Enhancing Foundation Models for Imbalanced SAR Ship Classification via Targeted Oversampling

Ch Muhammad Awais[1][0009−0001−4589−4103], Marco Reggiannini[1][0000−0002−4872−9541], and Davide Moroni[1][0000−0002−5175−5126]

Institute of Science and Technology, National Research Council, Pisa, Italy
fname.lname@cnr.it

Abstract. Remote-sensing foundation models offer strong representations for SAR imagery, but their behavior under severe long-tail class imbalance is still not well characterized. We benchmark DOFA and SAR-JEPA on the imbalanced OpenSARShip dataset and compare them with ImageNet-pretrained baselines under a fixed, training-efficient protocol that keeps the backbone frozen. To mitigate imbalance without fine-tuning, we apply four oversampling methods in embedding space exclusively to minority classes and train a lightweight classifier head on the augmented embeddings. Across both foundation models, oversampling improves Macro-F1 and test accuracy relative to their respective baselines, with the largest Macro-F1 gains observed for DOFA using ADASYN (34.39 to 38.56) and for SAR-JEPA using SVM-SMOTE (25.89 to 32.30). We also report class-wise behavior, showing that aggregate improvements can coexist with persistent failures on specific rare classes. Code for embedding extraction and reproducible multi-seed evaluation is provided to support rapid experimentation on free-tier hardware.

Keywords: SAR Ship Classification · Oversampling · Foundation Models

1 Introduction

Synthetic Aperture Radar (SAR) enables all-weather maritime monitoring, making it suitable for ship classification in operational settings. However, practical SAR ship datasets are typically long-tailed, with a few dominating vessel categories, while rare types remain underrepresented [4]. Under this imbalance, high test accuracy can coexist with poor recognition of minority classes [5]. SAR target recognition has been studied with both traditional machine learning approaches [7, 11, 19–21] and modern deep learning methods [1, 9, 16, 23, 27, 31, 33], yet severe imbalance remains a persistent practical obstacle.

Deep learning models generally benefit from large and diverse labeled datasets, but collecting SAR ship imagery at scale is difficult and costly. As a result, prior

work often relies on strategies such as transfer learning [16, 22], data augmentation [13, 24], multimodal fusion [28, 29, 34], and domain adaptation [2, 8]. More recently, foundation models pretrained on remote sensing data have emerged as a promising alternative to ImageNet-pretrained backbones for Earth observation tasks [17, 26]. Despite this promise, there is limited evidence on how such representations behave on SAR ship classification when class distributions are highly skewed, and whether their benefits persist when evaluation emphasizes minority performance.

A natural way to address imbalance without collecting new labels is to rebalance the training distribution. Oversampling methods such as SMOTE [3, 6] generate synthetic minority samples by interpolating between nearest neighbors in feature space, avoiding simple duplication. Several variants refine the synthesis process. SVM-SMOTE targets regions near decision boundaries, KMeans-SMOTE synthesizes within clusters, SMOTE-ENN removes ambiguous samples after oversampling, and ADASYN allocates more synthesis to locally difficult regions [12]. While these methods are well studied for conventional feature spaces, their behavior on embeddings produced by large pretrained models has received comparatively little attention, particularly for SAR where minority classes can be one to two orders of magnitude smaller than the majority ones.

This work addresses two questions. First, do remote-sensing foundation model embeddings provide better balanced performance than ImageNet-pretrained baselines under severe class imbalance? Second, can minority-only oversampling in embedding space artificially solve the balance issue in the original dataset and generate a reliable classifier, with high performance across categories without end-to-end backbone fine-tuning? To answer these questions, we benchmark Dynamic One-For-All (DOFA) [30] and SAR-JEPA [15] on OpenSARShip [14] under a controlled protocol. Embeddings are extracted once from a frozen backbone, oversampling is applied only to minority classes in feature space, and a lightweight classifier head is trained on the augmented embeddings. This design supports rapid experimentation and reproducible evaluation on free-tier hardware. In summary, our contributions are:

1. A benchmark of remote-sensing foundation model embeddings on an imbalanced SAR ship classification task, reporting Macro-F1 alongside accuracy.
2. A training-efficient imbalance mitigation pipeline that performs minority-only oversampling in embedding space and improves balanced performance without backbone fine-tuning.
3. An open-source, Colab-friendly implementation that pre-extracts embeddings and supports reproducible multi-seed evaluation. https://github.com/cmawais/SARShipfoundationModels

2 Methodology

Fig. 1 summarises the pipeline. The backbone is kept frozen throughout: embeddings are extracted once, minority-only oversampling is applied in feature space, and a lightweight classifier head is trained on the augmented embeddings. This

design avoids costly backbone fine-tuning and keeps experimentation tractable on free-tier hardware.

