Multi-modal Aerial View Image Challenge: SAR Classification

Spencer Low
Brigham Young University
Provo, Utah
spencerlow@byu.edu

Oliver Nina
Air Force Research Laboratory
Dayton, OH
oliver.nina.1@afresearchlab.com

Dylan Bowald
Air Force Research Laboratory
Dayton, OH
dylan.bowald.1@us.af.mil

Angel D. Sappa
ESPOL Polytechnic University, Ecuador
Computer Vision Center, Spain
sappa@ieee.org

Nathan Inkawhich
Air Force Research Laboratory
Rome, NY
nathan.inkawhich@us.af.mil

Peter Bruns
University of Utah
Salt Lake, UT
brunsp10@gmail.com

Abstract

This manuscript delineates the outcomes of the fourth Multi-modal Aerial View Image Challenge - Classification (MAVIC-C). The challenge is aimed at advancing the development of recognition models that leverage Synthetic Aperture Radar (SAR) and Electro-Optical (EO) imagery. Encouraging the integration of data from these two distinct modalities, the challenge seeks to foster the creation of multi-modal approaches that complement characteristics of SAR and EO information. Building upon the precedents set in previous years, the 2021 MAVOC challenge validated the potential of integrating SAR and EO modalities. The subsequent 2022 and 2023 challenges further explored the capabilities of multi-modal frameworks. In its latest iteration, the 2024 challenge presents an enhanced UNIfied COincident Optical and Radar for recognitioN (UNICORN) dataset alongside a revised competition format, focused on the task of SAR classification. The 2024 challenge evaluates model robustness through out-of-distribution measures, alongside traditional accuracy metrics. The core of this paper is devoted to analyzing the methodologies of the top-performing entries and their performance metrics on a blind test set.

1. Introduction

The primary goal of Object Recognition (OR) models is to effectively detect, identify, and classify target signatures within remotely sensed imagery, leveraging various sensing technologies as evidenced in literature [1, 3, 5, 9, 10, 14]. OR, akin to object detection and labeling in conventional imagery, uniquely operates on complex remote sensing (RS) platforms deployed on aerial vehicles. Or from aerial perspective introduces specific challenges for target recognition, including constraints such as limited sensor resolution, which may reduce targets to mere pixels, complicating their identification [4].

Synthetic Aperture Radar (SAR) has been explored as a standalone modality, offering advantages of all-weather, all-time surveillance capabilities. Nevertheless, it also faces challenges, including issues with signal distortion due to multi-bounce effects, shadows, and the delineation of adjacent objects [15]. The integration of Electro-Optical (EO) imagery with SAR data for OR purposes introduces a series of distinct research challenges.

A critical requirement for OR systems, distinguishing them from conventional object detection models, is their ability to handle out-of-distribution samples accurately [3]. Given the potential consequences of high-confidence misclassifications in OR applications, models must assign a credibility score to each detection, where lower scores signal potential uncertainty. The 2024 PBVS Multi-modal Aerial View Image Challenge (MAVIC) challenge offers a fertile ground for addressing these intricate challenges and exploring how multi-modal data can enhance OR models.

Remote sensing (RS) systems benefit significantly from multi-sensor fusion [13]. While certain sensors, like self-illuminating or microwave systems, provide all-weather capabilities and penetration through obstructions, passive sensors can measure environmental variables such as temperature. Despite these benefits, the integration of diverse sensor data into a cohesive OR system remains under explored in computer vision, largely due to the complexities associated with merging disparate data sources. Traditional RS approaches often rely on singular modalities, primarily, visual spectrum data. While multi-spectral (MSI) and hyperspectral (HSI) imaging extend the available spectral bands, they also introduce challenges related to band selection and increased computational demands. Yet, the strategic fusion of various sensor data holds the promise of significant enhancements in OR system performance [12].

In this paper, we outline the 2024 MAVIC-C which introduces a further refined version of the UNICORN Dataset. This dataset more rigorously test the generalizability of object recognition models, by splitting the train, test and validation data across different geographic regions. We observe strong performance, despite this change. Section 2 describes the challenge with Section 3 listing the top performing methods. Section 4 details three top methods and Section 5 provides the conclusions.

