4th Multi-modal Aerial View Image Challenge: SAR Classification - PBVS 2025

Nathan Inkawhich[1]*, Claire Thorp[1]*, Justice Wheelwright[1], Oliver Nina[1],
Dylan Bowald[1], Angel Sappa[2,3], Erik Blasch[1]
[1] Air Force Research Laboratory, USA
[2] ESPOL Polytechnic University, Ecuador
[3] Computer Vision Center, Spain

Abstract

The Multi-modal Aerial View Image Challenge - Classification Track (MAVIC-C) continues to push the boundaries of multi-modal object recognition by encouraging researchers to innovate models that leverage both Synthetic Aperture Radar (SAR) and Electro-Optical (EO) imagery. This paper analyzes the outcomes of the new iteration of this challenge and emphasizes the critical role of EO and SAR data fusion in remote sensing tasks. This year MAVIC-C saw impressive developments of sophisticated multi-modal approaches that address the distinct properties and challenges inherent to the data. This year's challenge notably builds on insights from previous iterates: in 2021 we demonstrated the potential of EO and SAR integration; in 2022 and 2023 we explored the capabilities of multi-modal frameworks; and in 2024 we examined model robustness in out-of-distribution scenarios. This year, we started with the same challenge design as 2024 and asked teams to further advance techniques for improving accuracy and of out-of-distribution detection, which builds model ***robustness****. Overall, this manuscript provides an in-depth investigation of the methodologies of top-performing teams and analyzes participant's performance on a sequestered test set.*

1. Introduction

The ability to accurately detect, identify, and classify objects within imagery is a cornerstone of modern remote sensing (RS) applications. Object Recognition (OR) models, which strive to perform these tasks automatically and are often built with Machine Learning (ML) models [2, 11], may leverage a variety of RS modalities and play a crucial role in the automatic exploitation of the data [28]. While OR shares similarities with traditional object detection tasks, its application to aerial platforms presents unique challenges. As viewed with an Electro-Optical (EO) sensor (see rows 1 and 3 of Figure 1), the perspective from aerial vehicles often results in limited sensor resolution, potentially reducing targets to just a few pixels. This, along with factors like varying illumination and atmospheric conditions, demands robust OR models specifically designed to handle the complexities of aerial imagery.

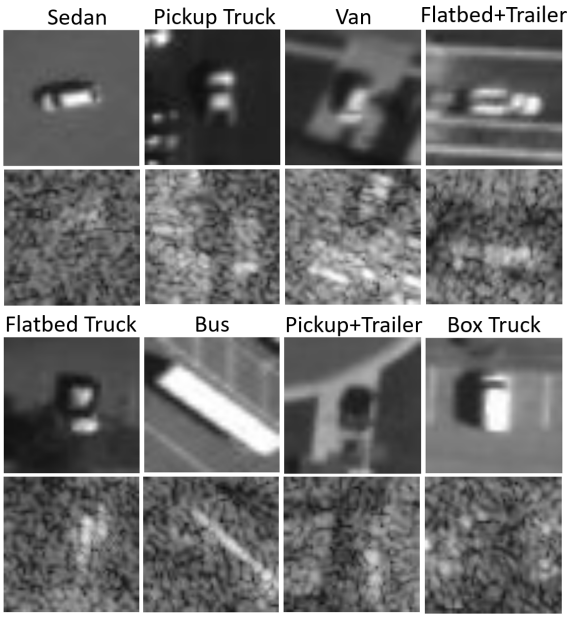

Figure 1. Example EO/SAR pairs for eight classes of data in the challenge dataset.

Synthetic Aperture Radar (SAR) offers a compelling solution for its all-weather, all-time surveillance capabilities as well as the resolution being independent of range. However, SAR data is not without its limitations (see rows 2 and 4 of Figure 1). Atmospheric conditions can introduce attenuation, shadows can obscure targets, and multi-bounce effects can complicate signal interpretation [14, 33]. Integrating EO imagery with SAR data provides a powerful means to overcome these challenges, but their fusion introduces new difficulties.

* equal contribution

Besides the properties of the sensing modalities, a key challenge for effective OR *algorithms* is the ability to handle out-of-distribution (OOD) samples - inputs that are anomalous w.r.t. the training data - without compromising in-distribution accuracy [12, 13]. In high-stakes scenarios, misclassifications can have significant consequences. However, it is well known that ML models are bad at distinguishing these confusing inputs [8], making OOD detection a critical algorithmic challenge [9] and a key aspect of model *robustness*. Often, ML practitioners tackle the robustness challenge by providing credibility scores which indicate the level of uncertainty associated with the prediction. Interestingly, the 2025 MAVIC-C challenge operates at the crossroads of the aforementioned sensor modality issues and OR algorithmic issues, and provides a valuable research-playground for exploring how to develop highly-performant and robust OR algorithms with multi-modal data.