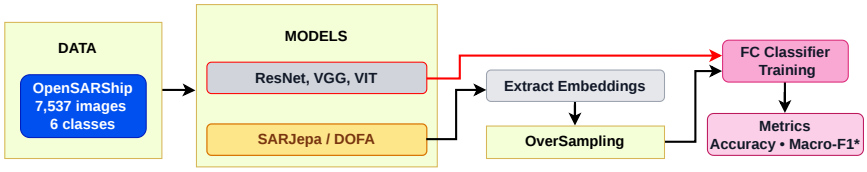

Fig. 1. Overview of the proposed evaluation pipeline. Given OpenSARShip, we extract frozen embeddings from each backbone (ImageNet baselines and Earth Observation foundation models), apply minority-only oversampling in embedding space, and train a lightweight fully connected classifier head on the augmented embeddings.

2.1 Dataset and Foundation Models

The proposed method is evaluated on OpenSARShip [14], a SAR ship classification dataset with six classes: Cargo (5303 samples), Tanker (1825), Dredging (142), Fishing (139), Passenger (66), and Tug (62). The dataset is split into train, validation, and test sets using an 80/10/10 ratio. Given a pretrained encoder $f(\cdot)$, each SAR chip is mapped to a d-dimensional embedding, which is extracted for all splits and saved to disk before any oversampling or classifier training takes place.

We benchmark two remote-sensing foundation models: Dynamic One-For-All (DOFA) [30], a unified multimodal framework designed for diverse earth observation tasks, and SAR-JEPA [15], a joint-embedding predictive architecture trained directly on SAR data. Both are compared against three ImageNet-pretrained baselines, ResNet, VGG [18], and ViT-224 [32], evaluated under the same protocol.

2.2 Oversampling in Embedding Space

Classes with less than 200 training samples are treated as minority classes. The subset related to the minority classes is thus given by $\mathcal{M}$ = Dredging, Fishing, Passenger, Tug. For each minority class, oversampling is applied exclusively within that class in feature space, until the class size reaches 3 times its original value. Majority classes are never modified, and validation and test embeddings are left untouched.

All four oversampling methods generate synthetic points by interpolating between same-class nearest neighbours in $\mathbb{R}^d$:

$$\tilde{\mathbf{z}} = \mathbf{z}_i + \lambda(\mathbf{z}_j - \mathbf{z}_i), \qquad \lambda \sim \mathcal{U}(0,1). \tag{1}$$

The methods differ in how the pairs $(\mathbf{z}_i, \mathbf{z}_j)$ are selected: SVM-SMOTE restricts synthesis to regions near the support-vector decision boundary; KMeans-SMOTE first clusters the minority classes, then synthesises within each cluster, keeping synthetic samples close to the true data manifold; SMOTE-ENN starts with a KMeans-SMOTE approach, then it additionally removes ambiguous samples using Edited Nearest Neighbours; and ADASYN weights synthesis density by local class difficulty i.e., generating more samples in regions where minority examples overlap with the majority ones [12].

2.3 Classifier and Training Protocol

The classifier head $g(\cdot)$ is a small three-layer perceptron with batch normalisation, ReLU activations, and dropout ($p = 0.1$). It is trained on the augmented training set using cross-entropy loss with Adam (learning rate 10^{-3}, batch size 256, 50 epochs). Results are reported as averages over three random seeds, varying model initialisation and sampler randomness while keeping the data split fixed.

Table 1. Overall performance metrics.

Model	Method	Accuracy	Macro-F1
DOFA	Base	60.65±1.12	34.39±0.26
DOFA	ADASYN	72.42±0.84	**38.56±1.85**
DOFA	KMeans	**72.82±0.75**	35.59±0.62
DOFA	SENN	66.96±1.17	33.85±1.23
DOFA	SVM	72.32±1.36	36.87±1.05
SAR-JEPA	Base	48.02±2.53	25.89±1.20
SAR-JEPA	ADASYN	69.74±0.98	29.28±0.54
SAR-JEPA	KMeans	**71.76±1.22**	30.86±2.64
SAR-JEPA	SENN	66.37±0.14	28.11±1.25
SAR-JEPA	SVM	71.20±0.42	**32.30±0.26**
ResNet-50	Base	70.24±0.25	13.75±0.42
VGG-16	Base	71.30±0.33	19.93±0.11
ViT-224	Base	73.74±0.39	**27.03±0.01**

Base: baseline; KMeans: KMeans-SMOTE; SENN: SMOTEENN; SVM: SVMSMOTE

3 Results

We evaluated five classification pipelines across two foundation models and three ImageNet baselines. Table 1 summarizes overall test accuracy and Macro-F1 scores. Among baseline architectures, ViT-224 achieved the highest accuracy at 73.74% with a Macro-F1 of 27.03, substantially outperforming ResNet and VGG.

