High-Performance Lightweight Vision Models for Efficient Land Cover Classification with Coresets & Compression

IIT Madras Zanzibar

Tushar Shinde, IIT Madras Zanzibar

ICML International Conference On Machine Learning

Our Framework

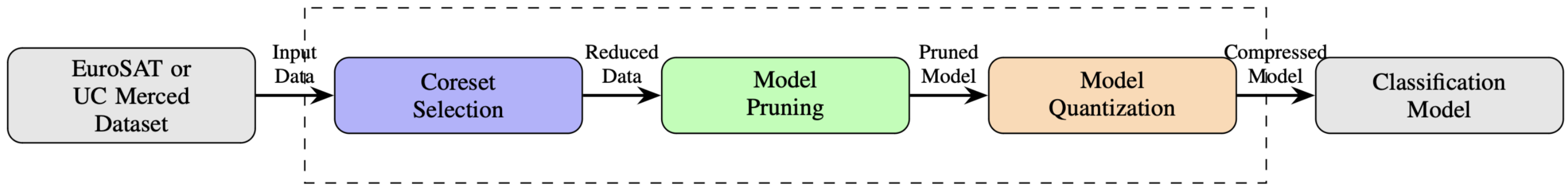

Fig. 1. Proposed framework: coreset selection + adaptive model compression.

Introduction

Background: Land cover classification using satellite imagery is vital for environmental monitoring and urban planning. Deployment of deep learning models on edge devices faces memory and compute limitations.

Key Insight: We present an efficient framework reducing training data and model size while maintaining >92% accuracy.

Challenges:

- Limited generalization across diverse sensors and environments.
- High resource demands prevent edge deployment.
- Efficiency in both training and inference is essential.

Contributions

- Unified framework combining coreset selection and adaptive model compression.
- Achieved competitive accuracy with reduced training data and model size.
- Highlight: **98.10% Accuracy with 9× Compression Ratio on UC Merced.**
- Over 92% accuracy with up to 6× model size reduction using 10% training data.

Method and Mathematical Framework

Our framework optimizes both coreset selection and model compression.

Coreset Selection: Given dataset $\mathcal{D}$, select subset $\mathcal{C}$ of size $\alpha \in (0, 1]$ such that $L(\mathcal{C}) \approx L(\mathcal{D})$.

- **Random:** $\mathcal{C}_{\text{random}} = \{(x_i, y_i)\}_{i \in \mathcal{S}}, \quad |\mathcal{S}| = \alpha N.$
- **Forgetting-based:** $\mathcal{C}_{\text{forget}} = \{(x_i, y_i) \mid f_i \in \text{Top-}M\}, \quad f_i = \#$ forget events.
- **Margin-based:** $m_i = p_i^{(1)} - p_i^{(2)}, \quad \mathcal{C}_{\text{margin}} = \{(x_i, y_i) \mid m_i \in \text{Bottom-}M\}.$

Model Compression: Minimize model size while maintaining accuracy:

$$\tilde{\theta} = \mathcal{Q}(\mathcal{P}(\theta))$$

- **Pruning:** $\mathcal{M}_i = \{w \in W_i \mid |w| < \tau_i\}, \quad w \in \mathcal{M}_i \Rightarrow w = 0.$
- **Quantization:** $\hat{w}_i = \text{clip}\left(\text{round}\left(\frac{w_i}{\Delta}\right), q_{\min}, q_{\max}\right) \cdot \Delta.$
- **Adaptive:** Layer-wise pruning and quantization guided by importance metrics.

Datasets and Experimental Setup

- **Datasets:**
  - EuroSAT: 27K RGB images, 10 classes, 64×64.
  - UC Merced: 2.1K images, 21 classes, 256×256.
- **Models:** ConvNeXt-Tiny (28M), Swin-Tiny (28M), EfficientNetV2-S (24M), RegNetY-3.2GF (27M).
- **Training Setup:** Batch size: 64, epochs: 10, Optimizer: AdamW, LR: 0.0005, Loss: cross-entropy with class weights.
- **Coreset Fractions:** 100%, 10%, 5%.
- **Compression Methods:**
  - Pruning: fixed ($k \in \{1.5, 1.0, 0.5, 0.25, 0\}$), Adaptive Pruning (LAP).
  - Quantization: fixed-bit (8, 4, 2, 1), Adaptive Quantization (LAQ).
- **Metrics:** Accuracy, Compression Ratio (CR).