This paper provides a comprehensive analysis of the new iteration of the MAVIC-C challenge and builds on the results and experiences of the previous four years [22, 24–26]. Similar to the 2024 challenge, the 2025 version prioritizes high classification accuracy and reliable OOD detection. The challenge utilizes the UNIfied COincident Optical and Radar for recognitioN (UNICORN-V2) dataset as a rigorous benchmark for evaluating performance [26]. Uniquely, this dataset offers *paired* SAR and EO images for 20 categories of objects, which are deliberately split into in distribution detection (IDD) vs OOD sets across train, val, and test partitions. Figure 1 shows example (EO, SAR) pairs for several categories. This year's results were particularly interesting. We observed strong performance from the top entrants who utilized new ML techniques to improve performance by leveraging the multi-modal feature information. Of particular note, this year we saw an improvement in OOD detection scores yielding an improvement in robust performance. Details on methodologies and quantitative results will be discussed in the following sections.

The remainder of the paper is organized as follows. Section 2 provides key details of the 2025 challenge, including dataset and category splits. Section 3 outlines the quantitative results of the top-performing methods. Section 4 delves into the details of the leading approaches. Finally, Section 5 concludes the paper with key takeaways and future directions.

2. Challenge

Accurately classifying objects from aerial imagery is crucial for a wide range of applications, from environmental monitoring to disaster response. The Multi-modal Aerial View Image Classification Challenge (MAVIC-C) pushes the boundaries of this field by focusing on the combined potential of SAR and EO imagery. Building on the successes of previous MAVIC-C and MAVOC challenges [22, 24–26], the 2025 iteration emphasizes robust performance and multi-modal learning. Participants grapple with the unique properties of SAR data, particularly the presence of SAR shadows and tilted perspectives, demanding innovative solutions for robust SAR image classification. Similarly, the challenge's EO data is of low resolution and many of the object categories are similar in nature, making for a fine-grained recognition challenge.

Figure 2. The aligned scene of the full UNICORN dataset before chipping is performed [17].

Practically, the training and testing data are presented to the OR model in image *chips*, meaning the localization step has already happened and the ML-based OR model must perform classification. Classifier performance is rigorously assessed on traditional top-1% accuracy and also on the ability to identify OOD samples, as measured by the Area Under the Receiver Operating Characteristic (AUC) curve [9].

2.1. The Train/Test Split

During the training phase, participants are given access to (SAR, EO) pairs for the ten IDD object categories shown in Table 1. We intuit that the combination of SAR and EO imagery for training can allow the OR models to learn useful features from the multi-modal representations. In the testing phase, participants are only given SAR data. By using this framework the computational expense of the traditional pre-processing can be eliminated while greatly improving decision-making processes. The main challenge, however, is working with a multi-modality training dataset and trying to learn features from the SAR+EO data that can assist in the SAR-only recognition.

Table 1. Details of the UNICORN V2 Dataset used as the in-distribution classes in this challenge (counts represent the number of (EO, SAR) pairs).

Class #	Vehicle Type	# Train	# Val	# Test
0	sedan	364,291	77	200
1	SUV	43,401	77	200
2	pickup truck	24,158	77	200
3	van	16,890	77	200
4	box truck	2,896	77	200
5	motorcycle	1,441	77	200
6	flatbed truck	898	77	200
7	bus	612	77	200
8	pickup truck w/ trailer	695	77	200
9	semi truck w/ trailer	353	77	200
	Total	455,635	770	2000

Table 2. Details of the UNICORN V2 Dataset used as the out-of-distribution classes in this challenge (counts represent the number of (EO, SAR) pairs).