The DOFA baseline achieved lower accuracy than the ImageNet baselines in Table 1, but oversampling improved it substantially. KMeans-SMOTE produced the best accuracy at 72.82%, while ADASYN achieved the highest Macro-F1 score of 38.56. Both ADASYN and SVMSMOTE converged to similar accuracy levels around 72.3%. SMOTEENN showed degraded performance relative to other oversampling variants, reaching only 66.96% accuracy.

SAR-JEPA baseline performance was markedly weaker, achieving 48.02% accuracy and 25.89 Macro-F1. Oversampling substantially improved results, with KMeans-SMOTE reaching 71.76% accuracy and SVMSMOTE attaining 32.30 Macro-F1. The large gap between SAR-JEPA baseline and oversampled variants suggests SAR-JEPA is more sensitive to the choice of the oversampling method under our fixed classifier setup. ADASYN and SVMSMOTE both reached approximately 71% accuracy, while SMOTEENN again underperformed.

Table 2. Per-class F1-scores across configurations.

Class	Baseline	ADASYN	KMeans	SENN	SVM
DOFA					
Cargo	73.15±1.27	82.52±0.67	83.09±0.64	81.32±0.34	**83.54±0.19**
Dredging	18.74±3.64	**19.93±5.09**	8.28±4.46	19.71±1.14	8.99±0.14
Fishing	27.47±2.05	24.70±5.47	30.56±2.82	32.03±6.72	**34.31±4.52**
Passenger	24.98±7.47	**38.41±7.17**	28.83±3.88	26.08±2.40	30.79±1.12
Tanker	**52.40±1.39**	51.30±0.75	49.91±1.64	30.75±2.51	49.22±0.63
Tug	19.00±3.24	18.53±3.57	18.93±10.1	20.66±7.04	**24.50±0.71**
SAR-JEPA					
Cargo	63.57±2.37	81.94±0.51	**82.83±0.56**	81.34±0.26	82.46±0.16
Dredging	11.24±2.18	8.25±2.24	5.08±0.90	**18.29±2.99**	13.50±3.09
Fishing	22.90±2.84	30.45±0.94	36.02±8.25	27.44±3.45	**39.92±1.27**
Passenger	7.04±1.41	6.90±4.88	6.70±1.83	**10.82±1.84**	0.00±0.00
Tanker	**44.57±1.43**	34.50±0.92	38.84±0.43	25.65±2.26	40.08±2.10
Tug	9.61±2.55	17.39±0.00	19.05±0.00	11.21±3.00	**21.64±0.83**

Per-class results reveal the underlying challenges of the dataset. Table 2 presents F1-scores across vessel types and methods, highlighting differences between backbones. The Cargo class demonstrates consistently strong performance across methods. For DOFA, Cargo F1 is above 81% for all methods, while for SAR-JEPA it ranges from 63.57% (Baseline) to 82.83% (KMeans-SMOTE). This suggests that majority classes benefit uniformly from the oversampling strategies employed.

DOFA exhibits more balanced per-class performance overall. Passenger shows the largest spread across oversampling methods within each backbone. For DOFA, ADASYN yields the strongest Passenger F1 (38.41%), whereas for SAR-JEPA some oversampling choices degrade Passenger severely, including a null F1 in case SVMSMOTE is adopted, despite a moderate overall accuracy. This highlights why overall accuracy can be misleading under severe imbalance and why per-class reporting is necessary even when Macro-F1 is included.

Minority classes follow distinct patterns by method. The Dredging class shows marked volatility in DOFA, dropping to 8.28% with KMeans-SMOTE but reaching 19.93% with ADASYN. SAR-JEPA performs even worse on Dredging, with all methods producing F1-scores below 18.29%. The Fishing class exhibits more consistent improvements, with SAR-JEPA SVMSMOTE achieving the highest performance at 39.92%. Tanker class performance degrades with SMOTEENN for both architectures, suggesting this method may not suit denser minority manifolds.