Results and Analysis

We present a comprehensive evaluation of our framework across datasets, coreset selection strategies, and compression settings

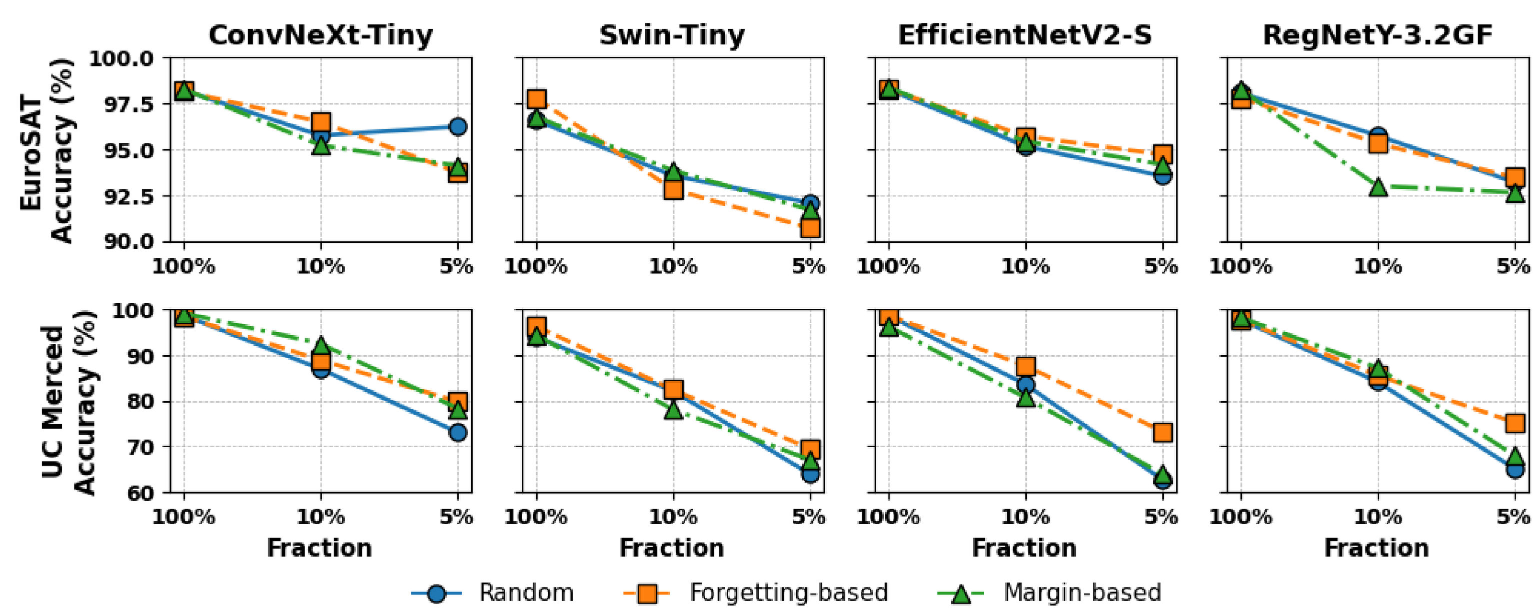

Fig. 2. Accuracy vs. coreset fraction.

