GEOMETRY-PRESERVING CORESETS FOR QUANTIZED FOUNDATION MODELS IN REMOTE SENSING

Tushar Shinde
IIT Madras Zanzibar, Tanzania
{shinde}@iitmz.ac.in

ABSTRACT

We reveal a fundamental yet overlooked coupling in foundation model deployment: *data selection and quantization cannot be optimized independently*. Through comprehensive experiments on remote sensing classification under extreme constraints (5% labeled data, INT8/binary quantization), we demonstrate that standard coreset selection strategies, while effective at full precision, suffer catastrophic accuracy collapse once models are quantized, with binary networks degrading to near-chance performance. This failure occurs because conventional methods prioritize decision uncertainty while ignoring representation geometry, which quantization fundamentally distorts. We introduce *Entropy-Based Density-Weighted Coresets* (EntropyBDWC), a geometry-aware selection strategy that explicitly preserves local embedding structure under discretization. Evaluated across three datasets, four architectures, and multiple precision regimes, EntropyBDWC consistently outperforms entropy-based and random sampling under INT8 quantization and substantially stabilizes binary networks. Critically, we show that performing selection in frozen foundation model embeddings (DINO) amplifies this robustness, establishing a new role for foundation models as *data coresets* rather than trainable backbones. Our work establishes that quantization-aware data curation is not optional but essential, with implications extending beyond remote sensing to any resource-constrained deployment of foundation models.

1 INTRODUCTION

Foundation model deployment is increasingly constrained by two orthogonal pressures: limited labeled data for task adaptation and aggressive quantization for edge inference. The conventional approach treats these as independent optimization problems, first select informative training data, then compress the trained model. Our work demonstrates that *this independence assumption is fundamentally wrong*. Data selected to maximize full-precision accuracy can induce representations that collapse catastrophically under quantization, particularly for binary neural networks where accuracy often degrades to chance levels.

Existing coreset selection strategies prioritize decision-boundary uncertainty Sener & Savarese (2017); Shinde; Shinde & Madabhushi, identifying samples where the classifier is least confident. While this maximizes informativeness under full-precision training, it ignores a critical fact: quantization is not a uniform noise process but a *geometry-altering transformation* that compresses margins, amplifies perturbations, and fundamentally reshapes feature space. Uncertainty-selected coresets are often geometrically fragile, high-entropy samples near decision boundaries that, once quantized, become uninterpretable or redundant. This fragility is invisible at FP32 but catastrophic at 1-bit precision.

We reframe coreset selection as the problem of *preserving representation geometry under quantization*. Rather than selecting samples that challenge the classifier, we select samples that maintain distributional coverage and local density structure, properties that remain meaningful even after aggressive discretization. Our method, EntropyBDWC, couples predictive uncertainty with embedding density, ensuring selected subsets span the feature manifold while avoiding redundancy. Crucially, we perform this selection in frozen foundation model embeddings (DINO), which provide a quantization-invariant geometric scaffold that transcends task-specific overfitting.

Contributions.

- We establish that data selection and quantization are inseparable design choices, not sequential optimizations. Standard uncertainty sampling fails catastrophically under aggressive quantization.
- We propose EntropyBDWC, which explicitly couples decision uncertainty with local density to maintain feature geometry post-quantization. This simple principle yields consistent gains across datasets, architectures, and precision regimes.
- We demonstrate that frozen self-supervised embeddings (DINO) outperform task-native features for coreset selection, revealing a new role for foundation models in controlling *what to train on* rather than *how to train*.
- We provide the first controlled evaluation of remote sensing classification under joint data scarcity and post-training quantization, spanning multiple architectures, datasets, budgets, and precision levels.

2 METHODOLOGY

We formalize the coupled optimization: given a labeled dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ and budget $k \ll N$, select coreset $\mathcal{C} \subset \mathcal{D}$, $|\mathcal{C}| = k$, such that a model trained on $\mathcal{C}$ retains high accuracy *after post-training quantization*. We target three deployment regimes: FP32, INT8, and binary neural networks (BNN). Unlike classical coreset selection, which assumes FP32 deployment Sener & Savarese (2017); Mirzasoleiman et al. (2020), we require *quantization-stable coresets* whose induced representations remain separable once weights are discretized. This reframes the problem: coreset quality is measured not by training loss but by *post-quantization generalization*.

