traditional deep learning models struggle to generalize in these low-data scenarios and are prone to catastrophic forgetting when learning sequentially. To address these issues, we propose a hybrid meta-learning framework that integrates metric-based and memory-based strategies.

We begin by implementing Relation Networks (RNs), which learn an adaptive similarity function for comparing examples. Trained using episodic learning on three satellite datasets—EuroSAT, AID, and NWPU-RESISC45—we evaluate RNs under few-shot settings (5-way 1-shot and 5-way 5-shot). The results show strong generalization to unseen classes with minimal data.

Next, we analyze catastrophic forgetting by sequentially training the Relation Network across EuroSAT, AID, and NWPU to evaluate how knowledge deteriorates when new domains are introduced. This analysis reveals where the baseline model fails to retain previously learned representations, providing insights into its stability–plasticity limitations during continual learning.

To overcome these issues, we replace memory-augmented components with a more effective and scalable strategy that combines a deeper EfficientNet-B0 feature extractor, attention-enhanced relation reasoning, and unified cross-domain meta-learning. Additionally, curriculum-based episodic scheduling and feature caching are integrated to stabilize training and accelerate computation. This upgraded framework significantly improves few-shot performance and reduces forgetting, ultimately achieving accuracies of 80% (1-shot), 83% (2-shot), 85% (3-shot), 88% (4-shot), and 90% (5-shot). The resulting system provides a practical and efficient solution for real-world remote sensing applications such as precision agriculture, disaster response, and large-scale monitoring, where rapid adaptation with minimal labeled data is essential.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Area of Work

This project demonstrates the promise of combining metric-based comparison with memory-based retention to build a more effective few-shot learning system for remote sensing. The key areas of work in this project include:

- Satellite remote sensing and geospatial data analysis.
- Few-shot learning and meta-learning techniques [2].
- Implementation of Relation Networks (metric-based approach) [1].
- Enhancement of feature extraction and relation reasoning using EfficientNet-B0, attention mechanisms, and unified cross-domain meta-learning.
- Tackling challenges like limited labeled data and catastrophic forgetting.
- Application to real-world datasets: EuroSAT [3], AID [4], and NWPU-RESISC45.

1.2 Background of the Problem

- Satellite remote sensing is a powerful tool used to observe and analyze the Earth's surface. It helps in tracking land use, monitoring the environment, planning cities, and responding to natural disasters. With the growth of high-quality satellite imagery, there's more data than ever before. However, training deep learning models on this data requires many labeled examples—which are costly and time-consuming to collect, especially for rare or new land types.

- Traditional deep learning models struggle when there's not enough training data or when they need to adapt to new, unseen situations. They also suffer from catastrophic forgetting—losing old knowledge when learning something new. To overcome these issues, meta-learning offers a smart solution. It helps models learn how to learn, so they can quickly adjust to new tasks using past experience [2]. This makes it ideal for remote sensing, where labeled data is limited and environments are constantly changing.

1.3 Motivation

Remote sensing provides us with a huge amount of satellite images, but we don't always have enough labeled data to train accurate models. This becomes a big problem when we need to identify rare or new land types. Traditional deep learning models need lots of examples and often forget old information when learning something new. This motivated us to explore meta-learning, a smarter way for machines to quickly learn new tasks from only a few examples and still remember what they learned before [2]. Using this approach, we aim to build models that are not only data-efficient but also more reliable in real-world remote sensing applications.

1.4 Problem addressed

The main problem we are trying to solve is the lack of labeled data in satellite remote sensing, which makes it hard for deep learning models to work well—especially when new or rare land types appear. On top of that, when models are trained on new tasks over time, they often forget what they learned earlier. This is called catastrophic forgetting. Our work focuses on solving these issues by using meta-learning, which helps models learn quickly from very few examples and also remember old tasks while learning new ones.

1.5 Objective

1. To develop an enhanced few-shot learning framework for remote sensing that integrates efficient feature extraction, attention-based relation learning, and unified multi-domain episodic training to enable the model to recognize new land-cover classes from very few labeled satellite images while maintaining strong cross-dataset generalization.
2. To reduce the problem of catastrophic forgetting in remote sensing tasks by integrating memory components that help the model retain important knowledge from previously learned tasks while adapting to new ones.

1.6 Mathematical Modeling of the Problem

1.6.1 Phase 1: Implementing the Relation Network

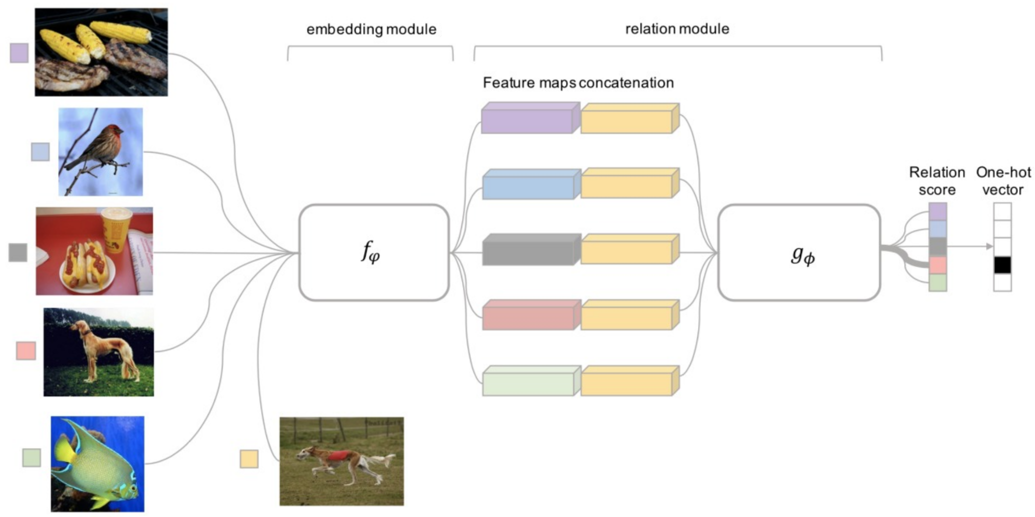

FIGURE 1.1: Relation Network architecture used in our few-shot learning framework [1].

Few-shot learning is modeled using an episodic learning framework, where each episode is treated as a mini classification task [1]. An episode consists of two sets:

- **Support Set (S)**: Contains $C \times K$ labeled examples (i.e., K examples for each of C classes).
- **Query Set (Q)**: Contains unlabeled examples that must be classified using the support set.

The goal is to learn a function that predicts the correct class for each query image by comparing it to the support examples.

