ADDRESSING SAR SHIP CLASS IMBALANCE VIA TARGETED OVERSAMPLING OF FOUNDATION MODEL EMBEDDINGS

Ch Muhammad Awais, Marco Reggiannini, Davide Moroni
Institute of Information Science and Technologies- ISTI
National Research Council of Italy (CNR, Pisa, Italy
{firstname.lastname}@isti.cnr.it

Oktay Karakus
Department of Computer Science,
Cardiff University, Cardiff, UK
o.karakus@cardiff.ac.uk

ABSTRACT

Foundation models pretrained on remote sensing data have shown promise for downstream tasks, yet their behaviour under class imbalance remains underexplored. We benchmark two foundation models, DOFA and SAR-JEPA, against ImageNet-pretrained models on the severely imbalanced OpenSARShip dataset. We apply four feature-space oversampling techniques exclusively to minority classes, scaling them to three times their original size. Our approach achieves up to 8.34% Macro-F1 and 7.34% accuracy improvements over baseline foundation models, demonstrating that targeted oversampling enables better balanced performance on SAR ship classification. We provide code to pre-extract embeddings, and reproducible experiments optimised for free-tier Google Colab https://github.com/cm-awais/SARShipfoundationModels.

1 INTRODUCTION AND METHODOLOGY

Synthetic Aperture Radar (SAR) imaging enables all-weather maritime surveillance. Recent foundation models pretrained on remote sensing data offer potential advantages for SAR tasks, but two practical questions motivate this work: (1) Do foundation models handle class imbalance better than ImageNet-pretrained models? (2) Can targeted oversampling improve their performance without full model fine-tuning? We test these questions using DOFA (Xiong et al., 2024) and SAR-JEPA (Li et al., 2024) foundation models on the 6 ship classes of the OpenSARShip (Li et al., 2017) dataset: Cargo (5303 samples), Tanker (1825), Dredging (142), Fishing (139), Passenger (66), and Tug (62).

We selected four advanced SMOTE variants (Kovács, 2019) that generate synthetic samples by interpolating between k-nearest neighbors in embedding space. SVM-SMOTE and KMeans-SMOTE use support vectors and cluster centroids to identify safe regions for synthesis, while SMOTE-ENN applies Edited Nearest Neighbours to remove noisy borderline samples. ADASYN (Adaptive Synthetic Sampling) generates more samples near decision boundaries. Unlike naive oversampling, these methods first fit on the complete training set to learn global structure, then generate synthetic samples exclusively for minority classes. This ensures synthetic samples respect decision boundaries learned from majority classes.

Algorithm 1 Minority Class Oversampling for Foundation Models

1: **Input:** Pretrained model f, training set $\mathcal{D}_{\text{train}}$, threshold $\tau = 200$
2: Extract embeddings: $\mathbf{X}_{\text{train}} = \{f(\mathbf{x}_i)\}_{i=1}^{N}$, Identify minority classes: $\mathcal{M} = \{c : N_c < \tau\}$
3: **for** each class $c \in \mathcal{M}$ **do**
4: Apply oversampling method to obtain $N'_c = 3N_c$ samples
5: **end for**
6: Train classifier $g : \mathbb{R}^d \to \mathbb{R}^C$ on augmented embeddings
7: **Return:** Classifier g

ImageNet-pretrained ResNet, VGG (Mascarenhas & Agarwal, 2021), and ViT-224 (Yuan et al., 2021) served as baselines. Algorithm 1 shows the experiment pipeline. The utilised dataset was split 80/10/10 for training, validation, and test sets, respectively. Final layer embeddings were extracted and stored for all splits.

The classifier g is a three-layer fully connected network with batch normalization, ReLU activations, dropout, and trained with weighted cross-entropy loss (learning rate 0.0005, batch size 64, 100 epochs). Four minority classes (Dredging, Fishing, Passenger, Tug) were oversampled using the oversampling methods. Each was scaled to three times its original size in embedding space. Validation and test used only the original embeddings.

2 Results and Discussion

Table 1 shows performance across models and methods. ImageNet-pretrained ViT achieves 73.74% accuracy but only 27.03 Macro-F1, revealing poor minority class performance. Foundation models show better baseline Macro-F1 (DOFA: 32.12, SAR-JEPA: 26.96) than ImageNet models (ViT: 27.03, ResNet: 13.75) despite lower accuracy, indicating less bias toward majority classes. Oversampling methods that learn global structure (SVM-SMOTE, KMeans-SMOTE) outperform those with cleaning steps (SMOTE-ENN). DOFA SMOTE-ENN drops to 28.83 Macro-F1, below baseline, suggesting borderline sample removal harms minority classes in embedding space.