SAR-JEPA baseline underperforms substantially across most classes, particularly on Passenger (7.04%) and Tug (9.61%), indicating that this backbone is more sensitive to class imbalance under our fixed classifier setup. DOFA baseline provides more competitive baseline scores, with Tanker at 52.40% and Cargo at 73.15%. The oversampling techniques improve SAR-JEPA performance across the board, but introduce high variance in specific class-method pairs. ADASYN and KMeans-SMOTE provide the most consistent improvements, though their effectiveness varies by class and architecture.

4 Discussion

This study benchmarks remote-sensing foundation model embeddings for imbalanced SAR ship classification and shows that applying proper minority-only oversampling methods in embedding space, can improve Macro-F1 in this benchmark, without backbone fine-tuning. Under a fixed classifier head and training protocol, oversampling consistently increases Macro-F1 for both DOFA and SAR-JEPA, while keeping the overall pipeline lightweight and reproducible.

A first outcome is that accuracy alone is not representative of performance under the OpenSARShip long-tail distribution. The ImageNet baselines obtain the highest accuracies, with ViT-224 reaching 73.74%, yet their Macro-F1 scores remain low (Table 1). This indicates that correct predictions are concentrated on the frequent classes. In contrast, DOFA exhibits a higher Macro-F1 at baseline despite lower accuracy, suggesting that its embedding space preserves more separability for minority classes even before any oversampling is applied.

A second outcome is that the choice of oversampling method affects accuracy and Macro-F1 differently, and the best configuration depends on the target metric. For DOFA, KMeans-SMOTE yields the highest accuracy, whereas ADASYN achieves the highest Macro-F1 (Table 1). For SAR-JEPA, KMeans-SMOTE again maximizes accuracy, while SVM-SMOTE attains the best Macro-F1. This separation reinforces the need to report balanced metrics when comparing models for long-tail SAR classification, since improvements in accuracy can be decoupled from improvements on rare classes.

The class-wise view clarifies where these gains originate. Table 2 shows that Cargo remains consistently strong across methods, which is expected given its large support. The more informative differences appear among the minority classes. Passenger is a particularly challenging class: DOFA combined with ADASYN yields the strongest Passenger F1, while SAR-JEPA remains weak and can fail

entirely for specific oversampling choices. Dredging also remains challenging for both backbones, and its response to oversampling is inconsistent. Fishing follows a different profile, with larger improvements in some SAR-JEPA settings. These patterns indicate that minority classes are not interchangeable and that a single oversampling strategy may not be uniformly beneficial.

Since the backbone is frozen, all improvements are attributable to changes in the effective training distribution seen by the classifier head. Interpolation-based synthesis is likely to help when same-class neighborhoods in embedding space are locally coherent, because synthetic points expand the decision region with samples that remain consistent with the class manifold. Conversely, when a minority class is fragmented, overlapping, or poorly represented, synthetic points can amplify noise and degrade class discrimination. The per-class failures visible in Table 2 are consistent with this limitation and motivate reporting class-wise results alongside aggregate metrics.

From a practical standpoint, this pipeline uses frozen pretrained embeddings as a lightweight testbed for imbalance mitigation, letting us leverage strong representations while avoiding the cost and complexity of end-to-end fine-tuning. It also enables rapid comparison of oversampling techniques under controlled conditions. In this benchmark, ADASYN is the strongest choice for maximizing Macro-F1 with DOFA, while SVM-SMOTE is the strongest choice for SAR-JEPA. If accuracy is prioritized, KMeans-SMOTE is the best configuration for both foundation models (Table 1). These selections should still be validated against per-class behavior when the operational requirements emphasize specific rare categories.

There are several limitations to consider. Results are averaged over three random seeds with a fixed dataset split, so the reported variability reflects initialization and sampler randomness but does not capture split sensitivity. Oversampling is applied with a single minority threshold and a fixed target multiplier, even though different classes may require different augmentation strengths. Finally, feature-space interpolation assumes that linear combinations of neighbors remain plausible within the representation space, which may not hold equally for all classes and backbones.

Future work can extend this benchmark by making the oversampling policy class-adaptive, selecting both method and oversampling ratio using validation performance per class. Additional studies can combine embedding-space oversampling with loss reweighting or focal objectives while keeping the backbone frozen. A further direction is to introduce parameter-efficient fine-tuning, such as adapters or LoRA [10,25], to test whether the observed minority gains persist when representation learning is allowed to adjust.