Standard entropy-based selection ranks samples by predictive uncertainty: $H(x_i) = -\sum_{c=1}^{C} p(c \mid x_i) \log p(c \mid x_i)$. High-entropy samples near decision boundaries are informative for gradient-based optimization but geometrically unstable, they occupy low-density regions where quantization noise has maximal impact. Moreover, entropy ignores redundancy: selecting all high-uncertainty samples from a cluster yields no additional geometric coverage.

We preserve representation geometry by jointly modeling decision uncertainty and local embedding density. For each sample, we extract embedding $z_i \in \mathbb{R}^d$ from either the penultimate layer (BASE) or frozen DINO encoder. Local density is estimated via K-nearest neighbors:

$$\rho(x_i) = \left(\frac{1}{K} \sum_{j \in \mathcal{N}_K(i)} d_{\cos}(z_i, z_j) \right)^{-1}, \quad S(x_i) = H(x_i) \cdot \rho(x_i), \tag{1}$$

where $\mathcal{N}_K(i)$ denotes the K nearest neighbors of x_i in normalized embedding space and $d_{\cos}$ is cosine distance. The product $S(x_i)$ prioritizes samples that are both decision-critical (high H) and locally representative (high ρ), yielding coresets that maintain manifold coverage under quantization-induced geometry distortion.

DINO embeddings provide a task-agnostic, quantization-invariant geometry. Self-supervised models learn representations optimized for distributional structure rather than classifier confidence, making them robust to downstream precision constraints. By selecting coresets in DINO space, we decouple data curation from task overfitting and position foundation models as *data selection coresets*, controlling what to learn rather than how to learn it. This reframes foundation model utility: they need not be fine-tuned to be valuable.

3 EXPERIMENTAL SETUP

Datasets. We evaluate our method on three widely used remote sensing benchmarks covering binary and multi-class land-use classification under varying levels of visual complexity and spatial heterogeneity Penatti et al. (2015); Helber et al. (2019).

The *Brazilian Coffee Scenes* dataset Penatti et al. (2015) focuses on fine-grained agricultural pattern recognition and is treated as a binary classification task (2 classes, ∼2,800 images). The *Joinville*

Table 1: Quantization sensitivity across datasets and architectures under full-data training. Comparison of FP32, INT8, and BNN accuracies for four backbones on three remote sensing datasets using full training data (ratio = 1.0, input size = 96).

Model	**Brazilian Coffee**			**Joinville Urban**			**EuroSAT**		
	FP32	INT8	BNN	FP32	INT8	BNN	FP32	INT8	BNN
ResNet-18	89.2	88.3	47.5	89.8	92.9	32.7	98.0	97.1	9.3
ConvNeXt-Tiny	87.5	90.8	47.5	93.9	86.7	32.7	98.4	98.6	7.4
Swin-Tiny	91.7	91.7	52.5	89.8	91.8	67.3	98.2	98.7	9.3
MobileViT-v2-S	90.8	85.0	52.5	93.9	92.9	32.7	98.3	95.6	11.1

Urban dataset addresses urban versus non-urban scene discrimination (2 classes, $\sim$3,000 images), while the *EuroSAT* dataset Helber et al. (2019) consists of Sentinel-2 imagery spanning 10 land-use and land-cover categories ($\sim$27,000 images). For Joinville Urban and EuroSAT, we adopt the official training and testing splits to ensure strict reproducibility.

Architectures. We consider four representative convolutional and transformer-based backbones: ResNet-18 He et al. (2016), ConvNeXt-Tiny Liu et al. (2022), Swin-Tiny Liu et al. (2021), and MobileViT-v2-S Mehta & Rastegari (2021). These architectures span distinct inductive biases, depth-to-width trade-offs, and receptive field structures, allowing us to probe how architectural design interacts with data selection and quantization sensitivity Gholami et al. (2022). All models are initialized from ImageNet-pretrained weights, and the final classification layer is replaced to match the number of classes.