2. Challenge

Building on the foundation laid by the 2023 Multi-modal Aerial View Object Classification (MAVOC) challenge [8], the 2024 MAVIC-C (Multi-modal Aerial View Image Classification Challenge) progresses the initiative in conjunction with the 20th Perception Beyond the Visible Spectrum (PBVS) workshop. The MAVIC-C challenge aims to drive advancements in multi-modal image classification, focusing on the synergistic use of Synthetic Aperture Radar (SAR) and Electro-Optical (EO) image pairs. Participants are assessed based on top-1% accuracy and the Area Under the Receiver Operating Characteristic (AUC) curve, promoting the development of models proficient in precise labeling and the identification of out-of-distribution samples.

The provision of SAR images introduces distinct challenges for competitors, attributed to their inherent self-illumination and coherent properties. These characteristics produce images marked by notable *SAR shadows* and a unique tilted perspective, testing the adaptability and innovation of participating teams. The 2024 challenge is focused on SAR image classification.

2.1. SAR Classification

The MAVIC-C is designed to facilitate development of classifiers capable of being trained using both Synthetic Aperture Radar (SAR) and Electro-Optical (EO) data but operational exclusively on SAR data during testing. The essence of this track is to leverage the integrated features of SAR and EO imagery within the training phase to enable the classifier to function without EO data at deployment. This approach aims to expedite decision-making processes by eliminating the need for the computationally intensive rectification preprocessing required to align SAR with EO data. The heterogeneous nature of the training dataset, encompassing different modalities, poses a significant challenge to achieving this objective.

2.2. Area Under the ROC

The AUC metric serves as a pivotal evaluation criterion in the 2024 MAVIC-C challenge, assessing a model's proficiency in identifying out-of-distribution samples. With a scale ranging from 0 to 1, where 0.5 equates to random chance and 1 denotes flawless performance, the AUC provides a nuanced measure of model credibility. The challenge intricately integrates out-of-distribution or negative samples within the validation and test sets—absent during the training phase—to discourage reliance on generic catch-all classes for these samples.

2.3. Dataset

The MAVIC-C challenge is based by the UNIfied CO-incident Optical and Radar for recognitioN (UNICORN) dataset, featuring a curated SAR-EO dataset that is publicly accessible and aligned, with hand-annotated classes. The 2024 iteration employs the UNICORN V2 dataset, an advanced version of the 2008 dataset, comprising Wide Area Motion Imagery (WAMI) from large-format EO sensors and Wide Area Synthetic Aperture Radar (SAR) data, sourced from aerial surveys over Dayton, Ohio [11]. Initial evaluations with a baseline pre-trained Resnet-50 model indicate comparable performance across both UNICORN dataset versions, with accuracy rates ranging from 15-18% on a reserved test set. Additionally, methodologies and models from the 2022 challenge exhibited consistent performance on the updated dataset.

Despite covering analogous fields of view, the SAR and EO data differ in resolution, with SAR images presenting finer detail but with greater introduced noise. The images are meticulously rectified and aligned using advanced homography algorithms, as illustrated in Fig. 1. The competition dataset comprises smaller segments (chips) extracted from these aligned larger images. Each chip showcases one of 10 object categories, featuring EO chips of 31×31 pixels and SAR chips of 55×55 pixels, reflecting the resolution disparity. Enhanced homography and alignment processes underpin the creation of the UNICORN V2 dataset, facilitating the identification of more chips and achieving more precise alignments. The training set exhibits a long-tail distribution among the classes, while the validation and test sets are uniformly distributed across all categories, ensuring an equitable assessment of model accuracy.

A significant methodological advancement for the 2024 challenge is the adoption of a more sophisticated approach

Figure 1. The aligned scene of the full UNICORN dataset before chipping is performed[7].

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

to splitting the dataset into training, validation, and testing sets. Instead of the conventional random sampling method, the split is based on geographical regions, resulting in distinct test and validation sets that pose a heightened challenge for participants. This innovative splitting strategy demands that models exhibit enhanced generalizability across a limited geographical dataset, a departure from previous competitions that aims to better assess the robustness of the models to novel environments. Table 1 outlines the distributions and split of the dataset.

The out-of-distribution images are pulled from other classes in the UNICORN V2 dataset. These are classes with fewer examples than the semi truck w/ trailer class. These classes are shown in Table 2.

