# A Three-Stage Framework with Reliable Sample Pool for Long-Tailed Classification

Feng Cai*, Keyu Wu*, Haipeng Wang†, Feng Wang†
Fudan University
*{fcai21, kywu21}@m.fudan.edu.cn, †{hpwang, fengwang}@fudan.edu.cn

## Abstract

*Synthetic Aperture Radar (SAR) imagery presents a promising solution for acquiring Earth surface information regardless of weather and daylight. However, the SAR dataset is commonly characterized by a long-tailed distribution due to the scarcity of samples from infrequent categories. In this work, we extend the problem to aerial view object classification in the SAR dataset with long-tailed distribution and a plethora of negative samples. Specifically, we propose a three-stage approach that employs a ResNet101 backbone for feature extraction, Class-balanced Focal Loss for class-level re-weighting, and reliable pseudo-labels generated through semi-supervised learning to improve model performance. Moreover, we introduce a Reliable Sample Pool (RSP) to enhance the model's confidence in predicting in-distribution data and mitigate the domain gap between the labeled and unlabeled sets. The proposed framework achieved a Top-1 Accuracy of 63.20% and an AUROC of 0.71 on the final dataset, winning the first place in track 1 of the PBVS 2023 Multi-modal Aerial View Object Classification Challenge.*

## 1. Introduction

Synthetic Aperture Radar (SAR) imagery has garnered significant attention [1, 2, 3] for its unique ability to provide all-time, all-weather imaging of the Earth's surface, which compensates for traditional Electro-Optical (EO) imagery in some aspects. Notwithstanding the advantages, the classification of SAR images remains a formidable challenge owing to the complex and heterogeneous nature of the data [4]. In this paper, the focus lies on the task of aerial view object classification under long-tailed distribution.

This task poses several challenges. 1) The long-tail distribution of the data set. In addition to the inter-class long-tail distribution, there is also an intra-class long-tail distribution. The former refers to the uneven distribution of data samples among different classes, while the latter refers to the uneven distribution of data samples within the same class. The intra-class long tail distribution is also defined as Attribute-wise Long Tail [5], where the attributes can be those of the vehicle itself, as well as image-level attributes such as the background. 2) The presence of out-of-distribution (OOD) samples in the test set. These negative samples do not belong to any of the classes in the training set. The classifier needs to correctly classify the images within the distribution and identify patterns that do not exist in the training data. 3) The SAR images in the dataset are of poor quality and may contain a significant amount of noise. It is challenging to distinguish targets from clutter, particularly when targets exhibit low signal-to-noise ratios. These challenges pose significant obstacles to achieving high-performance classification models.

A novel three-stage approach is proposed in this paper to address the above challenges. Specifically, a ResNet101 [6] backbone pre-trained on ImageNet is employed for feature extraction. Rather than relying solely on label frequencies of training samples for loss re-weighting, Class-balanced Focal Loss [7] is utilized to re-balance classes by adjusting loss values more efficiently for different classes during training. In the second and third stages, the reliable pseudo-labels generation and semi-supervised learning are leveraged to improve model performance and mitigate overfitting on labeled data. Additionally, a Reliable Sample Pool (RSP) is introduced to enhance the model's confidence in predicting in-distribution data and alleviate the domain gap between the labeled and unlabeled sets. RSP stores the top-N samples with the highest confidence scores for each class prediction, which are continuously updated with the training iterations and used for fine-tuning to further improve the model's performance.

Our contributions can be summarized as follows:

1) We rethink the strategy for addressing long-tail distribution and propose a three-stage framework that can overcome both class imbalance and out-of-distribution sample interference issues.

*Equal contribution.
†Corresponding Author.

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61901122 and 62271153), and the Natural Science Foundation of Shanghai (Grant No. 20ZR1406300 and 22ZR1406700).

2) We propose a flow-state Reliable Sample Pool that stores high-confidence samples. By utilizing trustworthy predictions, the pool can increase the model's trust degree in data within the distribution and mitigates the domain gap between the labeled and unlabeled sets.
3) The simplicity and versatility of the proposed method allow for easy integration with different approaches to further enhance classification effectiveness.

Overall, this work provides an idea for training strategies in dealing with long-tailed distributions. The structure of this paper is organized as follows: Section 2 provides a comprehensive review of related work; Section 3 outlines the proposed approach; Section 4 presents the experimental results and ablation study; Section 5 draws comprehensive conclusions and presents future research prospects.