5 Conclusion

This work provides a benchmark of remote-sensing foundation model embeddings for imbalanced SAR ship classification and evaluates minority-only oversampling in embedding space as a simple alternative to end-to-end fine-tuning.

Under a fixed classifier head and training setup, oversampling improves balanced performance for both DOFA and SAR-JEPA. DOFA achieves its best Macro-F1 with ADASYN (38.56), while SAR-JEPA achieves its best Macro-F1 with SVM-SMOTE (32.30). In terms of accuracy, KMeans-SMOTE yields the strongest results for both foundation models (72.82 for DOFA and 71.76 for SAR-JEPA), highlighting that the configuration that maximizes accuracy is not necessarily the one that maximizes Macro-F1.

Class-wise results further show that improvements are not uniform across categories. Cargo remains consistently strong, while rare classes such as Passenger and Dredging remain challenging and can be sensitive to the oversampling choice, including occasional severe degradations for specific combinations. These findings support reporting both aggregate and per-class metrics when evaluating imbalanced SAR recognition systems. Future work will address the development of ad-hoc techniques, including class-adaptive oversampling and parameter-efficient adaptation.

References

1. Awais, C.M., Reggiannini, M., Moroni, D.: A framework for imbalanced SAR ship classification: Curriculum learning weighted loss functions and a novel evaluation metric. In: Proceedings of the Winter Conference on Applications of Computer Vision (WACV) Workshops. pp. 1570–1578 (2025)
2. Awais, C.M., Reggiannini, M., Moroni, D., Galdelli, A.: Sar-to-infrared domain adaptation for maritime surveillance with limited data. In: Proceedings. vol. 129, p. 66. MDPI (2025)
3. Awais, C.M., Reggiannini, M., Moroni, D., Karakuş, O.: Feature-space oversampling for addressing class imbalance in sar ship classification. In: IGARSS 2025-2025 IEEE International Geoscience and Remote Sensing Symposium. pp. 2010–2014. IEEE (2025)
4. Awais, C.M., Reggiannini, M., Moroni, D., Salerno, E.: A survey on sar ship classification using deep learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing pp. 1–26 (2026). https://doi.org/10.1109/JSTARS.2026.3695704
5. Awais, C., Reggiannini, M.: Deep learning for sar ship classification: Focus on unbalanced datasets and inter-dataset generalization. In: 2024 International Conference on Electromagnetics in Advanced Applications (ICEAA). pp. 1–1. IEEE (2024)
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research **16**, 321–357 (2002)
7. Fu, K., Dou, F.Z., Li, H.C., Diao, W.H., Sun, X., Xu, G.L.: Aircraft recognition in sar images based on scattering structure feature and template matching. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **11**(11), 4206–4217 (2018)
8. Gao, G., Dai, Y., Zhang, X., Duan, D., Guo, F.: ADCG: A cross-modality domain transfer learning method for synthetic aperture radar in ship automatic target recognition. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–14 (2023). https://doi.org/10.1109/TGRS.2023.3313204

9. He, J., Chang, W., Wang, F., Liu, Y., Wang, Y., Liu, H., Li, Y., Liu, L.: Group bilinear CNNs for dual-polarized SAR ship classification. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2022). https://doi.org/10.1109/LGRS.2022.3178080
10. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. Iclr **1**(2), 3 (2022)
11. Ikeuchi, K., Shakunaga, T., Wheeler, M.D., Yamazaki, T.: Invariant histograms and deformable template matching for sar target recognition. In: Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 100–105. IEEE (1996)
12. Kovács, G.: Smote-variants: A python implementation of 85 minority oversampling techniques. Neurocomputing **366**, 352–354 (2019)
13. Lang, H., Yang, G., Li, C., Xu, J.: Multisource heterogeneous transfer learning via feature augmentation for ship classification in SAR imagery. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–14 (2022). https://doi.org/10.1109/TGRS.2022.3178703
14. Li, B., Liu, B., Huang, L., Guo, W., Zhang, Z., Yu, W.: 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)
15. Li, W., Yang, W., Liu, T., Hou, Y., Li, Y., Liu, Z., Liu, Y., Liu, L.: 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)
16. Lu, C., Li, W.: Ship classification in high-resolution SAR images via transfer learning with small training dataset. Sensors **19**(1) (2019). https://doi.org/10.3390/s19010063, https://www.mdpi.com/1424-8220/19/1/63
17. Lu, S., Guo, J., Zimmer-Dauphinee, J.R., Nieusma, J.M., Wang, X., VanValkenburgh, P., Wernke, S.A., Huo, Y.: Vision foundation models in remote sensing: A survey. IEEE Geoscience and Remote Sensing Magazine **13**(3), 190–215 (2025). https://doi.org/10.1109/MGRS.2025.3541952
18. Mascarenhas, S., Agarwal, M.: 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). vol. 1, pp. 96–99. IEEE (2021)
19. Meth, R., Chellappa, R.: Feature matching and target recognition in synthetic aperture radar imagery. In: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258). vol. 6, pp. 3333–3336. IEEE (1999)
20. Nicoli, L.P., Anagnostopoulos, G.C.: Shape-based recognition of targets in synthetic aperture radar images using elliptical fourier descriptors. In: Automatic Target Recognition XVIII. vol. 6967, pp. 148–159. SPIE (2008)
21. Park, J.I., Park, S.H., Kim, K.T.: New discrimination features for sar automatic target recognition. IEEE Geoscience and Remote Sensing Letters **10**(3), 476–480 (2012)
22. Relekar, H., Shanmugam, P.: Transfer learning based ship classification in Sentinel-1 images incorporating scale variant features. Advances in Space Research **68**(11), 4594–4615 (2021). https://doi.org/https://doi.org/10.1016/j.asr.2021.08.042, https://www.sciencedirect.com/science/article/pii/S027311772100692X