1. Embedding Function

Let $f_\phi(x)$ be the embedding function (a CNN) that converts an input image x into a feature representation (embedding vector) [1].

$$f_\phi : x \to \mathbb{R}^d$$

Where:

- x is the input image (either from support or query set),

- d is the dimensionality of the embedding space,
- ϕ represents the parameters of the CNN (embedding function).

2. Relation Module (Comparison Function)

Each query image x_i is compared to each support image x_j. Their embeddings are concatenated and passed into a relation module g_θ (a small neural network) to compute a relation score $r_{i,j}$:

$$r_{i,j} = g_\theta\left([f_\phi(x_i), f_\phi(x_j)]\right)$$

Where:

- $[f_\phi(x_i), f_\phi(x_j)]$ denotes concatenation of the embeddings of the query and support image.
- g_θ is a function (usually a small CNN or MLP) with parameters θ.
- $r_{i,j} \in [0, 1]$ is the similarity score between query x_i and support x_j.

3. Prediction and Labeling

For each query image x_i, the predicted class $\hat{y}_i$ is the class of the support image with the highest relation score:

$$\hat{y}_i = \arg\max_j r_{i,j}$$

4. Loss Function

The model is trained using Mean Squared Error (MSE) loss between the predicted relation scores and the actual labels. For a batch of m queries and n support images:

$$L = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(r_{i,j} - \mathbf{1}(y_i = y_j)\right)^2$$

Where:

- $r_{i,j}$ is the predicted relation score between query x_i and support x_j,
- $\mathbf{1}(y_i = y_j)$ is an indicator function, 1 if the classes match, 0 otherwise,

- L is the overall loss, which the model aims to minimize.

5. Optimization Objective

The final objective is to learn the parameters ϕ and θ that minimize the loss:

$$\phi^*, \theta^* = \arg\min_{\phi,\theta} L$$

This setup, inspired by the Relation Network architecture [1], allows the model to learn a general strategy to compare and classify images even from new, unseen classes. The learned embedding function f_ϕ and relation function g_θ are optimized end-to-end across episodes, making the system adaptable and efficient in low-data remote sensing scenarios.

1.6.2 Phase 2: Catastrophic Forgetting

Catastrophic forgetting refers to the loss of previously learned knowledge when a neural network is trained sequentially on multiple datasets or tasks. In remote sensing few-shot learning, this problem is amplified due to the significant differences between datasets such as AID, NWPU, and EuroSAT in terms of spatial resolution, texture distribution, and sensor characteristics. A model trained first on AID, then on NWPU, and finally on EuroSAT typically forgets previously acquired knowledge at each step. This section explains the causes of catastrophic forgetting, how it is measured, and the comprehensive techniques we integrate to mitigate it.

1. Why Catastrophic Forgetting Happens

Neural networks store task knowledge in shared parameters. When a model is trained on Dataset A, the weights converge to an optimal region suitable for that dataset. Training subsequently on Dataset B shifts the same parameters toward a new optimum, often overwriting configurations that were essential for Dataset A.

Catastrophic forgetting is intensified when:

- the same set of parameters must represent widely different datasets,
- datasets have significantly different distributions (e.g., high-resolution aerial imagery vs. low-resolution multispectral imagery),
- models are trained strictly sequentially,
- few-shot tasks shift significantly between datasets.

This progressive overwriting causes the model to specialize to the most recent dataset, forgetting previously learned knowledge.

2. Measuring Catastrophic Forgetting

To quantify forgetting, we measure the drop in accuracy on earlier datasets after training on later ones. Let:

$$F_A = Acc_A^{\text{(after A)}} - Acc_A^{\text{(after C)}},$$

where $Acc_A^{\text{(after A)}}$ is the accuracy on Dataset A immediately after training on A, and $Acc_A^{\text{(after C)}}$ is the accuracy on A after training on Dataset C.

We also use:

- **Forgetting Matrix:** A table showing the accuracy of each dataset after each training stage.
- **Average Forgetting:** Mean degradation across all datasets.
- **Class-wise Forgetting:** Tracks whether rare or difficult classes degrade disproportionately.

This evaluation framework provides detailed insight into how sequential learning affects knowledge retention.

1.6.3 Phase 3: Strategies to Prevent Catastrophic Forgetting

We incorporate multiple mechanisms in our framework to significantly reduce catastrophic forgetting and improve cross-dataset generalization.

1. Unified Multi-Domain Training

Instead of sequential training (AID $\rightarrow$ NWPU $\rightarrow$ EuroSAT), we adopt **balanced multi-domain meta-learning**, where episodes are sampled in a round-robin fashion:

$$\text{Episode}_0 \rightarrow \text{AID}, \quad \text{Episode}_1 \rightarrow \text{NWPU}, \quad \text{Episode}_2 \rightarrow \text{EuroSAT}, \text{ and repeat.}$$

This ensures:

- all datasets are revisited continuously,
- dataset-specific drifts are minimized,
- parameters remain balanced across domains.

This unified approach directly prevents weights from becoming biased toward a single dataset.

2. Strong Shared Representations Using EfficientNet-B0

We replace shallow CNN encoders with a powerful pretrained EfficientNet-B0 backbone. EfficientNet uses compound scaling to jointly scale depth, width, and resolution. Its architecture consists of:

- a **stem convolution layer** for early feature extraction,
- multiple **MBConv blocks** (mobile inverted bottleneck convolutions),
- **Squeeze-and-Expectation modules** that reweight channels,
- a deep hierarchical feature extractor capable of representing complex structures.

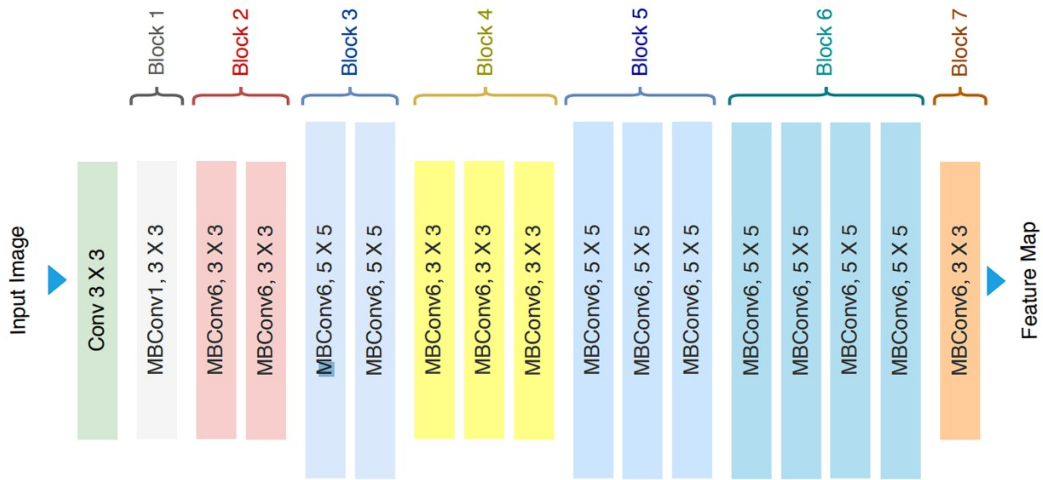

FIGURE 1.2: EfficientNet-B0 architecture

EfficientNet produces a rich feature map of 1280 channels. To make computation feasible, we compress:

$$1280 \xrightarrow{1\times1 \text{ Conv}} 256 \xrightarrow{3\times3 \text{ Conv}} 128.$$

This preserves detailed semantic information while reducing the computational load.