## 2. Related Works

In the following, we briefly discuss various research works addressing the class-imbalance problem that are relevant to this paper.

### 2.1. Transfer Learning

Transfer learning[8, 9] has become a ubiquitous strategy for boosting the training of models in classification tasks with long-tailed distributions, which transfers knowledge from the source domain to refine the classification performance on the target domain.

Transfer learning from head to tail classes aims to leverage the knowledge learned from abundant head classes to improve the recognition performance on under-represented tail classes. Liu et al. [10] utilized geometric information from relatively larger head class classifiers to enhance the weights of tail class classifiers. Online feature augmentation (OFA)[11] enhances tail classes by combining class-specific features from tail class samples with class-agnostic features from head class samples. Similarly, Sarah et al. [12] selected and recombined classifier features from common classes to obtain stronger tail class representations. Major-to-minor translation (M2m)[13] augments tail classes by translating samples from head classes through perturbation-based optimization. Model pre-training is also a transfer learning approach to address the problem of long-tailed distributions. Domain-specific transfer learning (DSTL)[8] proposes obtaining pre-trained models from the training set of the long-tailed distribution first and then fine-tuning them on a balanced dataset. Inspired by DSTL[8], we train the entire training set in the first stage and transfer the model in three stages.

Before clustering in the second stage, we expand the labeled set by sampling pseudo-labeled unlabeled data, using a sampling strategy inspired by class-rebalancing self-training (CReST)[14], which proposes selecting more tail class samples as pseudo-labels for training in each iteration.

### 2.2. Cost-sensitive Learning

Cost-sensitive learning [15, 16] adjusts the loss values of different classes to alleviate the problem of long-tailed distribution. Several methods have been proposed to address this issue. For example, LADE [17] and Balanced Softmax [18] directly apply loss re-weighting by training on the label frequencies of the samples. Focal Loss [19] focuses on the difficulty of sample classification by assigning higher weights to the tail classes that are harder to predict, and lower weights to the head classes that are easier to predict. Class-Balanced Loss [7] associates each example with a small neighborhood rather than an individual data point, measures whether there is overlap between data, and then reassigns the weights of each class's loss based on the effective number of samples for each class. Label-distribution-aware margin (LDAM)[20] proposes to first let the model learn the initial feature representation in an initial stage, and then perform re-weighting or re-sampling. The proposed method employs cost-sensitive learning in the initial stage, building upon Focal Loss [19] by introducing a sample weighting factor to alleviate the impact of long-tailed distribution of the dataset on model training.

### 2.3. Scheme-oriented Sampling

Scheme-oriented sampling (SOS) is a sample selection strategy of re-sampling [21, 22, 23, 24, 25, 26] that selects suitable samples from imbalanced datasets by considering the correlations and differences between different classes. Partitioning reservoir sampling (PRS)[24] balances the samples of each class in continual learning tasks by partitioning the dataset and maintaining a sample buffer for each partition. Bilateral-branch network (BBN)[25] proposes two network branches, where the traditional branch uses uniform sampling to simulate the original long-tailed training distribution, and the rebalancing branch uses a reverse sampler to sample more tail class samples. Dynamic curriculum learning (DCL)[26] dynamically constructs a curriculum based on the difficulty of each sample, allowing the model to gradually learn increasingly challenging examples. In this work, RSP is introduced as a novel method for pseudo-labeled sample sampling, which can alleviate the long-tailed distribution and narrow the domain gap between the labeled and unlabeled sets.