Class #	Vehicle Type	# Train	# Val	# Test
0	van w/ trailer	-	77	-
1	other	-	77	2000
2	dismount	-	77	-
3	semi	-	7	-
4	SUV w/ trailer	-	77	-
5	flatbed truck w/ trailer	-	77	-
6	plane	-	77	-
7	bicycle	-	24	-
8	dump truck	-	77	-
9	sedan w/ trailer	-	77	-
	Total	-	647	2000

2.2. UNICORN V2 Dataset

Last year the MAVIC-C challenge introduced the second version of the UNIfied COincident Optical and Radar for recognitioN (UNICORN V2) dataset [26]. The original source dataset, UNICORN, was developed in 2008 and contains both EO and SAR data. It was collected through large-format sensors and sourced from aerial surveys over Dayton, OH. It contains Wide Area Motion Imagery (WAMI) EO data and Wide Area SAR data [29]. The EO and SAR data was rigorously aligned through both geo-registration and homography techniques. It was then labeled by human annotators and chipped for the challenge. Figure 2 shows an overlay of the EO and SAR full scene data (note, the chips in Figure 1 are contained within this figure). There are 20 total categories of objects in UNICORN V2. The top-10 most prevalent categories are used for the IDD set and described in Table 1. The remaining 10 less-prevalent categories are used for the OOD set and represent *confusers*. These OOD classes are shown in Table 2. The dataset itself is publicly accessible. The dataset was validated on models from the 2022 challenge prior to being introduced.

2.3. Metrics & Evaluation

MAVIC-C challenge submissions are evaluated on two main metrics: top-1 accuracy and AUC. The AUC provides a nuanced measurement of the model's ability to distinguish OOD samples (unknowns) from In Distribution (ID) samples (knowns). AUC demonstrates the rate the model outputs true positives vs false positives using a 0 to 1 scale. An AUC score close to 1 indicates the True Positive Rate (TPR) is high and the False Positive Rate is low. Meaning that most positives and negatives are being predicted correctly. Therefore, the closer the AUC is to 1 the better the model is performing. AUC is meant to measure the *robust* performance of the model in the face of unknowns. The other metric is of course a standard accuracy of the classification model.

The final evaluation of contestant's submission is a weighted average of ID accuracy and OOD detection measured on a sequestered set of 4,000 total test images. Within this test set there are 200 examples for each of the 10 ID classes (so, 2000 total ID samples) and 2000 OOD samples. The weighting of the final evaluation is as follows:

$$\text{Score} = (0.75 \cdot \text{ Accuracy}) + (0.25 \cdot \text{ AUC}). \quad (1)$$

Our idea behind the weighting is that if accuracy is too low, the model will not be usable for target recognition by users. So, in the scoring we give a higher weight to accuracy and a slightly lower priority to OOD detection.

During the testing phase of the competition, teams are allowed up to ten submissions per day. During the evaluation phase, teams submit their label predictions and credibility score to be evaluated on the competition server. Teams are allowed up to 12 submissions, which prevents them from effectively fine-tuning on the test dataset. Results are made visible during both phases.

2.4. Challenge Phases

The challenge began in mid January 2025, and the test data was released in late February 2025. The testing phase ended on March 12, 2025 with team submissions finalized.

3. Challenge Results

The 2025 MAVIC-C challenge received substantial engagement, affirming the community's interest in developing multi-modal solutions for OR in RS tasks. A total of 105 teams registered for participation. During the development

Table 3. Top-10 Teams for 2025 MAVIC-C Challenge (**bold** represents best score, blue is second best).

Rank	Team	Total Score	Accuracy	AUC
1	**TongJi-CV**	**0.43**	**31.78**	0.77
2	up6	0.42	27.56	0.87
3	crys4al	0.41	29.61	0.76
4	momo_mo	0.41	28.78	0.76
5	shdsff	0.39	22.61	**0.88**
6	TianyiYu	0.39	22.83	0.87
7	FangzhouHan	0.37	20.06	**0.88**
8	HuangRoman	0.34	21.11	0.73
9	KNUNIST	0.34	22.39	0.68
10	Yc2	0.33	21.83	0.66

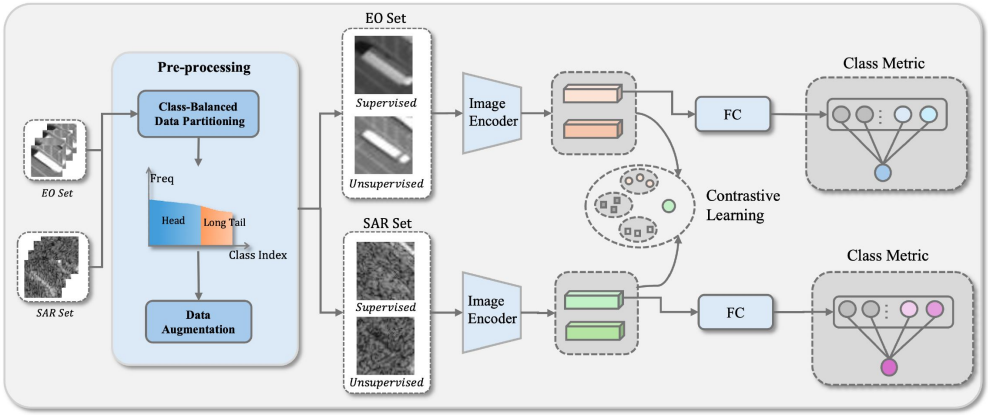

Figure 3. Overview of the TongJi-CV architecture (1^{st} place overall).