23. Wang, C., Zhang, H., Wu, F., Zhang, B., Tian, S.: Ship classification with deep learning using COSMO-SkyMed SAR data. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). pp. 558–561 (2017). https://doi.org/10.1109/IGARSS.2017.8127014
24. Wang, L., Qi, Y., Mathiopoulos, P.T., Zhao, C., Mazhar, S.: An improved SAR ship classification method using text-to-image generation-based data augmentation and squeeze and excitation. Remote Sensing **16**(7) (2024). https://doi.org/10.3390/rs16071299, https://www.mdpi.com/2072-4292/16/7/1299
25. Wang, Z., Shen, Z., He, Y., Sun, G., Wang, H., Lyu, L., Li, A.: Flora: Federated fine-tuning large language models with heterogeneous low-rank adaptations. Advances in Neural Information Processing Systems **37**, 22513–22533 (2024)
26. Xiao, A., Xuan, W., Wang, J., Huang, J., Tao, D., Lu, S., Yokoya, N.: Foundation models for remote sensing and earth observation: A survey. IEEE Geoscience and Remote Sensing Magazine (2025)
27. Xie, N., Xiong, M., Wei, F., Li, J., Zhang, T., Yu, W.: MFSAF: A plug-and-play module for SAR ship classification. In: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium. pp. 9075–9079 (2024). https://doi.org/10.1109/IGARSS53475.2024.10642068
28. Xie, N., Zhang, T., Guo, W., Zhang, Z., Yu, W.: Dual branch deep network for ship classification of dual-polarized sar images. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–15 (2024). https://doi.org/10.1109/TGRS.2024.3372021
29. Xie, N., Zhang, T., Wei, F., Yu, W.: Pfdn: A polarimetric feature-guided deep network for dual-polarized SAR ship classification. IEEE Geoscience and Remote Sensing Letters **21**, 1–5 (2024). https://doi.org/10.1109/LGRS.2024.3367759
30. Xiong, Z., Wang, Y., Zhang, F., Stewart, A.J., Hanna, J., Borth, D., Papoutsis, I., Le Saux, B., Camps-Valls, G., Zhu, X.X.: Neural plasticity-inspired foundation model for observing the earth crossing modalities. CoRR (2024)
31. Xu, J., Lang, H.: A unified multiple proxy deep metric learning framework embedded with distribution optimization for fine-grained ship classification in remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **17**, 5604–5620 (2024). https://doi.org/10.1109/JSTARS.2024.3365472
32. Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z.H., Tay, F.E., Feng, J., Yan, S.: 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)
33. Zhang, T., Zhang, X.: Injection of traditional hand-crafted features into modern CNN-based models for SAR ship classification: What, why, where, and how. Remote Sensing **13**(11) (2021). https://doi.org/10.3390/rs13112091, https://www.mdpi.com/2072-4292/13/11/2091
34. Zheng, H., Hu, Z., Yang, L., Xu, A., Zheng, M., Zhang, C., Li, K.: Multifeature collaborative fusion network with deep supervision for SAR ship classification. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–14 (2023). https://doi.org/10.1109/TGRS.2023.3297648