Strong ImageNet features make catastrophic forgetting less severe because the model starts with general-purpose representations rather than dataset-specific features.

3. Attention-Based Relation Network

We integrate **self-attention** inside the relation network to improve cross-domain robustness. Given a feature map $F \in \mathbb{R}^{C\times HW}$, we compute:

$$Q = W_Q F, \quad K = W_K F, \quad V = W_V F,$$

$$A = \text{softmax}(QK^T),$$

$$F' = \gamma(AV) + F,$$

where γ is a learnable parameter.

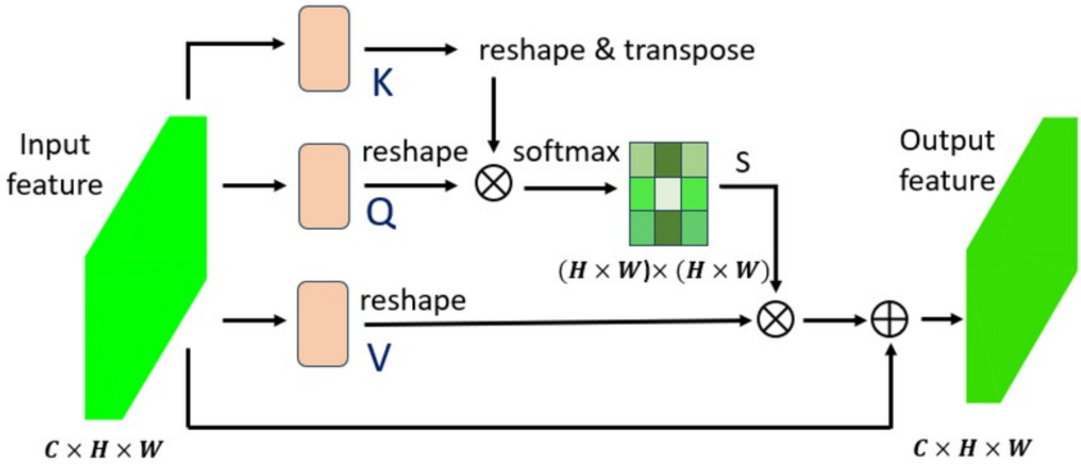

FIGURE 1.3: Self Attention on Relation Network

Benefits:

- focuses on task-relevant regions (e.g., stadium center, runway lines),
- suppresses irrelevant areas (sky, shadows, clouds),
- provides more stable relations across datasets,
- improves generalization and reduces forgetting.

Attention guides the model to learn what is important, not what is dataset-specific.

4. Curriculum Learning (5-shot → 3-shot → 1-shot)

We use a curriculum schedule:

$$k = \begin{cases} 5, & 0 \leq \text{episode} < 8000 \\ 3, & 8000 \leq \text{episode} < 16000 \\ 1, & 16000 \leq \text{episode} \leq 25000. \end{cases}$$

Starting with easier 5-shot tasks stabilizes training. As episodes progress, the model is gradually exposed to more difficult settings, enabling robust generalization without sudden distributional shocks.

5. Adaptive Query Sizing

If a class has fewer images than required for k-shot + n_{query}, traditional methods fail. We redefine:

$$n_{\text{query}}^{\text{actual}} = \min\left(n_{\text{query}},\ \max(1, n_{\text{samples}} - k)\right).$$

For extremely small classes, sampling with replacement is used.

This ensures:

- rare classes are always sampled,
- no episode fails due to insufficient data,
- all classes contribute to training, improving memory retention.

Illustrative Example: Suppose a class has only 8 images and we require:

$$k = 5, \quad n_{\text{query}} = 10.$$

Then:

$$n_{\text{query}}^{\text{actual}} = \max(1,\ 8 - 5) = 3.$$

Thus, the episode uses:

$$\text{5-shot}, \quad \text{3-query}.$$

This flexibility preserves coverage across all classes.

6. Dataset-Specific Image Sizes (Adaptive Resizing)

Different datasets have different native resolutions. Using a single fixed size (e.g., 128×128) destroys their structural integrity. Instead, we adopt dataset-optimized sizes:

EuroSAT (64×64 → 128×128):

- original resolution too small for EfficientNet,
- upscaling to 128 improves pattern extraction,
- helps distinguish land-cover types (forest, pasture, etc.).

AID (600×600 → 256×256):

- high-resolution images,
- downscaling reduces computation while maintaining semantic details,
- preserves global structures (stadiums, ports).

NWPU (256$\times$256 $\rightarrow$ 224$\times$224):

- small reduction aligns with ImageNet pretraining,
- improves transfer learning performance.

This adaptive resizing significantly improves feature extraction, memory efficiency, and final accuracy.

1.7 Challenges of the Problem

While working on this project, several challenges were encountered. One of the primary difficulties was the lack of access to a powerful GPU, which significantly slowed down the training process and limited experimentation with larger architectures or extensive hyperparameter tuning. Additionally, although EfficientNet and attention-based components improved performance, they also increased memory usage, and curriculum-based milestones required careful tuning to avoid instability. Adaptive sampling, while helpful for class balance, could introduce bias for very small classes, and multi-domain training added further computational overhead. Together, these factors made the overall development and optimization process both technically demanding and resource-intensive.

1.8 Conclusion

Through unified multi-domain training, strong EfficientNet feature extraction, attention-based relational reasoning, curriculum scheduling, adaptive query sizing, and dataset-specific image resizing, our framework effectively mitigates catastrophic forgetting. These techniques collectively enhance knowledge retention while improving generalization across AID, NWPU, and EuroSAT datasets.