phase, 36 of these teams submitted their algorithms for preliminary evaluation. The participation slightly adjusted in the testing phase, with 14 teams submitting valid algorithms for rigorous assessment. Over the course of the challenge, we received 1187 total submissions.

The results from the top-10 performers are shown in Table 3. This year we saw comparable performance in accuracy as last year but a notable improvement in OOD detection (max AUC=0.88). Interestingly, the most accurate models did not necessarily have the best OOD detection scores. This indicates that there is still substantial room for improvement in coming years and can only be explained by parsing through the methodologies of the individual submissions. As a highlight, this year we saw many exciting and new approaches to the challenge which are leveraging modern concepts in ML and Deep Learning (DL). For example, top teams leveraged contrastive learning objectives, feature matching terms, knowledge distillation, and vision transformers. We leave it to the next section to discuss more details.

4. Challenge Methods

This section describes the methodologies from some of the top performing teams. The descriptions have been adapted from the individual author submissions and we thank them for their willingness to contribute.

4.1. Rank 1: TongJi-CV

Team Members: Hongli Liu, Yu Wang and Shengjie Zhao

4.1.1. Overview

The Tongji-CV team proposes a cross-modal semi-supervised collaborative optimization domain adaptation framework to enhance transferable representations and generalization capability across heterogeneous modalities.

As illustrated in Fig. 3, the core process includes several components. First, is a semi-supervised class-balanced data partitioning and data augmentation strategy to handle labeled and unlabeled data. Second, is a dual-stream feature extraction network which adopts a dual-branch architecture with Efficient-Net [34] and ResNet-50 [7] to extract local

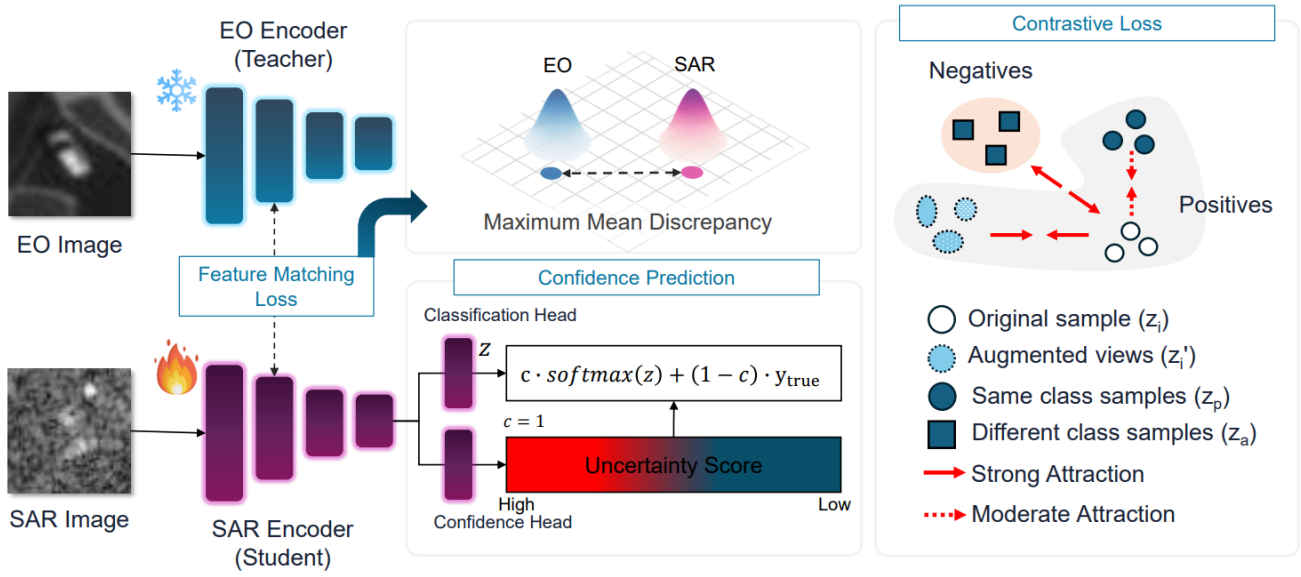

Figure 4. *KNUNIST* overview with knowledge distillation from EO to SAR. The EO encoder (teacher) is first pre-trained and frozen, and then the SAR encoder (student) is trained with multiple components: classification loss, feature matching loss, contrastive loss, and confidence prediction.

and global features from EO and SAR modalities. Third, is a cross-modal contrastive domain adaptation module which integrates contrastive loss to guide domain-invariant feature learning. Finally, a category metric learning term ensures accurate classification of aerial targets.