Implementation Details. All experiments are implemented in PyTorch 2.x and executed on a single NVIDIA P100 GPU. Models are trained using the Adam optimizer with a learning rate of 10^{-4} and cross-entropy loss. Each configuration is trained for 10 epochs without early stopping, intentionally operating in a low-compute regime consistent with prior work on data-efficient learning.

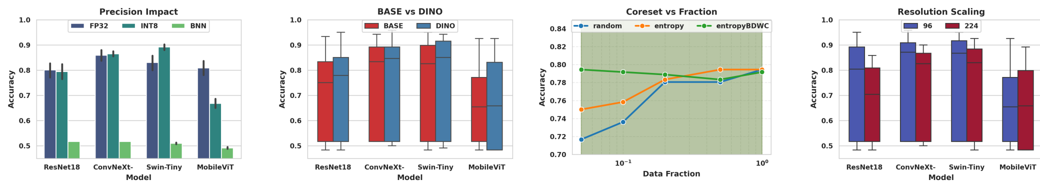
Coreset Budgets and Selection Strategies. We evaluate four data budgets corresponding to 5%, 10%, 20%, and 100% of the available training data. Three selection strategies are compared: uniform random sampling, entropy-based uncertainty sampling Gal & Ghahramani (2016), and the proposed EntropyBDWC method. EntropyBDWC jointly scores samples using predictive uncertainty and local embedding density, encouraging selection of samples that are both informative and geometrically representative. Coreset selection is performed either in the model's penultimate feature space (BASE) or in a frozen DINOv1 embedding space Caron et al. (2021), enabling a direct comparison between task-specific and foundation-model-derived representations.

Precision Settings and Evaluation Metrics. We evaluate all models under three precision settings: full-precision FP32, post-training INT8 quantization Jacob et al. (2018), and binary neural networks (BNN) obtained via naive post-hoc layer-wise binarization Courbariaux et al. (2016). Performance is measured using classification accuracy on a held-out test set.

4 RESULTS

Quantization Effects under Full-Data Training. Table 1 reports the impact of post-training quantization across datasets and architectures using the full training set. Across all models, INT8 quantization preserves FP32-level accuracy and in several cases yields marginal improvements. However, naive binary quantization leads to severe performance degradation across datasets, with accuracy often collapsing toward chance levels, a known failure mode when representational geometry is not preserved.

Coreset Selection under Joint Data Scarcity and Quantization. Table 2 examines the interaction between data budget, coreset strategy, embedding space, and precision on the Brazilian Coffee Scenes dataset. At low data fractions (5–10%), entropy-based sampling consistently outperforms random selection. However, entropy-only selection exhibits unstable behavior at higher budgets and under quantization, indicating that uncertainty alone is insufficient to guarantee representational robustness once feature geometry is distorted by low-bit precision.



(a) **Precision impact.** (b) **BASE vs DINO.** (c) **Coreset vs fraction.** (d) **Resolution scaling.**

Figure 1: Ablation of precision, embedding space, coreset selection, and input resolution under extremely low-resource remote sensing classification. (a) INT8 quantization closely matches FP32 accuracy, while BNNs exhibit severe degradation. (b) DINO embeddings outperform native features across models. (c) EntropyBDWC yields consistent gains at small data fractions. (d) Higher input resolution improves performance but does not offset poor data selection.

EntropyBDWC and Embedding Geometry. Across all architectures, EntropyBDWC achieves the most consistent performance, particularly under INT8 quantization. The gains are highlighted when coreset selection is performed in frozen DINO embedding space, where accuracy improvements persist even at extremely low data budgets. This suggests that self-supervised foundation model embeddings encode a geometry that is both semantically meaningful and resilient to quantization-induced distortion.

Figure 1 summarizes these interactions across architectures and precision regimes. DINO embeddings consistently outperform native features across models, EntropyBDWC dominates alternative coreset strategies at low data fractions, and INT8 quantization remains robust provided the selected data preserves embedding geometry. Resolution scaling improves accuracy (Fig. 1c) but cannot compensate for poor data selection or unstable embedding spaces, further reinforcing that data curation, not resolution alone, is the dominant factor under extreme constraints.