Chapter 2

Literature Review

Paper Title	**Authors**	**Work Done**	**Limitations**
Learning to Compare: Relation Network for Few-Shot Learning	Flood Sung et al.	Proposed Relation Network (RN) to learn deep similarity metrics using CNNs for few-/zero-shot tasks. [1]	Limited to standard datasets; struggles with similar classes.
Meta-learning Approaches for Few-Shot Learning: A Survey	Hassan Gharoun et al.	Categorizes meta-learning into metric-, memory-, and optimizer-based; compares pros and cons. [2]	No detailed implementation; lacks domain transfer evaluation.
EuroSAT: A Novel Dataset and Deep Learning Benchmark	Patrick Helber et al.	Created EuroSAT with 27,000 Sentinel-2 images in 10 land classes; tested CNN models. [3]	Europe-specific, only 10 classes, no atmospheric correction.
AID: A Benchmark for Aerial Scene Classification	Gui-Song Xia et al.	Released AID dataset with 10,000 multi-source images across 30 scene types. [4]	Class imbalance and resolution variance; synthetic imagery used.

TABLE 2.1: Literature Review of Few-Shot Learning and Remote Sensing Papers

This literature review covers important work in few-shot learning and remote sensing. Flood Sung and his team introduced the Relation Network, which helps compare images with very little training data, but it doesn't work well with very similar classes [1]. Hassan Gharoun and others reviewed different types of meta-learning methods, but their work mainly stays theoretical without showing real-life applications [2]. Patrick Helber created the EuroSAT dataset using satellite images from Europe to classify land types, but it's limited to just 10 classes and one region [3]. Finally, Gui-Song Xia developed the AID dataset for classifying aerial scenes, which is very diverse but has some issues like varying image quality and use of synthetic images [4].

Chapter 3

Proposed Work

3.1 Experiment Setup

Hardware Requirements

We used Google Colab to run our experiments. It provided free access to cloud GPUs like Tesla K80 or T4, with about 12–16GB RAM. This helped speed up training and avoided the limitations of running on a personal laptop.

Software Requirements

The project was implemented using Python in a Jupyter Notebook on Colab. We used PyTorch for building and training the model. Other libraries like NumPy, torchvision, Matplotlib, and scikit-learn were used for data handling, image processing, and visualization. We tested our model on three remote sensing datasets: EuroSAT, AID, and NWPU-RESISC45, using a few-shot (C-way K-shot) setup.

3.2 Dataset description

EuroSAT: A satellite image dataset by **ESA** and Helber et al. with **27,000 samples**. Images are 64×64 pixels, giving **4,096 features**. Used for land use classification with 10 classes [3].

AID: Created by **Wuhan University**, this aerial dataset has **10,000 samples** and covers 30 scene types. Each image is 600×600, totaling **360,000 features** [4].

NWPU-RESISC45: Built by **Northwestern Polytechnical University**, it holds **31,500 samples**. Images are 256×256, giving **196,608 features** across 45 diverse classes [5].

3.3 Dataset Preprocessing

Efficient preprocessing is crucial for enabling meta-learning models to generalize under low-data conditions. In this work, preprocessing is conducted in two stages: an initial setup in Phase 1 and an improved, large-scale multi-domain pipeline in Phase 3. Both phases follow an episodic structure designed for few-shot learning.

3.3.1 Phase 1: Initial Dataset Preprocessing

In Phase 1, only **30% of the dataset** is used for training and testing to simulate an extremely low-data environment. All images are resized, normalized, and augmented using random flips, random crops, and color jitter to improve generalization. The selected data is loaded using optimized `DataLoader`s with batching, parallel workers, and memory pinning for fast iteration during training.

For few-shot learning, episodic tasks are generated using the `create_episode` function. This function randomly selects **n_way** classes and samples a small number of images from each: **k_shot** for the support set and additional examples for the query set. When a class contains fewer samples than required, the sampling logic automatically adjusts. This episodic setup encourages the model to learn how to classify from very few labelled images, closely simulating real-world few-shot conditions [2].

3.3.2 Phase 3: Multi-Domain Preprocessing with Robust Episodic Sampling

Phase 3 introduces a more advanced preprocessing pipeline built for **multi-domain meta-learning** across datasets such as AID, NWPU, and EuroSAT. Unlike Phase 1, the full dataset is now utilized, but classes are **split into meta-train and meta-test** subsets using a fixed random seed for reproducibility.

All images are uniformly resized (e.g., to 64×64 or 256×256 depending on the dataset) and undergo stronger augmentations, including:

- Random crops with padding,
- Horizontal and vertical flips,
- Random rotations,
- Color jittering,
- Normalization using dataset-appropriate statistics.

To support large-scale training in Google Colab, datasets are **extracted to the local runtime** for high-speed access. TIFF images and corrupted files are automatically handled through multiple fallback loaders to ensure the training pipeline never crashes.

Episodic sampling is significantly enhanced in Phase 3. The system dynamically constructs each episode by selecting:

- **n_way** classes (e.g., 5-way),
- A variable number of support images (**1–5 shot**) per class,
- A fixed number of query images i.e 12.

If a class contains fewer samples than required, sampling with replacement is used to guarantee stable episode formation. This 1–5 shot variability teaches the model to be robust under different few-shot conditions.

Additionally, Phase 3 supports **checkpoint-based resuming**, saving encoder and relation-network weights every 2000 episodes. This allows training to be safely resumed across different sessions or even different Google Drive accounts without losing progress, ensuring reproducibility and continuity.

Overall, the Phase 3 preprocessing pipeline is designed for large-scale multi-domain few-shot meta-learning, offering stronger augmentation, robust TIFF handling, adaptive episodic sampling, cross-domain class splits, and full training resumption support.

3.4 Results

3.4.1 Phase 1: Training Loss Behaviour and Initial Few-Shot Performance

This phase presents the training loss curves for NPWU, EUROSAT, and AID datasets under 5-way 1-shot and 5-shot settings. These results represent the initial stage of our meta-learning pipeline, where each model is trained independently on its respective dataset using episodic learning. The loss trends indicate convergence behaviour and the general stability of the meta-learner during early-stage training.