## 3. Approach

### 3.1. Overall Framework

A three-stage framework is proposed for long-tailed

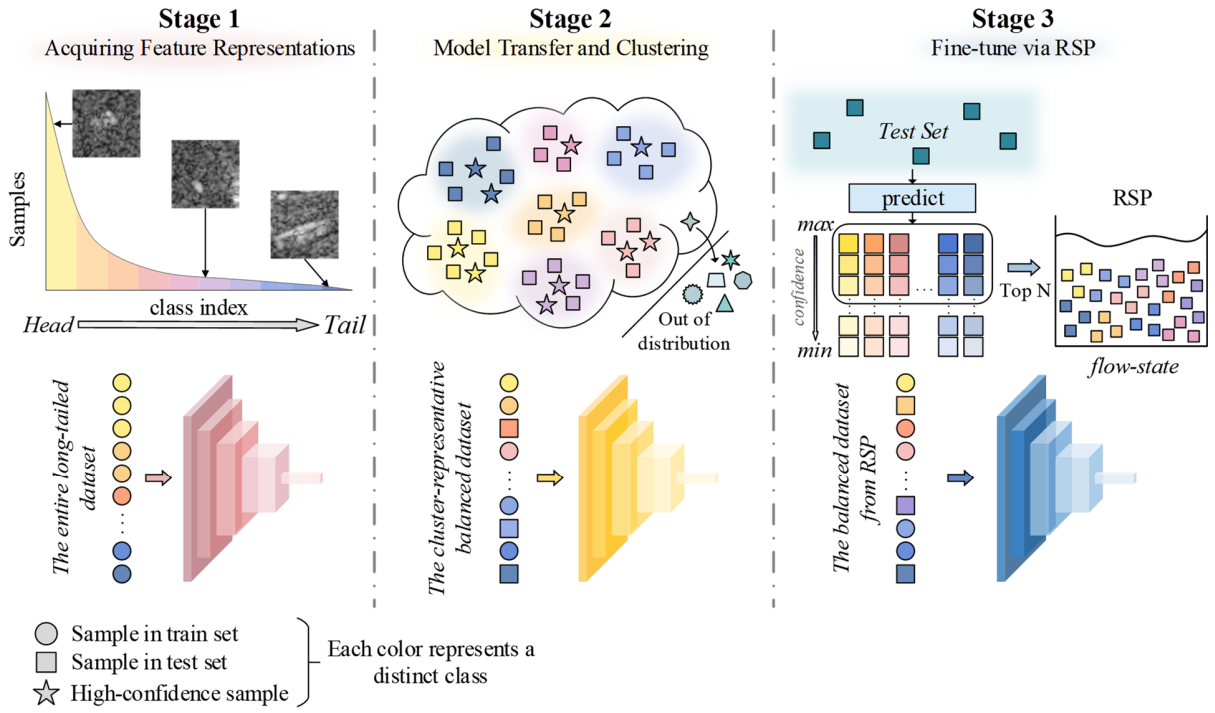


Figure 1. Overall framework of the proposed method. Stage 1: model pre-training with all long-tailed samples for representation learning. Stage 2: model transferring to a class-balanced dataset (consisting of representative samples from test set by clustering and samples from training set.). Stage 3: Fine-tune the second-stage model via Reliable Sample Pool.

image classification tasks. We utilize a pre-trained ResNet101 [6] backbone on ImageNet to extract features, and fully exploit the potential distribution and structure of the data through transfer learning and reliable pseudo-labeled samples. In addition, a novel semi-supervised sampling strategy is proposed to reduce dependence on labeled data and facilitate knowledge transfer from the labeled set domain to the unlabeled set domain, ultimately improving the model's performance. The overall framework of the proposed method is shown in Fig. 1.

In the first stage, we train the model on the entire long-tail dataset to learn rich feature representations. However, the data distribution exhibits severe imbalance, with the ratio of samples between the head and tail classes exceeding 500:1. Moreover, intra-class data imbalance is present, particularly in the head classes, with a significant number of redundant samples sharing similar attributes. To address this issue, we use Class-balanced Focal Loss [7] as the loss function, which re-weights the loss for each class based on its effective number of samples.

In the following stage, a class-balanced dataset is created by combining cluster-representative pseudo-labeled samples with selected samples from the training set. We then transfer the model obtained in the first stage on this dataset for further training, yielding a new classification model.

In the third stage, we propose a scheme-oriented sampling method called Reliable Sample Pool (RSP). Specifically, at each training round, we select the top-N unlabeled samples in test set with the highest model confidence in each class, and add their predicted labels to RSP to train the model for the next round. This allows the classifier to learn from different feature subsets in each training cycle. The model is fine-tuned with a few epochs using the reliable samples in RSP to leverage trustworthy predictions.