Discussion and Limitations. These results reveal that data selection and quantization should be treated as a coupled design problem rather than independent optimizations. While INT8 quantization is surprisingly robust, binary networks remain highly sensitive to both data scarcity and coreset quality. EntropyBDWC mitigates this sensitivity by explicitly preserving local feature geometry, particularly when applied in a foundation-model embedding space. Our study is limited to post-training quantization, image-level classification, and a fixed set of datasets; extending these findings to quantization-aware training, theoretical guarantees for geometry-preserving coresets, and temporal or multi-label remote sensing remains an important direction for future work.

5 Conclusion

Our work revisits remote sensing classification from a fundamentally constrained regime, few labeled samples, frozen backbones, and aggressive post-training quantization. Through extensive experiments across convolutional and transformer-based architectures, we show that standard coreset heuristics, while effective in full-precision settings, degrade sharply once representation geometry is distorted by low-bit quantization, with binary networks exhibiting near-complete accuracy collapse.

To address this challenge, we introduced *EntropyBDWC*, a quantization-robust coreset selection strategy that jointly models predictive uncertainty and local embedding density. Across datasets, architectures, and data budgets, EntropyBDWC consistently selects samples that remain linearly separable after quantization, yielding substantial gains in INT8 regimes and stabilizing performance under extreme data scarcity. Moreover, performing coreset selection in frozen foundation-model embedding spaces (DINO) further amplifies this robustness, revealing a new role for foundation models as *data selection* rather than trainable backbones in low-resource settings.

Our work opens a new direction for quantization-aware data curation, with implications extending beyond remote sensing to edge AI, foundation model adaptation, and efficient learning under systemic resource constraints.

Table 2: Accuracy (%) under joint data scarcity and quantization on the Brazilian Coffee Scenes dataset (96×96). Each cell reports **FP32 / INT8 / BNN**. Results compare coreset selection strategies using either native model embeddings (BASE) or frozen DINOv1 embeddings (DINO). **Bold** values indicate the best result within each model–coreset–budget block. Across architectures and data budgets, DINO-based selection consistently improves performance in low-data regimes and substantially enhances INT8 robustness. BNN results reflect naive post-hoc binarization without quantization-aware training and therefore saturate near chance level.