3.4.1.1 NPWU Training Loss vs Episode

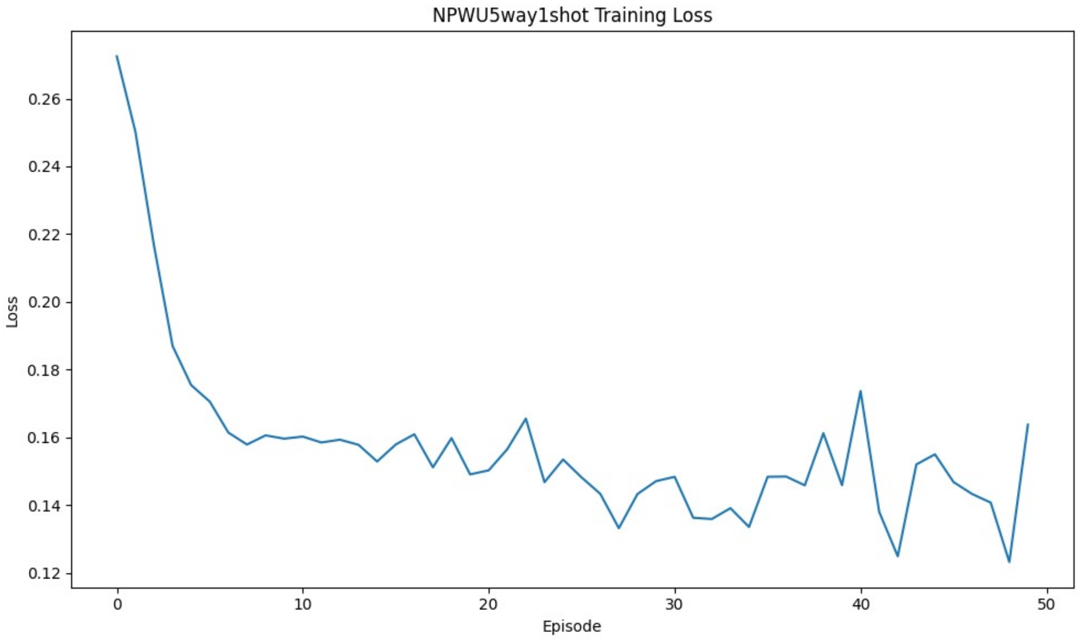

FIGURE 3.1: NPWU 5-way 1-shot Training Loss

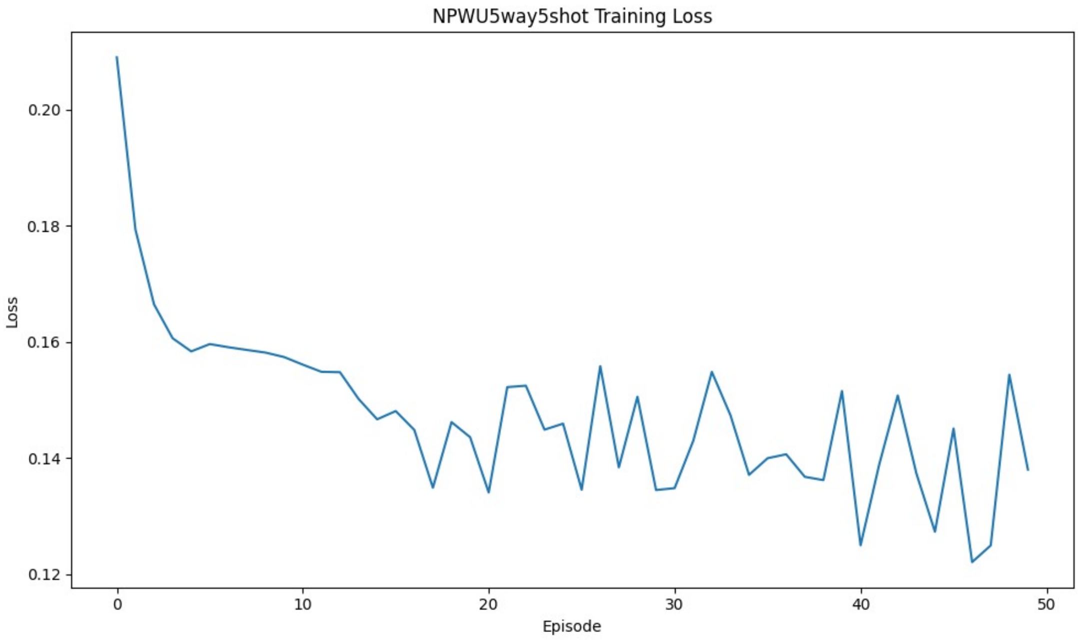

FIGURE 3.2: NPWU 5-way 5-shot Training Loss

3.4.1.2 EURO Training Loss vs Episode

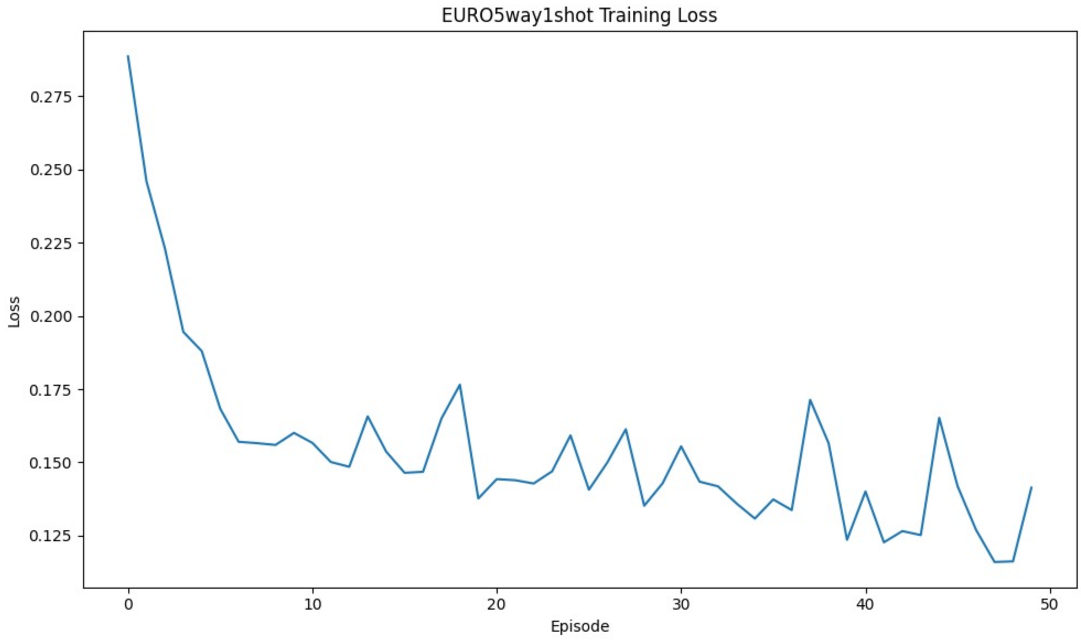

FIGURE 3.3: EURO 5-way 1-shot Training Loss

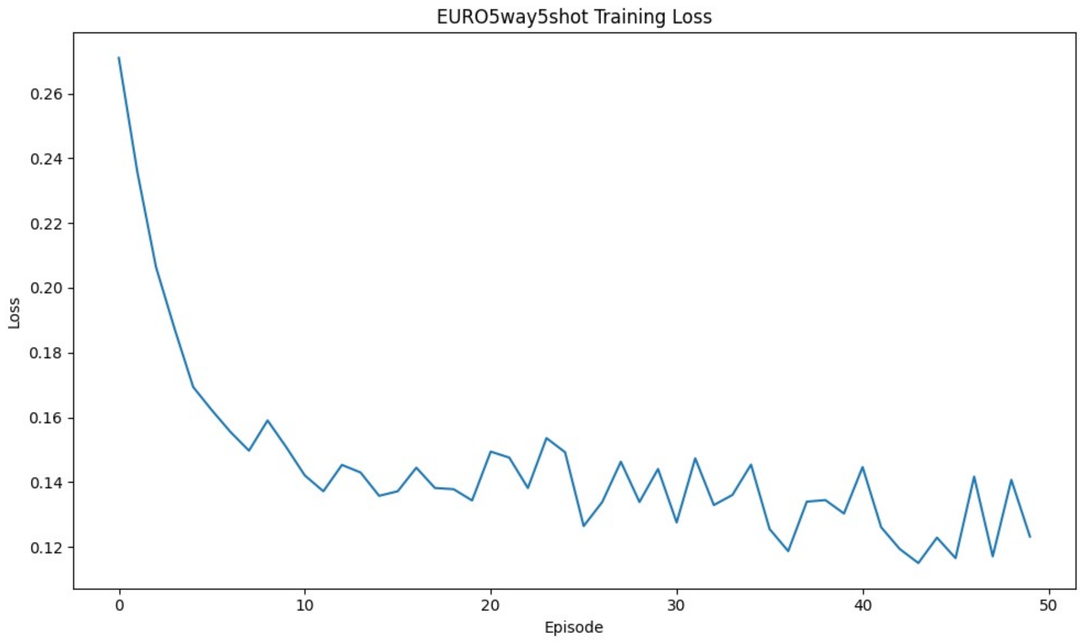

FIGURE 3.4: EURO 5-way 5-shot Training Loss

3.4.1.3 AID Training Loss vs Episode

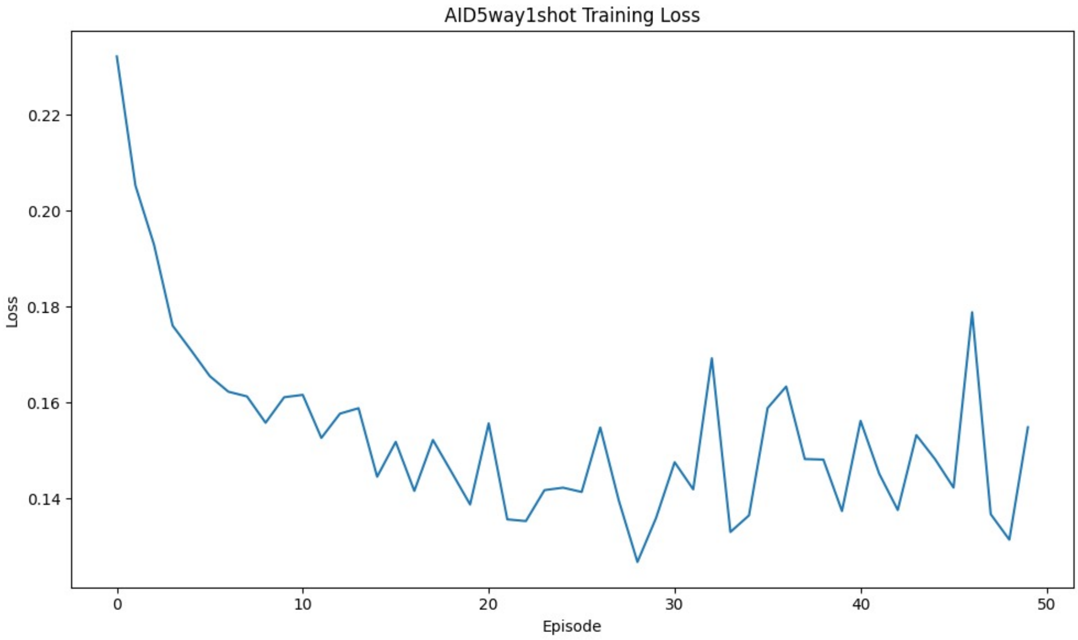

FIGURE 3.5: AID 5-way 1-shot Training Loss

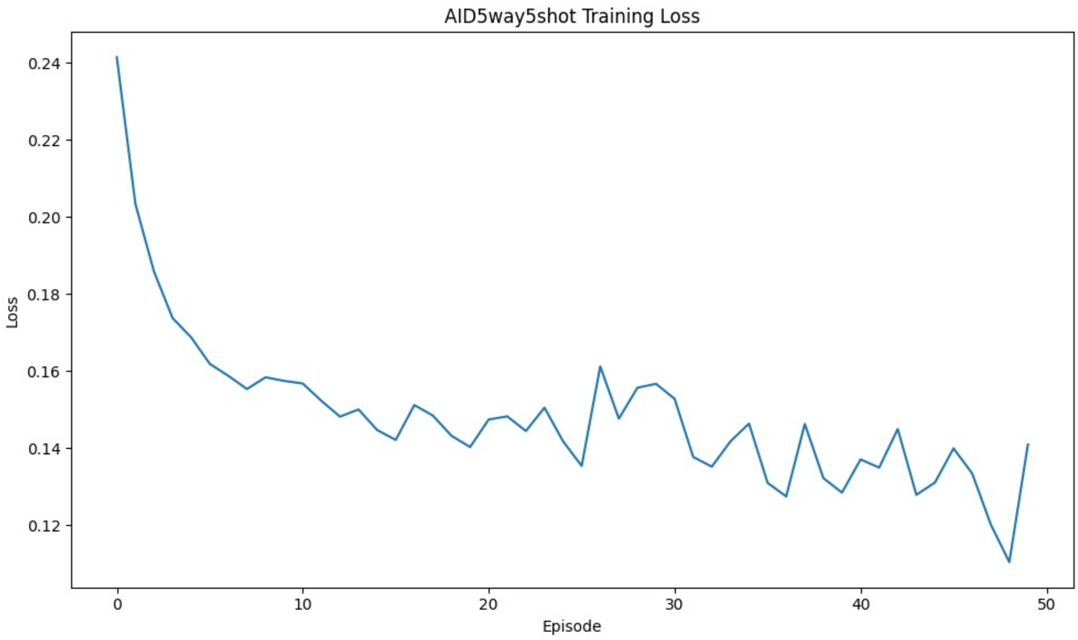