### 3.2. Cluster-based Pseudo Label Generation

The first-stage model acquires feature-abundant representations, which enables high-confidence samples to perform better in terms of representativeness and reliability, especially for tail-class samples. Such samples match the training set's features and are easier to classify correctly. To minimize noise introduction, we first use the model trained in the first stage to predict the labels of unlabeled data. Then, we select a subset of high-confidence samples and clustered them with the remaining unlabeled samples. Specifically, we utilize the pre-trained VGG-16 [27] to extract the features and integrate the same clustering results from DBSCAN [28] and k-means, thereby mining the cluster structure of unlabeled data to obtain representative

images and generate reliable pseudo-labels. Additionally, DBSCAN is used to discard certain outlier clusters that could have a detrimental effect on the model's classification performance in the next stage, due to the existence of a distribution shift between the labeled and unlabeled sets, which includes negative samples.

Note that the pseudo-labeled dataset used for training in the second stage contains a significant proportion of tail class samples. Considering that the first-stage model trained on the long-tailed training set may be biased towards the head classes, which tends to generate more pseudo-labels with high confidence for those classes. Hence, the high-confidence samples of head classes may not necessarily belong to the data distribution of the original training set. Finally, these selected pseudo-labels are adopted to create a class-balanced dataset, on which we aim to gain a model that can learn domain knowledge from the unlabeled set.

### 3.3. Reliable Sample Pool

Most existing methods for long-tailed distribution do not consider the distribution gap between the test set and the training set. In practical scenarios, data within the same class may originate from different domains, and the distribution of the test set is typically unknown in advance. In our task, there are numerous OOD samples in the test set, which may not belong to any class in the training set. To address this issue, we propose a Reliable Sample Pool (RSP) to serve as a trustworthy pseudo-label source for further model training.

During the third-stage training, we maintain a reliable and dynamically changing sample pool that is updated at every round to ensure that the included samples are always the most representative test samples that fit the distribution of training set. Specifically, at each round, we leverage the previous epoch's model predictions on the test set. The Reliable Sample Pool (RSP) stores the top-N samples with the highest confidence scores in each class. Assuming there are $M$ rounds, in the $(m-1)^{th}$ $(m \in \{1, \cdots, M\})$ epoch, $P^{m-1}(y|x)$ denotes the probability of input sample x belonging to class y. The pseudo-labels sample set stored in the RSP of class y in the $m^{th}$ round is defined as:

$$RSP_y^m = \left\{ x_i \middle| \begin{matrix} P^{m-1}(y|x_i) \geq P^{m-1}(y|x_j) \\ for\ all\ x_j \in \mathcal{X}_y\ and\ i \leq N \end{matrix} \right\}. \quad (1)$$

The Reliable Sample Pool (RSP) in the $m^{th}$ iteration is defined as:

$$RSP^m = [RSP_1^m, RSP_2^m, \cdots, RSP_C^m], \quad (2)$$

where $\mathcal{X}_y$ represents the set of all samples predicted as class y, $x_i$ denotes the $i^{th}$ sample in $P^{m-1}(y|x_i)$ ordered by its confidence score, $N$ is a pre-defined threshold that indicates the capacity of each class in RSP. $C$ is the number of classes. $RSP^m$ is the set of all reliable pseudo-labeled samples in the $m^{th}$ round, which will be added to the training set for the $m^{th}$ round.

Note that the initial dataset for the third stage of training is an evenly sampled balanced dataset obtained from the training set. The capacity of RSP is a hyperparameter that can be dynamically adjusted based on factors such as the dataset size, model complexity, and changing rate of prediction results between rounds. In the third stage, we use the samples in RSP together with the class-balanced dataset sampled from the training set to fine-tune the model. The mutually reinforcing positive process leads to a gradual improvement in the model's performance and the accuracy of predicted samples, which enhances the model's robustness against distribution shift between training and test sets, improves its ability to detect OOD samples, and boosts its confidence in in-distribution samples.

### 3.4. Loss Function

Due to the imbalanced class distribution, the first-stage model tends to over-emphasize classes with larger sample sizes while overlooking those with smaller ones. However, more data does not necessarily result in better performance, especially when there is information overlap among samples within the same class (exacerbates the intra-class long-tail distribution). To address this issue, we adopt the Class-balanced Focal Loss [7] as the loss function, which is a variant of the Focal Loss [19]. By utilizing Focal Loss, difficult-to-classify samples are assigned higher weights and gain more attention from the model. To tackle data overlap, the effective number of samples is used as the class weights to balance the training sample numbers across different classes in the dataset.