The Tongji team also utilizes Focal Loss [20] for classification optimization and Sliced Wasserstein Distance [30] for cross-modal alignment. In the computation of cross-modal alignment loss, weights of 0.8 and 0.2 are assigned to supervised and unsupervised scenarios, respectively. The model is first pre-trained on the EO dataset, and the learned weights are transferred to the SAR task to acquire cross-modal prior knowledge.

The Tongji model contains 46.54M parameters. It was trained with a batch size of 64 using the AdamW optimizer with learning rate set to 1e-3 and run for 100 epochs. Additionally, to enhance model stability, Exponential Moving Average (EMA) [1] and Gradient Clipping are employed during training to prevent gradient explosion.

The experiments are conducted on two NVIDIA RTX 3090 GPUs (each with 24GB of VRAM), with a total training time of approximately 28 hours. Quantitative results indicate that the Tongji-CV team achieved the best performance in the SAR Classification track, with a Total Score of 0.43 and a Top-1% Accuracy of 31.78%. These results demonstrate that the proposed method has achieved state-of-the-art performance in SAR classification. Source code can be found here: https://github.com/HongliLiu1/PBVS2025_SARClassification.

4.2. Rank 9: KNUNIST

Team Members: Jeongho Min, Hyeonjin Kim, Jaehyup Lee and Jaejun Yoo

4.2.1. Overview

PBVS 2025 MAVIC-C uniquely provides pairs of EO and SAR images during training and only SAR images during testing. While SAR images have low resolution and noise (e.g., speckle noise), EO images provide clearer structure information. Previous studies have confirmed that using SAR and EO together is better than using SAR images alone for training OR models [25]. The *KNUNIST* team leverages this feature by actively utilizing EO data to improve the classification performance of SAR images. They adopt a knowledge distillation framework in which the EO model acts as a teacher, transferring that knowledge to the student SAR model. As shown in Figure 4, their approach addresses the domain gap between SAR and EO through a multi-modal learning framework that combines classification learning with feature alignment and contrastive learning.

This team's main contributions are summarized as follows:

- They propose a feature matching approach based on maximum mean discrepancy (MMD). This aligns SAR features with pre-trained EO features to transfer knowledge across the modalities effectively;
- They utilize a contrastive learning objective to solve the problem of high intra-class variation and inter-class simi-

larity in SAR imagery, which enhances the model's ability to discriminate between similar-looking objects from different classes;
- Finally, they incorporate a confidence prediction mechanism to handle OOD samples, improving the robustness of the model when faced with ambiguous and noisy SAR inputs.

4.2.2. Proposed Approach

Architecture: The proposed architecture employs a teacher-student framework with ResNet101[6] backbones for both SAR and EO, as shown in Fig. 4. The EO model serves as the teacher, guiding the student SAR model via knowledge distillation. The EO encoder is first pre-trained and then frozen during distillation training. A key component of their approach is the confidence prediction head, which estimates the reliability of each classification decision. The confidence prediction head mechanism is particularly useful for SAR images where noise and lower resolution can lead to uncertain predictions. The confidence head provides uncertainty estimates that can be used to control the predictions of the model and improve its robustness. By explicitly modeling uncertainty, this system can identify when it is likely to make errors, especially in SAR images, when speckle noise is severe and features are ambiguous. The multi-component loss function consists of three main terms:

Classification Loss: To address severe imbalances in the dataset, they apply class weights with Cross-Entropy loss to ensure that minority classes receive adequate attention during training.

Feature Matching Loss (MMD): By aligning the feature representations between the SAR and EO domains, they support the SAR model in training more discriminative features by mimicking the EO feature distribution. The authors adopt the Generative Feature Matching Network (GFMN) [32] method, which compares the statistical moments of the feature distribution directly:

$$\mathcal{L}_{\text{MMD}} = \sum_{j=1}^{M} \|\mu_j^{\text{SAR}} - \mu_j^{\text{EO}}\|^2 + \|\sigma_j^{\text{SAR}} - \sigma_j^{\text{EO}}\|^2, \quad (2)$$

where μ_j and σ_j represent the mean and variance of features extracted from layer j of their feature extractor network ϕ. This approach effectively transfers the rich structural information from EO to SAR features by matching their statistical moments directly.