Table 1: Performance on OpenSARShip (Acc=Accuracy %). Best Macro-F1 in bold.

ImageNet + Foundation Baselines				**Foundation Models + Oversampling**			
Model	**Method**	**Acc**	**Macro-F1**	**Model**	**Method**	**Acc**	**Macro-F1**
ResNet	baseline	70.24	13.75	DOFA	SVM-SMOTE	66.47	39.24
VGG	baseline	71.30	19.93	DOFA	KMeans-SMOTE	66.07	**40.46**
ViT-224	baseline	73.74	27.03	DOFA	SMOTE-ENN	63.89	28.83
DOFA	baseline	59.92	32.12	DOFA	ADASYN	64.95	36.88
SAR-JEPA	baseline	52.18	26.96	SAR-JEPA	SVM-SMOTE	57.28	29.76
				SAR-JEPA	KMeans-SMOTE	59.13	30.62
				SAR-JEPA	SMOTE-ENN	59.52	27.36
				SAR-JEPA	ADASYN	56.15	29.44

DOFA benefits substantially from oversampling (+8.34%), surpassing all ImageNet baselines including ViT's 27.03 Macro-F1, while SAR-JEPA shows consistent but modest gains (+3.66%). All oversampling strategies improve accuracy over their respective foundation model baselines, with DOFA, and SAR-JEPA gaining upto 6.55% and 7.34%, respectively. KMeans-SMOTE and SVM-SMOTE consistently outperform SMOTE-ENN across both models, indicating that preserving minority samples matters more than aggressive noise removal in pretrained embedding spaces. Notably, DOFA+KMeans-SMOTE achieves 40.46% Macro-F1 versus ViT's 27.03% despite 7.67% lower accuracy, demonstrating that accuracy alone severely underestimates performance on imbalanced data. By extracting and storing embeddings once, our approach enables rapid experimentation: all 13 model-method combinations (5 baselines + 8 oversampling variants) complete training on free-tier Colab GPU/CPU, enabling rapid experimentation without costly foundation model fine-tuning.

3 Conclusion

We demonstrated that SAR foundation models suffer from class imbalance but respond well to targeted minority class oversampling, with DOFA achieving a +8.34% Macro-F1 improvement. DOFA+KMeans-SMOTE (40.46 Macro-F1) surpasses the best-performing ImageNet-pretrained ViT (27.03) by 13.43% despite 7.67% lower accuracy, showing that Macro-F1 is essential for imbalanced evaluation and that SAR-specific pretraining combined with embedding-space oversampling enables superior minority class recognition. Our public release of code to extract embeddings and free-tier Colab code facilitates reproducibility and accessibility for the remote sensing community.

Future work can extend this study in several directions by exploring classifier architecture search, class-specific oversampling ratios, a broader set of ImageNet-pretrained baselines, and designing oversampling methods specifically for SAR foundation model embeddings, since current techniques were developed for general feature spaces and may not capture SAR-specific characteristics; additionally, investigating other foundation models and combining oversampling with other imbalance mitigation strategies could provide complementary benefits.

REFERENCES

György Kovács. Smote-variants: A python implementation of 85 minority oversampling techniques. *Neurocomputing*, 366:352–354, 2019.

Boying Li, Bin Liu, Lanqing Huang, Weiwei Guo, Zenghui Zhang, and Wenxian Yu. Opensarship 2.0: A large-volume dataset for deeper interpretation of ship targets in sentinel-1 imagery. In *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*, pp. 1–5. IEEE, 2017.

Weijie Li, Wei Yang, Tianpeng Liu, Yuenan Hou, Yuxuan Li, Zhen Liu, Yongxiang Liu, and Li Liu. Predicting gradient is better: Exploring self-supervised learning for sar atr with a joint-embedding predictive architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218:326–338, 2024.

Sheldon Mascarenhas and Mukul Agarwal. A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification. In *2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)*, volume 1, pp. 96–99. IEEE, 2021.

Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired foundation model for observing the earth crossing modalities. *CoRR*, 2024.

Li Yuan, Yunpeng Chen, Tao Wang, Weihao Yu, Yujun Shi, Zi-Hang Jiang, Francis EH Tay, Jiashi Feng, and Shuicheng Yan. Tokens-to-token vit: Training vision transformers from scratch on imagenet. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 558–567, 2021.