Model	Coreset Method	5%	10%	20%	100%
ResNet-18	Random (BASE)	78.3 / 79.2 / 51.7	82.5 / 79.2 / 51.7	84.2 / 84.2 / 51.7	92.5 / 85.0 / 51.7
	Entropy (BASE)	83.3 / 88.3 / 51.7	48.3 / 85.8 / 51.7	80.0 / 67.5 / 51.7	88.3 / 90.8 / 51.7
	EntropyBDWC (BASE)	48.3 / 87.5 / 51.7	77.5 / 88.3 / 51.7	**93.3** / 87.5 / 51.7	90.0 / 88.3 / 51.7
	Random (DINO)	77.5 / 80.0 / 51.7	84.2 / 75.8 / 51.7	88.3 / 89.2 / 51.7	91.7 / 84.2 / 51.7
	Entropy (DINO)	79.2 / 90.8 / 51.7	80.8 / 89.2 / 51.7	90.0 / 90.0 / 51.7	89.2 / 90.8 / 51.7
	EntropyBDWC (DINO)	80.0 / 90.0 / 51.7	75.0 / 90.8 / 51.7	90.0 / 82.5 / 51.7	89.2 / **95.0** / 51.7
ConvNeXt-Tiny	Random (BASE)	85.0 / 84.2 / 51.7	89.2 / 83.3 / 51.7	92.5 / 81.7 / 51.7	93.3 / 90.0 / 51.7
	Entropy (BASE)	90.8 / 91.7 / 51.7	92.5 / 92.5 / 51.7	80.8 / 91.7 / 51.7	94.2 / 90.0 / 51.7
	EntropyBDWC (BASE)	77.5 / 90.8 / 51.7	58.3 / 90.0 / 51.7	71.7 / 90.8 / 51.7	92.5 / 86.7 / 51.7
	Random (DINO)	79.2 / 81.7 / 51.7	85.8 / 82.5 / 51.7	90.8 / 83.3 / 51.7	95.0 / 89.2 / 51.7
	Entropy (DINO)	79.2 / 89.2 / 51.7	90.0 / 89.2 / 51.7	90.8 / 90.0 / 51.7	95.0 / 90.8 / 51.7
	EntropyBDWC (DINO)	91.7 / 89.2 / 51.7	94.2 / 90.0 / 51.7	**95.0** / 90.0 / 51.7	95.0 / 89.2 / 51.7
Swin-Tiny	Random (BASE)	83.3 / 80.8 / 51.7	80.0 / 85.8 / 51.7	87.5 / 88.3 / 51.7	95.0 / 90.8 / 51.7
	Entropy (BASE)	90.8 / 91.7 / 51.7	84.2 / 91.7 / 51.7	69.2 / 94.2 / 51.7	90.0 / 88.3 / 51.7
	EntropyBDWC (BASE)	85.8 / 89.2 / 51.7	86.7 / 88.3 / 51.7	52.5 / 90.8 / 51.7	90.8 / 91.7 / 51.7
	Random (DINO)	78.3 / 85.0 / 51.7	86.7 / 82.5 / 51.7	91.7 / 90.8 / 51.7	92.5 / 94.2 / 51.7
	Entropy (DINO)	79.2 / 94.2 / 51.7	83.3 / 92.5 / 51.7	92.5 / 90.8 / 51.7	92.5 / 94.2 / 51.7
	EntropyBDWC (DINO)	92.5 / 94.2 / 51.7	93.3 / 92.5 / 51.7	93.3 / 91.7 / 51.7	91.7 / 94.2 / 51.7
MobileViT-v2-S	Random (BASE)	55.8 / 50.0 / 51.7	70.0 / 59.2 / 51.7	83.3 / 63.3 / 51.7	91.7 / 66.7 / 51.7
	Entropy (BASE)	87.5 / 63.3 / 51.7	50.0 / 61.7 / 51.7	75.0 / 65.0 / 51.7	92.5 / 69.2 / 51.7
	EntropyBDWC (BASE)	51.7 / 71.7 / 51.7	80.8 / 74.2 / 51.7	85.8 / 73.3 / 51.7	90.8 / 70.8 / 51.7
	Random (DINO)	54.2 / 53.3 / 48.3	75.0 / 56.7 / 48.3	82.5 / 65.0 / 48.3	90.8 / 75.0 / 48.3
	Entropy (DINO)	62.5 / 77.5 / 48.3	90.0 / 73.3 / 48.3	84.2 / 70.8 / 48.3	92.5 / 70.0 / 48.3
	EntropyBDWC (DINO)	90.8 / 68.3 / 48.3	91.7 / 68.3 / 48.3	90.8 / 67.5 / 48.3	90.8 / 68.3 / 48.3

REFERENCES

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

Matthieu Courbariaux, Itay Hubara, Daniel Soudry, Ran El-Yaniv, and Yoshua Bengio. Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1. *arXiv preprint arXiv:1602.02830*, 2016.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059. PMLR, 2016.

Amir Gholami, Sehoon Kim, Zhen Dong, Zhewei Yao, Michael W Mahoney, and Kurt Keutzer. A survey of quantization methods for efficient neural network inference. In *Low-power computer vision*, pp. 291–326. Chapman and Hall/CRC, 2022.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019.

Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2704–2713, 2018.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

Sachin Mehta and Mohammad Rastegari. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*, 2021.

Baharan Mirzasoleiman, Jeff Bilmes, and Jure Leskovec. Coresets for data-efficient training of machine learning models. In *International Conference on Machine Learning*, pp. 6950–6960. PMLR, 2020.

Otávio AB Penatti, Keiller Nogueira, and Jefersson A Dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 44–51, 2015.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017.

Tushar Shinde. High-performance lightweight vision models for land cover classification with coresets and compression. In *TerraBytes-ICML 2025 workshop*.

Tushar Shinde and Manasa Madabhushi. Data-efficient and robust coreset selection via sparse adversarial perturbations. In *NeurIPS 2025 Workshop: Reliable ML from Unreliable Data*.