Contrastive Loss (SupCon): The third term is a supervised contrastive loss to improve feature discrimination. This learning draws features from the same class close to each other while separating features from different classes [16]:

$$\mathcal{L}_{\text{SupCon}} = \sum_{i=1}^{N} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p/\tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a/\tau)}. \quad (3)$$

Here, z_i is the normalized feature of the i-th sample, $P(i)$ is the set of indices of samples in the same class as i, $A(i)$ is the set of all indices except i, and τ is the temperature parameter. This is particularly important for SAR images where intra-class variation is high and inter-class similarity can be deceptive. Their specialized contrastive approach uses hard negative mining with class balancing to focus on the most challenging examples. The approach implements varying degrees of attraction between samples: strong attraction between augmented views of the same image, moderate attraction between different images of the same class, and repulsion between samples from different classes, creating a more nuanced embedding space that better handles the complexities of SAR imagery and ID vs OOD data.

Additionally, they incorporate a confidence prediction mechanism [3] to handle OOD samples. The confidence prediction is directly integrated with the classification process, where the model's confidence estimates modify the classification outputs. For each prediction, the confidence score determines how much the model relies on its prediction versus the ground truth label during training:

$$\hat{y} = c \cdot \text{softmax}(z) + (1 - c) \cdot y_{\text{true}}. \quad (4)$$

Where c is the predicted confidence, z is the logit output, and y_{true} is the ground truth label. The Eq. (4) formulation enables the model to adaptively balance between its predictions and the ground truth based on its confidence level, preventing overconfidence in noisy regions of SAR images. The confidence value can be interpreted as an uncertainty score (where lower confidence indicates higher uncertainty), providing valuable information about prediction reliability. This uncertainty quantification is particularly beneficial for SAR imagery analysis, where image quality variations can significantly impact classification reliability.

Overall, their training strategy includes EO pre-training, adaptive confidence adjustment, class weighting to address imbalance, and multi-GPU distributed training for efficiency. They employ a two-stage training process: first, pre-training the EO model on the EO images, then freezing the EO encoder and training the SAR model with knowledge distillation from the EO model.

Implementation Details: The authors used ResNet101 [6] for both teacher and student networks. Training was performed with a batch size of 128 using AdamW optimizer on NVIDIA A5000 GPUs with 24GB memory. The model employs class weighting to address imbalance and basic data

augmentation techniques, including random cropping, flipping, and rotation. The source code for our implementation is publicly available at: https://github.com/PBVS2025/PBVS2025_kd.

4.3. Rank 12: kjhfkjn

Team Members: Xiaojie Liu

4.3.1. Proposed Approach

Dataset Curation: The team addresses challenges associated with the long-tail distribution problem and image similarity issues in the training dataset. To mitigate these problems, they first use similarity measures to remove duplicate images. Then, data augmentation techniques are applied to increase the number of images in smaller categories, ensuring a balanced dataset with 6000 images per class.

Methodology: Inspired by securing second place in last year's PBVS 2024 challenge [26], the team employs two independent ViT-Base models to handle SAR and EO inputs, respectively. The final output is obtained by combining the predictions from both models using an ensemble approach. See Figure 5 for the architecture overview. Due to the lack of EO data in the test set, the team uses the pix2pixHD [36] model to generate EO data. A pretrained model from the SAR translation competition (MAVIC-T [27]) is fine-tuned on the training set and used to convert SAR images in the test set into EO images. The codes are available: https://github.com/kjhfkjn/PBVS2025SARClassification.

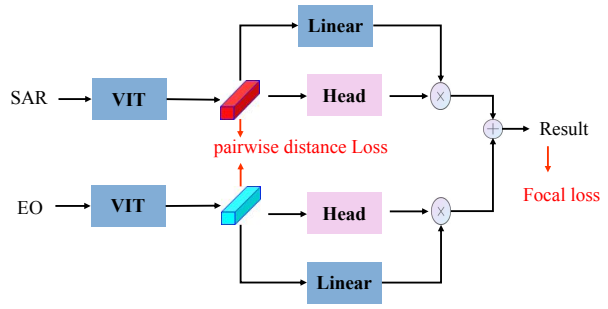

Figure 5. Overview of the kjhfkjn-architecture.