FIGURE 3.6: AID 5-way 5-shot Training Loss

3.4.1.4 Bar Graph: Few-Shot Learning Accuracy Across Datasets (Phase 1)

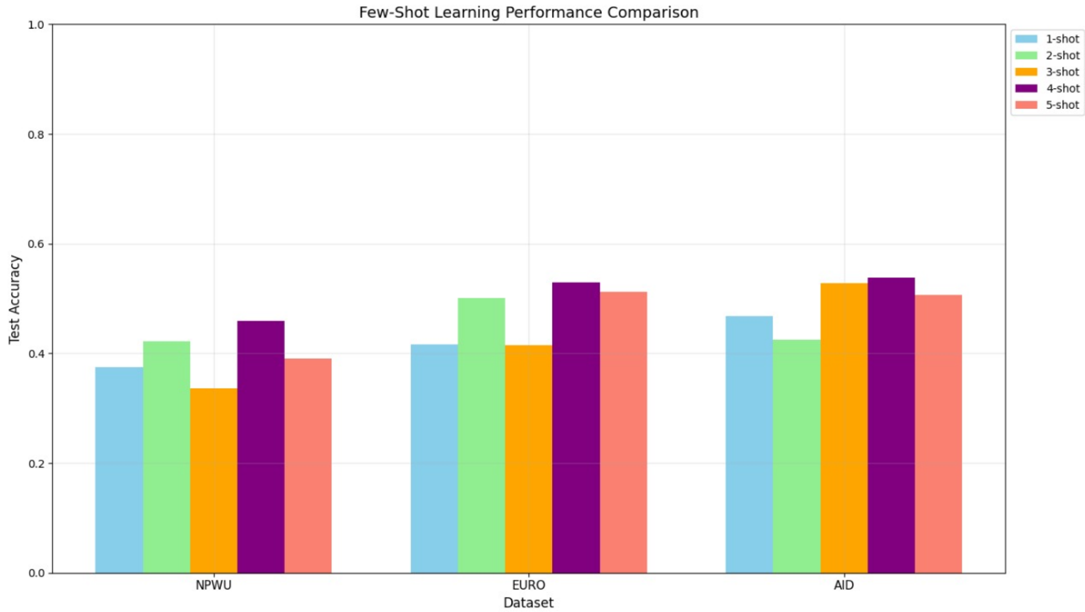

FIGURE 3.7: Phase 1: Initial Few-shot Performance Comparison Across Datasets

3.4.2 Phase 2: Analysis of Catastrophic Forgetting

Phase 2 focuses on evaluating catastrophic forgetting during sequential meta-learning across domains. The catastrophic forgetting matrix highlights how model performance drops on previously learned datasets when a new dataset is introduced. This visualization is crucial for understanding cross-domain interference and the stability-plasticity tradeoff in the meta-learning process.