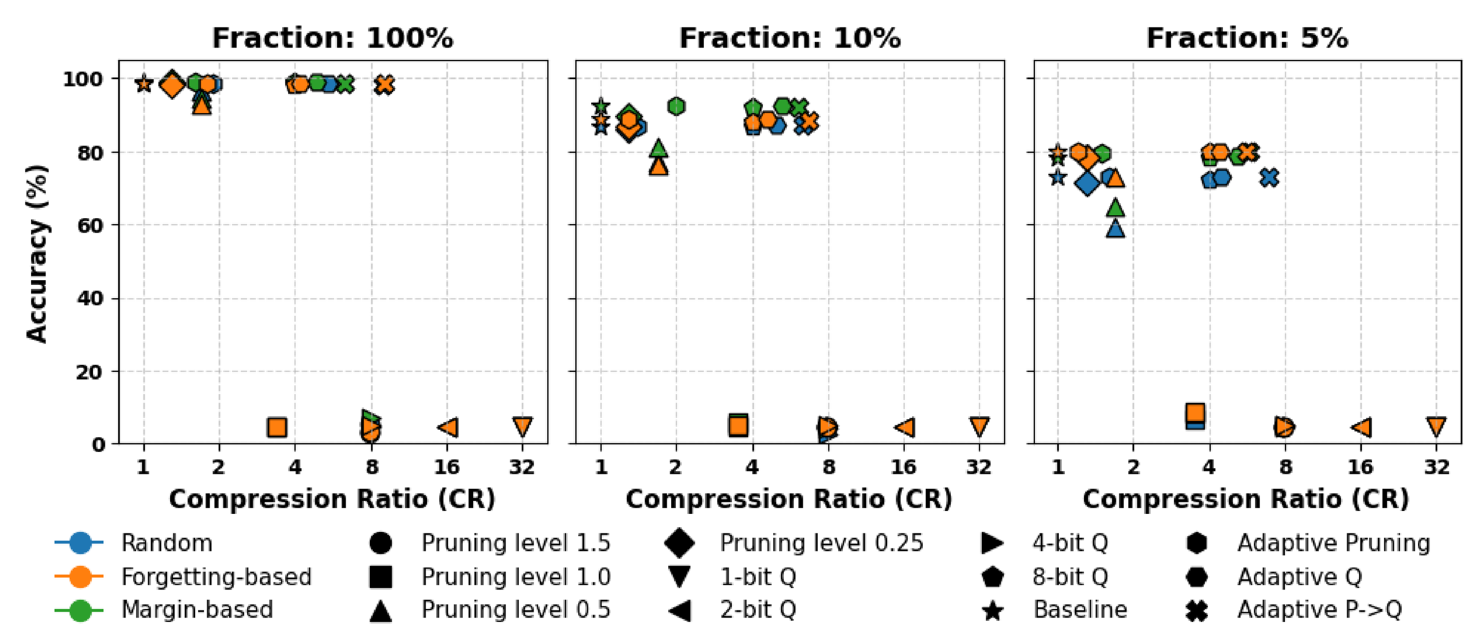

Fig. 3. Accuracy vs. Compression Ratio (ConvNeXt-Tiny, UC Merced).

Key Results:

- Forgetting-based coreset:
  - 96.46% (EuroSAT, 10%)
  - 88.81% (UC Merced, 10%)
- Margin-based coreset:
  - 94.09% (EuroSAT, 5%)
  - 78.33% (UC Merced, 5%)
- Adaptive compression:
  - 98.10% accuracy, 9× CR (UC Merced)
- Better trade-off vs. SwinV2-Tiny baseline [3].

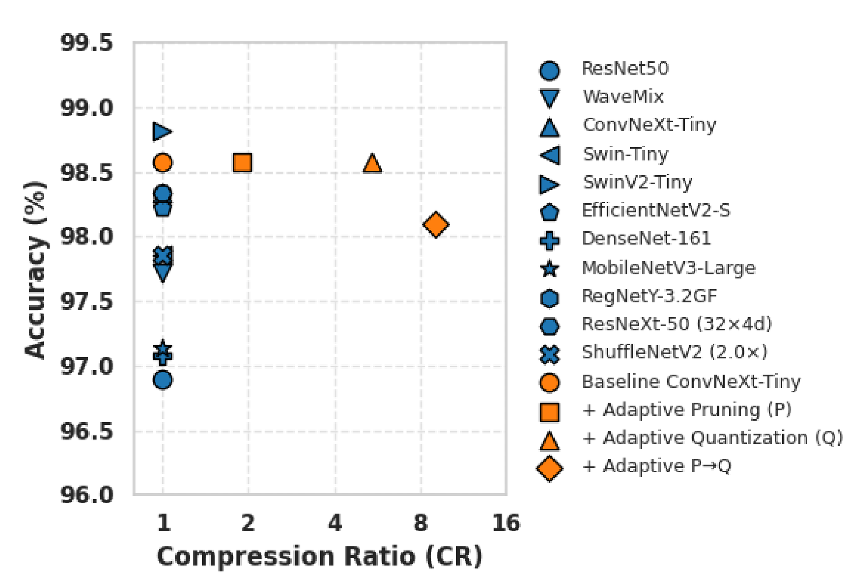

Fig. 4. Comparison with existing models on UC Merced.

Conclusions

- Coreset selection minimizes training data with <5% accuracy drop.
- Adaptive compression balances size reduction and performance.
- Enables efficient land cover classification on resource-constrained devices.
- Highlight: **98.10% Accuracy with 9× CR.**

Future Work:

- Joint coreset-compression optimization for edge devices.
- Integration of multi-modal data (optical + SAR).

References

1. Helber, P., et al. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE JSTARS, 2019.
2. Yang, Y. and Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. SIGSPATIAL GIS, 2010.
3. Jeevan, P. and Sethi, A. Which backbone to use: A resource-efficient domain specific comparison for computer vision. arXiv 2024.
4. Shinde, T. Adaptive quantization and pruning of deep neural networks via layer importance estimation. NeurIPS 2024 Workshop.