Class-balanced Focal Loss can be defined as:

$$\mathcal{L}_{cb\ focal}(y, k) = -\frac{1-\beta}{1-\beta^{n_k}} \mathcal{L}_{focal}(y, k), \quad (3)$$

$$\mathcal{L}_{focal}(y, k) = -\sum_{i=1}^{C} (1 - p_i^t)^{\gamma}\ log(p_i^t), \quad (4)$$

where $n_k$ is the number of samples in class $k$, $\mathrm{y} = [y_1, y_2, \cdots, y_C]^T$ represents the predicted outputs of all classes, and $C$ is the number of classes. The Class-balanced Focal Loss includes a weighting factor $\frac{1-\beta}{1-\beta^{n_k}}$, where $\beta \in [0, 1)$ is a hyperparameter that controls the rate at which hyperparameter of Focal Loss that controls the weights of hard-to-classify samples. $p_i^t$ represents the predicted the

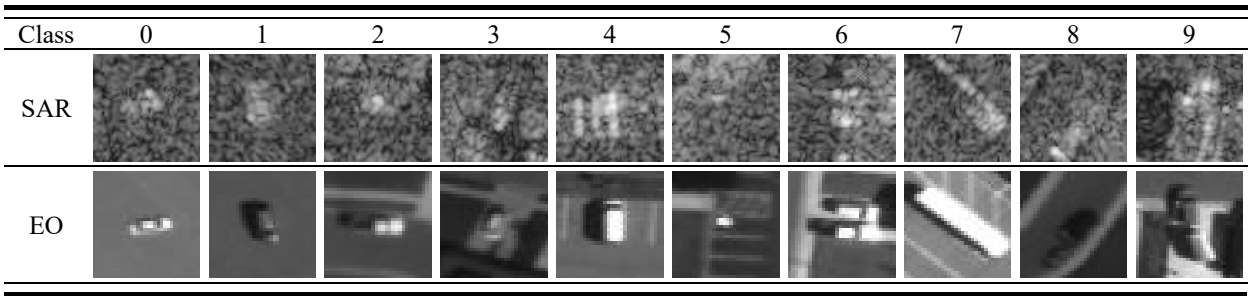


Figure 2. Examples of 10 classes in the training set with corresponding EO and SAR images.

effective samples increase with $n_k$. $\gamma$ is a probability of the model, where $i$ denotes the predicted class:

$$p_i^t = sigmoid(y_i^t) = \frac{1}{1 + exp(-y_i^t)}, \quad (5)$$

where $y_i^t$ is defined as:

$$y_i^t = \begin{cases} y_i, & if\ i = k. \\ -y_i, & otherwise. \end{cases} \quad (6)$$

## 4. Experiments

### 4.1. Experimental Setup

**Datasets.** The dataset for the PBVS 2023 Multi-modal Aerial View Object Classification Challenge consists of aerial view SAR images of 10 categories of vehicles. Fig. 2 illustrates the comparison between the images of each class captured by the optical sensor and SAR sensor. Table 1 presents the details of the training set, which comprises 459,262 images and suffers a severe long-tailed distribution. The proportion of class 0 is approximately 80%, while the proportion of class 9 is less than 0.16%. The sizes of the images in the dataset are not entirely uniform, with an average size of approximately 56 × 56 pixels. To facilitate experimentation, we resize all the images to 56 × 56 pixels.

**Evaluation metrics.** To quantitatively evaluate our proposed solution, four evaluation metrics were established for this challenge: (1) Top-1 Accuracy; (2) Area Under the Receiver Operating Characteristic curve (AUROC); (3) True negative rate (TNR) at 95% true positive rate (TPR); and (4) Total Score. AUROC, which displays the relationship between the true positive rate $TPR = TP/(TP + FN)$ and false positive rate $FPR = FP/(FP + TN)$, can be interpreted as the probability that a positive example has a higher value from the detector than a negative example [29]. The TNR in TNR at TPR 95% is defined as $TNR = TN/(TN + FP)$. Total Score is a comprehensive metric weighted by the organizers based on the first three metrics, which considers both error and success prediction and in- and out-of-distribution detection.