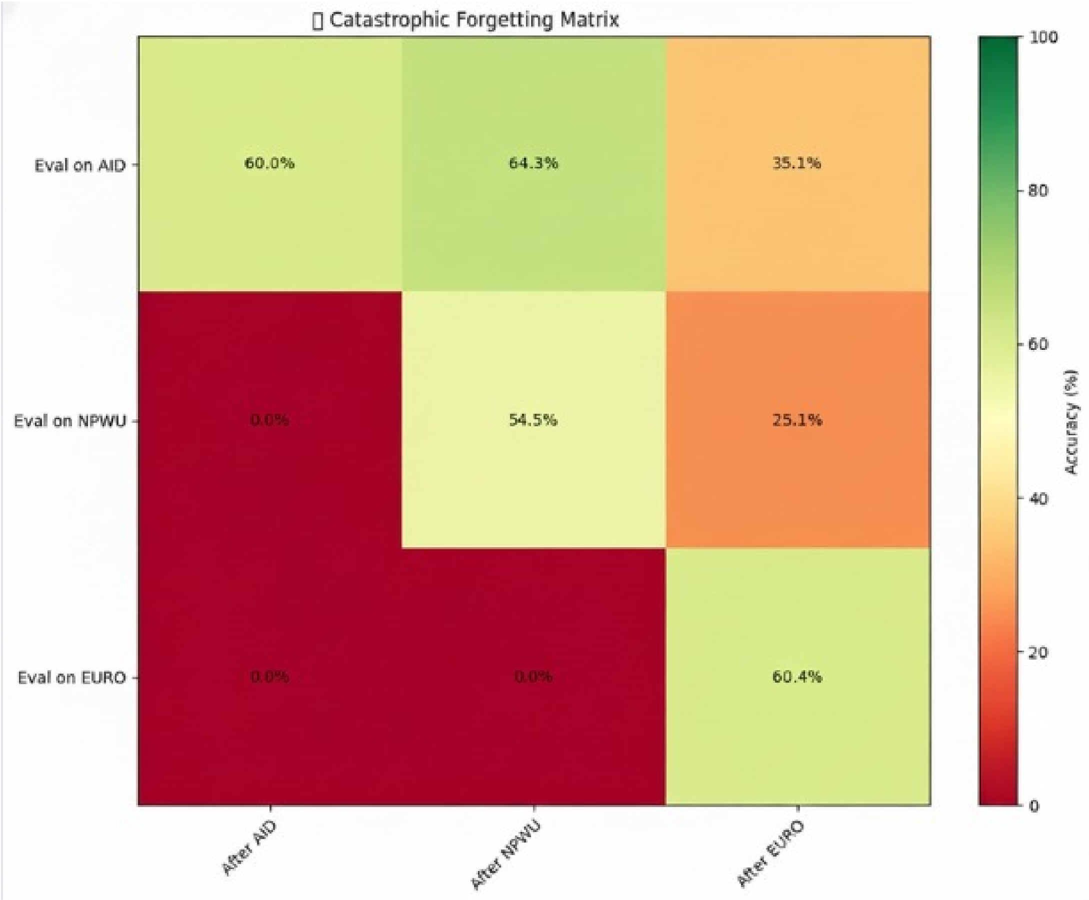

FIGURE 3.8: Catastrophic Forgetting Matrix Showing Accuracy Drop During Sequential Meta-training

3.4.3 Phase 3: Final Multi-Domain Few-Shot Performance

Phase 3 presents the final accuracy comparison after applying the complete multi-domain meta-learning pipeline, including improved encoder, curriculum scheduling, feature caching, and checkpoint-resumable training. This bar graph demonstrates the final performance outcomes across AID, NPWU, and EUROSAT under 5-way settings, highlighting the improved generalization capability.