Loss Functions: Two loss functions are employed in this approach. The first is a Focal Loss [21], which helps address the long-tail distribution problem by focusing on hard-to-classify examples. The second is a Pairwise Distance Loss, which synchronizes the feature distributions between SAR and EO domains. The pairwise distance loss can lead to sparse features in the model, which may hinder classification performance. Therefore, a coefficient of 0.1 is applied to mitigate this issue.

4.3.2. Hardware and Software

The team uses the following hardware and software tools in this competition:

- **Programming Language**: Python 3.11
- **Framework**: PyTorch 1.10.0+cu113 Ubuntu22.04
- **Hardware**: RTX 4090, 24GB GPU, 32GB RAM
- **Training Time**: 5 hours

5. Conclusion

This year's MAVIC-C challenge was a success! We had 105 teams participate in the competition and the winning teams demonstrated very impressive results, in both classification and OOD detection metrics. Interestingly, the methodologies employed by the top performers leverage new and exciting components from several areas across the deep learning community. We saw teams using combinations of contrastive learning, feature matching losses, uncertainty quantification techniques, specialized loss weightings to account for data imbalance, vision transformer backbones, student-teacher knowledge distillation, and model ensembling. Compared to last year, this year we saw a notable improvement in the OOD detection ability of methods, leading us towards more robust and reliable OR models for remote sensing in open-world environments.

Lastly, we give some comments on exciting new directions that future contestants and practitioners may consider to advance the performance further. Foundation models are a promising direction in many aspects [10, 18, 31]; they improve transfer learning, low-shot learning, and their representation quality can assist in OOD detection and generalization. Several recent advances in OOD detection, including improved architectures may be considered [4, 15, 37]. Teams may consider variants of online learning to adapt their offline-pretrained models to the test data distribution on the fly, potentially with or without human assistance [5, 19, 23, 35].

References

[1] Zhaowei Cai, Avinash Ravichandran, Subhransu Maji, Charless Fowlkes, Zhuowen Tu, and Stefano Soatto. Exponential moving average normalization for self-supervised and semi-supervised learning. In *CVPR*, 2021. 5

[2] S. Chen, H. Wang, F. Xu, and Y. Jin. Target classification using the deep convolutional networks for sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 54 (8):4806–4817, 2016. 1

[3] Terrance DeVries and Graham W Taylor. Learning confidence for out-of-distribution detection in neural networks. *arXiv preprint arXiv:1802.04865*, 2018. 6

[4] Xuefeng Du, Yiyou Sun, Xiaojin Zhu, and Yixuan Li. Dream the impossible: Outlier imagination with diffusion models. In *Advances in Neural Information Processing Systems*, 2023. 7

[5] Elvin Hajizada, Balachandran Swaminathan, and Yulia Sandamirskaya. Continual learning for autonomous robots: A prototype-based approach. *arXiv preprint arXiv:2404.00418*, 2024. 7

[6] 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*, pages 770–778, 2016. 6

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 4

[8] Matthias Hein, Maksym Andriushchenko, and Julian Bitterwolf. Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[9] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *Proceedings of International Conference on Learning Representations*, 2017. 2

[10] Nathan Inkawhich. A global model approach to robust few-shot sar automatic target recognition. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. 7

[11] Nathan Inkawhich, Eric Davis, Uttam Majumder, Chris Capraro, and Yiran Chen. Advanced techniques for robust sar atr: Mitigating noise and phase errors. In *IEEE International Radar Conference (RADAR)*, 2020. 1

[12] Nathan Inkawhich, Eric Davis, Matthew Inkawhich, Uttam K. Majumder, and Yiran Chen. Training sar-atr models for reliable operation in open-world environments. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3954–3966, 2021. 2

[13] Nathan Inkawhich, Jingyang Zhang, Eric K. Davis, Ryan Luley, and Yiran Chen. Improving out-of-distribution detection by learning from the deployment environment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:2070–2086, 2022. 2

[14] Eric Keydel, Shung Lee, and John Moore. Mstar extended operating conditions: a tutorial. In *SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery III*, 1996. 1

[15] Amol Khanna, Chenyi Ling, Derek Everett, Edward Raff, and Nathan Inkawhich. Multi-layer radial basis function networks for out-of-distribution detection. *ArXiv*, abs/2501.02616, 2025. 7