Table 1. Details of the training dataset used in PBVS 2023 Challenge Track 1.

| Class | Type | Samples | Percent (%) |
|---|---|---|---|
| 0 | sedan | 364,228 | 79.31 |
| 1 | suv | 43,642 | 9.50 |
| 2 | pickup truck | 24,420 | 5.32 |
| 3 | van | 17,159 | 3.74 |
| 4 | box truck | 3,414 | 0.74 |
| 5 | motorcycle | 2,351 | 0.51 |
| 6 | flatbed truck | 1,233 | 0.27 |
| 7 | bus | 1,130 | 0.25 |
| 8 | pickup truck with trailer | 971 | 0.21 |
| 9 | flatbed truck with trailer | 714 | 0.16 |

Table 2. Classification performance of different backbones using initial experimental settings in development phase.

| Backbone | Top-1 Accuracy (%) | AUROC | TNR at tpr95 | Total Score |
|---|---|---|---|---|
| MobileNetV3[30] | 50.91 | 0.76 | **0.15** | 0.57 |
| ResNet-34 [6] | 54.16 | 0.74 | 0.10 | 0.59 |
| ResNet-50 [6] | 53.12 | 0.73 | 0.11 | 0.58 |
| ResNet-101 [6] | **55.71** | **0.77** | 0.12 | **0.61** |

Note that higher values of these four metrics indicate better performance of the model.

**Implementation details.** Throughout our experiments, the SGD optimizer was utilized and a single GeForce RTX3090 GPU was used in all stages. Specifically, during the initial and secondary stages, we set the learning rate to 1e-3 and apply the CosineAnnealingLR scheduler to adjust it. The training process spanned a maximum of 300 iterations, using batch sizes of 128 and 32, respectively. In the third stage, the model was fine-tuned with a lower learning rate of 1e-5, leveraging the ReduceLROnPlateau scheduler. The batch size was set to 32. The number of epochs for fine-tuning the model in the third stage is related to the value of N in RSP, which can be adjusted according to N. When N is set to decrease with training starting from 80, the number of epochs is 50.

Table 3. Ablation study in development and test phases.

| Stage | | | Development Phase | | | | Test Phase | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | Top-1 Accuracy (%) | AUROC | TNR at tpr95 | Total Score | Top-1 Accuracy (%) | AUROC | TNR at tpr95 | Total Score |
| ✓ | | | 53.90 | 0.23 | 0.04 | 0.46 | - | - | - | - |
| ✓ | ✓ | | 56.88 | 0.82 | 0.07 | 0.63 | 61.20 | 0.70 | **0.04** | 0.63 |
| ✓ | ✓ | ✓ | **57.01** | **0.85** | **0.14** | **0.64** | **63.20** | **0.71** | 0.03 | **0.65** |

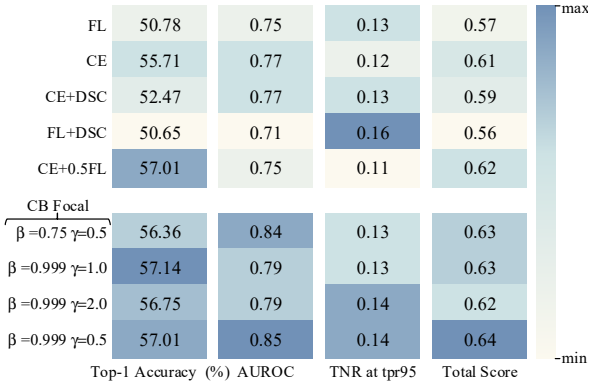


Figure 3. Classification performance of the proposed method trained with different loss functions in development phase.

## 4.2. Ablation Study

**Backbone selection.** During the development phase, a series of experiments were conducted to assess the performance of various backbones in the three-stage framework under the initial experimental settings. Multiple backbones, including MobileNetV3 [30], ResNet-34 [6], ResNet-50 [6], and ResNet-101 [6], were evaluated for their suitability in the proposed method.

As shown in Table 2, we compared the performance of these backbones on the validation set and observed that ResNet-101 outperformed other models, which consistently achieved better results across almost every evaluation metric. Therefore, we selected ResNet-101 [6] as the backbone for the proposed method.

**Investigation on loss function.** The first-stage model was trained on the entire dataset, which exhibited a long-tail distribution. To mitigate this, re-weighting was employed by utilizing different loss functions, including Cross-Entropy Loss (CE), Focal Loss (FL)[19], DSC Loss (DSC)[31], Class-Balanced Focal Loss (CB Focal)[7], as well as their multiple combinations. Furthermore, since the second and third-stage datasets were class-balanced, we used CE as the loss function in these stages.