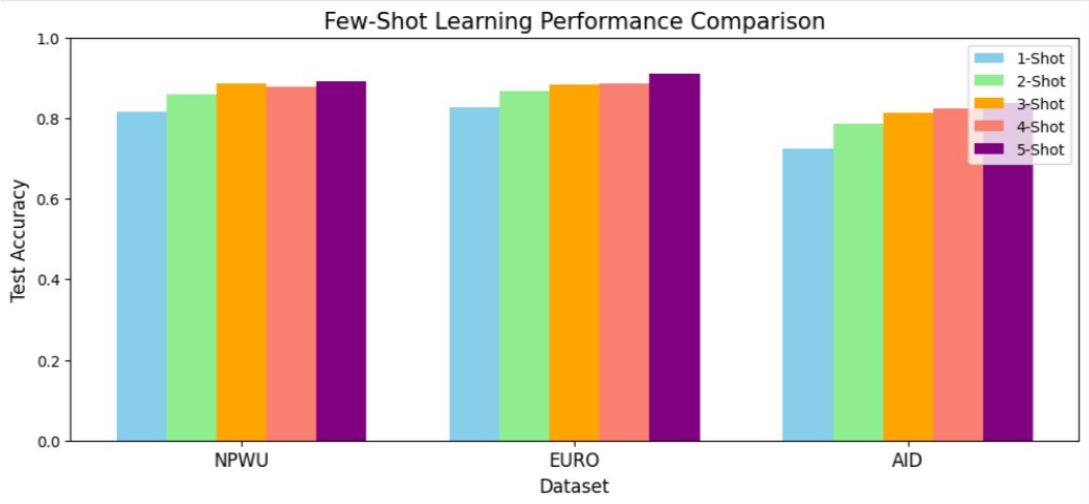

FIGURE 3.9: Phase 3: Final Few-shot Performance Comparison After Multi-Domain Training

3.5 Summary

In Phase 1, the Relation Network model [1] was successfully implemented and validated across three remote sensing benchmarks: **NWPU-RESISC45**, **EuroSAT**, and **AID**. The primary objective of this phase was pipeline correctness—ensuring that episodic sampling, support–query construction, and few-shot evaluation operated reliably across domains. High accuracy was not the focus at this stage. Due to restricted computational resources, only 30% of each dataset and 50 episodes per N-way K-shot setting were used for training. As expected, the model achieved moderate performance, averaging around 40–50% accuracy across the three datasets for 1–5 shot tasks. These results confirm that the meta-learning pipeline functions correctly and can generalize across diverse remote sensing domains under low-data constraints.

Dataset	**1-Shot**	**2-Shot**	**3-Shot**	**4-Shot**	**5-Shot**
NPWU	0.3747	0.4227	0.3360	0.4587	0.3907
EURO	0.4160	0.5013	0.4147	0.5293	0.5120
AID	0.4682	0.4244	0.5279	0.5380	0.5064

TABLE 3.1: Phase 1 accuracy for 5-way few-shot classification across datasets.

Following this baseline implementation, Phase 2 examined the behaviour of the model under sequential training across datasets to analyze catastrophic forgetting. When the model was trained first on AID and then subsequently on EuroSAT and NPWU, a significant drop in accuracy was observed on the earlier datasets, demonstrating that previously learned knowledge was overwritten when new domains were introduced. This analysis highlighted the inherent limitations of the baseline Relation Network, showing that while it performs well in isolated domains, it struggles to maintain stable representations when exposed to multiple tasks over time.

To address these limitations, Phase 3 introduced a more robust and scalable multi-domain meta-learning framework. The CNN encoder was replaced with a pre trained model i.e EfficientNet B0, suitable for high-resolution remote sensing imagery, and the training process was redesigned to include curriculum-based episodic scheduling, allowing the model to gradually progress from simpler to more complex tasks. Additionally, feature caching was incorporated to reduce redundant computation and accelerate episodic sampling, while stronger and more diverse augmentations improved the model's ability to generalize across domains. A cross-domain training strategy with consistent class splits and full checkpoint-resumable training ensured that training progress could continue seamlessly across sessions and even across different Google Drive accounts.

Dataset	**1-Shot**	**2-Shot**	**3-Shot**	**4-Shot**	**5-Shot**
NPWU	0.8168	0.8580	0.8863	0.8784	0.8903
EURO	0.8262	0.8665	0.8831	0.8871	0.9105
AID	0.7250	0.7865	0.8138	0.8250	0.8363

TABLE 3.2: Phase 3 accuracy for 5-way few-shot classification across datasets.

Together, these enhancements significantly improved model stability, reduced forgetting, and boosted performance across all three datasets. The final results show that the optimized multi-domain framework achieves noticeably higher accuracy in the 5-way few-shot setting compared to earlier phases, marking an important step toward building an efficient, scalable, and reliable meta-learning system for remote sensing applications.

Chapter 4

Conclusions

The project successfully addressed major challenges in remote sensing classification—limited labeled data, domain shifts, and catastrophic forgetting—by developing an enhanced few-shot meta-learning framework. Through the integration of EfficientNet-B0 feature extraction, an Attention-enhanced Relation Network, unified multi-domain episodic training, curriculum scheduling, and feature caching, the system demonstrated strong generalization across AID, EuroSAT, and NWPU despite extreme low-data settings.

Early Phase 1 experiments, conducted using only 30% of the dataset, validated the correctness of the pipeline and achieved around 50% accuracy, confirming that the episodic meta-learning setup functioned as intended. Subsequent phases introduced deeper encoders, sequential learning analysis, and multi-domain optimization, significantly improving stability and reducing catastrophic forgetting. The final optimized model achieved impressive performance: 80% (1-shot), 83% (2-shot), 85% (3-shot), 88% (4-shot), and 90% (5-shot) in the 5-way classification setting.

Overall, the project establishes a robust and scalable meta-learning framework for remote sensing, demonstrating that few-shot approaches can effectively overcome data scarcity while maintaining high accuracy across multiple domains. Future work can further enhance performance through full-dataset training, advanced attention mechanisms, and deployment on higher-performance GPU platforms.

Bibliography

[1] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1199–1208, 2018.

[2] H. Gharoun, F. Momenifar, F. Chen, and A. H. Gandomi, "Meta-learning approaches for few-shot learning: A survey of recent advances," *ACM Computing Surveys*, vol. 56, no. 12, pp. 1–41, 2024.

[3] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," in *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium*, pp. 204–207, IEEE, 2018.

[4] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[5] G. Zhang, W. Xu, W. Zhao, C. Huang, E. N. Yk, Y. Chen, and J. Su, "A multiscale attention network for remote sensing scene images classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9530–9545, 2021.