[16] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33:18661–18673, 2020. 6

[17] Colin Leong, Todd Rovito, Olga Mendoza-Schrock, Christopher Menart, Jason Bowser, Linda Moore, Steve Scarborough, Michael Minardi, and David Hascher. Unified coincident optical and radar for recognition (unicorn) 2008 dataset, 2008. 2

[18] Weijie Li, Wei Yang, Yuenan Hou, Li Liu, Yongxiang Liu, and Xiang Li. Saratr-x: Toward building a foundation model for sar target recognition. *IEEE Transactions on Image Processing*, 34:869–884, 2025. 7

[19] Jian Liang, Ran He, and Tieniu Tan. A comprehensive survey on test-time adaptation under distribution shifts. *International Journal Of Computer Vision*, 2023. 7

[20] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *CoRR*, abs/1708.02002, 2017. 5

[21] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 7

[22] Jerrick Liu, Nathan Inkawhich, Oliver Nina, Radu Timofte, Sahil Jain, Bob Lee, Yuru Duan, Wei Wei, Lei Zhang, Songzheng Xu, Yuxuan Sun, Jiaqi Tang, Xueli Geng, Mengru Ma, Gongzhe Li, Huanqia Cai, Chengxue Cai, Sol Cummings, Casian Miron, Alexandru Pasarica, Cheng-Yen Yang, Hung-Min Hsu, Jiarui Cai, Jie Mei, Chia-Ying Yeh, Jenq-Neng Hwang, Michael Xin, Zhongkai Shangguan, Zihe Zheng, Xu Yifei, Lehan Yang, Kele Xu, and Min Feng. NTIRE 2021 multi-modal aerial view object classification challenge. *CoRR*, abs/2107.01189, 2021. 2

[23] Lisa Loomis, David Wise, Nathan Inkawhich, Clare Thiem, and Nathan McDonald. Task-agnostic feature extractors for online learning at the edge. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications VI*, 2024. 7

[24] Spencer Low, Oliver Nina, Angel D. Sappa, Erik Blasch, and Nathan Inkawhich. Multi-modal aerial view object classification challenge results-pbvs 2022. In *Proceedings of the IEEE conference on computer vision and pattern recognition - workshop*, 2022. 2

[25] Spencer Low, Oliver Nina, Angel D Sappa, Erik Blasch, and Nathan Inkawhich. Multi-modal aerial view object classification challenge results-pbvs 2023. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 412–421, 2023. 5

[26] Spencer Low, Oliver Nina, Dylan Bowald, Angel D. Sappa, Nathan Inkawhich, and Peter Bruns. Multi-modal aerial view image challenge: Sar classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3105–3112, 2024. 2, 3, 7

[27] Spencer Low, Oliver Nina, Dylan Bowald, Angel D. Sappa, Nathan Inkawhich, and Peter Bruns. Multi-modal aerial view image challenge: Sensor domain translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3096–3104, 2024. 7

[28] Uttam K. Majumder, Erik P. Blasch, and David A. Garren. *Deep Learning for Radar and Communications Automatic Target Recognition*. Artech House, 2020. 1

[29] Kannappan Palaniappan, Mahdieh Poostchi, Hadi Aliakbarpour, and et al. Moving object detection for vehicle tracking in wide area motion imagery using 4d filtering. In *International Conference on Pattern Recognition (ICPR)*, 2016. 3

[30] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing.

In *Scale Space and Variational Methods in Computer Vision*, 2012. 5

[31] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. *CoRR*, abs/2103.00020, 2021. 7

[32] Cicero Nogueira dos Santos, Youssef Mroueh, Inkit Padhi, and Pierre Dognin. Learning implicit generative models by matching perceptual features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4461–4470, 2019. 6

[33] Merrill Skolnik. *Radar Handbook, 3rd Edition*. McGraw-Hill, 2008. 1

[34] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning (ICML)*, 2019. 4

[35] Claire Thorp, Sean Sisti, Lesrene Browne, Casey Schwartz, Nathan Inkawhich, and Walter Bennette. Efficient fine-grained automatic target recognition through active learning for defense applications. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications VI*. SPIE, 2024. 7

[36] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 7

[37] Jingyang Zhang, Jingkang Yang, Pengyun Wang, Haoqi Wang, Yueqian Lin, Haoran Zhang, Yiyou Sun, Xuefeng Du, Yixuan Li, Ziwei Liu, Yiran Chen, and Hai Li. Openood v1.5: Enhanced benchmark for out-of-distribution detection. *arXiv preprint arXiv:2306.09301*, 2023. 7