Fig. 3 summarizes the performance of our method using different loss functions in the first stage on the validation set. It indicated that the model trained with FL alone did not outperform the model trained with CE. However, a combination of both could achieve better performance. Notably, the model trained with CB Focal achieved the best performance. Hence, CB Focal was selected as the loss function for the proposed method.

We further investigated the impact of two hyperparameters in CB Focal on the classification performance. $\beta$ was set to 0.999, while $\gamma$ was varied between 0.5, 1.0, and 2.0. As shown in Fig. 3, the model achieved the highest values for all metrics when $\beta$ was set to 0.999 and $\gamma$ was set to 0.5. Moreover, when $\gamma$ was fixed at 0.5 and $\beta$ was set to 0.75, the model also yielded competitive results. Ultimately, $\beta = 0.999$ and $\gamma = 0.5$ were selected as the hyperparameters for CB Focal.

**The effectiveness of different stages.** The experiments were conducted to evaluate the effectiveness of each stage of the proposed architecture. The experimental results are summarized in Table 3. In the first stage, the model was trained on an imbalanced dataset, resulting in poor performance, particularly with a low AUROC. In the second stage, a balanced dataset was created by generating pseudo-labeled samples, leading to a significant improvement in model performance. In the third stage, we employed fine-tuning with RSP to further enhance the performance of the model. The results of ablation study demonstrate the effectiveness of the three-stage framework in improving model performance on long-tailed datasets.

## 4.3. Comparison with Other Methods

Table 4 presents the top four results during the test phase of this challenge. The proposed method ranks first in terms of Total Score and achieves a competitive Top-1 Accuracy of 63.20%. Our method exceeds the second-place by 3.35% in Top-1 Accuracy and achieves comparable performance in AUROC. Additionally, during the development phase, we compared our method with two other approaches, A Two-Stage Shake-Shake Network [32] and Bilateral-branch Network (BBN)[25]. The former is a top-performing method of this challenge in 2022, and the latter is one of the state-of-the-art methods for addressing long-tailed distribution. As shown in Table 5, our method achieved higher Total Score, Top-1 Accuracy, and AUROC than these two methods, which further validates the superiority of our approach in addressing the challenges.

Table 4. Competition results in test phase.

| Team | Top-1 Accuracy (%) | AUROC | TNR at tpr95 | Total Score |
|---|---|---|---|---|
| Team A | 59.20 | 0.80 | 0.35 | 0.64 |
| Team B | 53.15 | **0.85** | **0.50** | 0.61 |
| Team C | 59.85 | 0.64 | 0.05 | 0.61 |
| Ours | **63.20** | 0.71 | 0.03 | **0.65** |

Table 5. Comparison with the proposed method with other methods in development phase.

| Method | Top-1 Accuracy (%) | AUROC | TNR at tpr95 | Total Score |
|---|---|---|---|---|
| BBN [25] | 52.46 | 0.70 | 0.06 | 0.57 |
| A Two-Stage Shake-Shake Network [32] | 46.23 | 0.65 | **0.14** | 0.51 |
| Ours | **57.01** | **0.85** | **0.14** | **0.64** |

## 5. Conclusion

We proposed a three-stage framework to address the challenges in aerial view object classification under long-tail distribution. The proposed approach not only mitigates the interference of negative samples, but also effectively tackles domain shift between different sets. We introduce a Reliable Sample Pool to enhance the model's confidence in predicting in-distribution data. The experimental results demonstrate the outstanding performance of our approach, which ranked first in the PBVS 2023 Multi-modal Aerial View Object Classification Challenge Track 1. Future work needs to investigate the sampling variation method of the Reliable Sample Pool, considering both confidence and predictive distribution probability.

## References


[1] F. Chen, H. Balzter, F. Zhou, P. Ren, and H. Zhou, "DGNet: Distribution Guided Efficient Learning for Oil Spill Image Segmentation," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-17, 2023.

[2] X. Leng, K. Ji, S. Zhou, and X. Xing, "Ship detection based on complex signal kurtosis in single-channel SAR imagery," IEEE Transactions on Geoscience Remote Sensing, vol. 57, no. 9, pp. 6447-6461, 2019.

[3] L. Zhang et al., "Domain knowledge powered two-stream deep network for few-shot SAR vehicle recognition," IEEE Transactions on Geoscience Remote Sensing, vol. 60, pp. 1-15, 2021.

[4] S. Chen, H. Wang, F. Xu, Y.-Q. Jin, and r. sensing, "Target classification using the deep convolutional networks for SAR images," IEEE transactions on geoscience and remote sensing, vol. 54, no. 8, pp. 4806-4817, 2016.

[5] K. Tang, M. Tao, J. Qi, Z. Liu, and H. Zhang, "Invariant feature learning for generalized long-tailed classification," In: ECCV, pp. 709-726, 2022.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In: CVPR, pp. 770-778, 2016.

[7] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," In: CVPR, pp. 9268-9277, 2019.

[8] Y. Cui, Y. Song, C. Sun, A. Howard, and S. Belongie, "Large scale fine-grained categorization and domain-specific transfer learning," In: CVPR, pp. 4109-4118, 2018.

[9] J. Wang, T. Lukasiewicz, X. Hu, J. Cai, and Z. Xu, "Rsg: A simple but effective module for learning imbalanced datasets," In: CVPR, pp. 3784-3793, 2021.

[10] J. Liu, Y. Sun, C. Han, Z. Dou, and W. Li, "Deep representation learning on long-tailed data: A learnable embedding augmentation perspective," In: CVPR, pp. 2970-2979, 2020.

[11] P. Chu, X. Bian, S. Liu, and H. Ling, "Feature space augmentation for long-tailed data," In: ECCV, pp. 694-710, 2020.

[12] S. Parisot, P. M. Esperança, S. McDonagh, T. J. Madarasz, Y. Yang, and Z. Li, "Long-tail recognition via compositional knowledge transfer," In: CVPR, pp. 6939-6948, 2022.

[13] J. Kim, J. Jeong, and J. Shin, "M2m: Imbalanced classification via major-to-minor translation," In: CVPR, pp. 13896-13905, 2020.

[14] C. Wei, K. Sohn, C. Mellina, A. Yuille, and F. Yang, "Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning," In: CVPR, pp. 10857-10866, 2021.

[15] H.-J. Ye, H.-Y. Chen, D.-C. Zhan, and W.-L. Chao, "Identifying and compensating for feature deviation in imbalanced deep learning," arXiv preprint arXiv:2001.01385, 2020.

[16] Y. Sun, M. S. Kamel, A. K. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," Pattern recognition, vol. 40, no. 12, pp. 3358-3378, 2007.

[17] Y. Hong, S. Han, K. Choi, S. Seo, B. Kim, and B. Chang, "Disentangling label distribution for long-tailed visual recognition," In: CVPR, pp. 6626-6636, 2021.

[18] J. Ren, C. Yu, X. Ma, H. Zhao, and S. Yi, "Balanced meta-softmax for long-tailed visual recognition," In: NeurIPS, vol. 33, pp. 4175-4186, 2020.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," In: ICCV, pp. 2980-2988, 2017.

[20] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," In: NeurIPS, vol. 32, 2019.

[21] B. Kang et al., "Decoupling representation and classifier for long-tailed recognition," In: ICLR, 2020.

[22] J. Byrd and Z. Lipton, "What is the effect of importance weighting in deep learning?," In: ICML, pp. 872-881, 2019.

[23] T. Wang et al., "The devil is in classification: A simple framework for long-tail instance segmentation," In: ECCV, pp. 728-744, 2020.

[24] C. D. Kim, J. Jeong, and G. Kim, "Imbalanced continual learning with partitioning reservoir sampling," In: ECCV, pp. 411-428, 2020.

[25] B. Zhou, Q. Cui, X.-S. Wei, and Z.-M. Chen, "Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition," In: CVPR, pp. 9719-9728, 2020.

[26] Y. Wang, W. Gan, J. Yang, W. Wu, and J. Yan, "Dynamic curriculum learning for imbalanced data classification," In: ICCV, pp. 5017-5026, 2019.
[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
[28] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," In: KDD, vol. 96, no. 34, pp. 226-231, 1996.
[29] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," In: ICLR, 2017.
[30] A. Howard et al., "Searching for mobilenetv3," In: ICCV, pp. 1314-1324, 2019.
[31] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced NLP tasks," In: ACL, 2020.
[32] G. Li, L. Pan, L. Qiu, Z. Tan, F. Xie, and H. Zhang, "A Two-Stage Shake-Shake Network for Long-Tailed Recognition of SAR Aerial View Objects," In: CVPRW, pp. 249-256, 2022.