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

2.4. Evaluation

Submissions are evaluated using a weighted average of the top-1% accuracy and AUC of the model. The test set contains 2,000 unlabelled (SAR, EO) chip pairs, with 200 examples for each of the 10 classes, and 2000 out-of-distribution samples. The weighting is shown in Eq. 1.

$$\text{Score} = 0.75 \text{ Accuracy} + 0.25 \text{ AUC} \tag{1}$$

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 teams from effectively fine-tuning on the test dataset. Results are made visible during both phases.

2.5. Challenge Phases

The challenge began January 11, 2024, and the test data was released February 22, 2024. The testing phase ended on March 5, 2024 with team submissions finalized.

3. Challenge Results

The 2024 MAVIC-C challenge witnessed substantial engagement, reflecting the growing interest and advancements in multi-modal aerial view image classification challenge problems.

A total of 146 teams registered for participation. During the development phase, 96 of these teams proceeded to submit their algorithms for preliminary evaluation. The participation slightly adjusted in the testing phase, with 50 teams submitting valid algorithms for rigorous assessment.

4. Challenge Methods

The top three performing methods are outlined.

Table 3. Top-10 Teams for MAVIC-C.

Rank	Team	Total Score	Accuracy	AUC
1	**IQSKJSP**	**0.49**	**37.9**	**0.83**
2	MITHF	0.46	38.85	0.69
3	GuanYu	0.39	35.10	0.49
4	GWLoong	0.35	38.80	0.24
5	http	0.33	10.40	1.00
6	unknown	0.33	10.05	1.00
7	findanswear	0.32	10.00	1.00
8	NJUST-KMG	0.31	8.45	1.00
9	yuhyun	0.30	21.45	0.56
10	bingxiang	0.30	18.90	0.62
	2023 Best (fcai)	0.65	63.2	0.71

4.1. Rank 1: IQSKJSP

Large foundation models pre-trained on web-scale datasets have revolutionized the field of computer vision, showing powerful zero-shot and few-shot generalizations. These models have the potential to generalize tasks and data distributions beyond those seen during training. They have been successfully applied in various fields, including visual recognition, dense prediction, Reinforcement Learning (RL), and Robotics.

Recent results from BALLAD, RAC, VL-LTR, and LPT demonstrate that properly fine-tuning pre-trained foundation models can improve long-tail object recognition accuracy. For instance, LPT fine-tunes the ViT pre-trained on ImageNet, utilizing prompt tuning via two-stage training.

Due to the large differences between general images and SAR images and the absence of labeled SAR image datasets, there is no fine-tuned vision foundation model available for fine-grained object recognition in SAR images. And existing fine-tuned models are designed from the perspective of the model structure, ignoring the unique scattering information of SAR images.

Building upon these considerations, team IQSKJSP proposes a fine-tuned foundation model called Scattering Prompt Tuning (SPT) specifically tailored for object recognition in SAR images. As shown in Fig. 2, the scattering information extracted from SAR images is converted into a textual description, serving as the prompt alongside the visual image description. These inputs are processed by the text encoder to extract semantic features, initializing the model head for better convergence. To guide the image encoder in learning scattering information, they introduce trainable parameters as a scattering characteristics prompt in the input space, facilitating the fine-tuning of the ViT model.

Acknowledging the disparities between web images and SAR images, they introduce a lightweight module named Residual AdapterMLP (RAMLP) within each transformer block, comprising a fully connected layer, nonlinear activation function, and residual feature fusion factors. The pre-trained transformer is fine-tuned by updating RAMLP. Additionally, they implement a sequential feature aggregation module to selectively fuse feature outputs from different transformer blocks. This module adaptively extracts rich hierarchical information from SAR images. To address the challenge of extreme inter-class similarity and intra-class differences, they develop the Dynamic Distributional Contrast Loss. This loss function ensures that features of objects from the same class and different classes maintain appropriate distances, enhancing class distinguishability while preserving intra-class differences to some extent.

IQSKJSP's fine-tuned foundation model, SPT, diverges from existing methods by effectively leveraging the scattering properties of SAR images. SPT mitigates domain shift issues between general web images and SAR images, proving to be effective and well-suited for fine-grained object recognition in SAR images.

In summary, their contributions are outlined as follows:

1. They propose a Scattering Prompt Tuning (SPT) based foundation model.
2. Different from existing methods, their method uniquely leverages SAR image scattering properties to markedly improve object recognition performance.

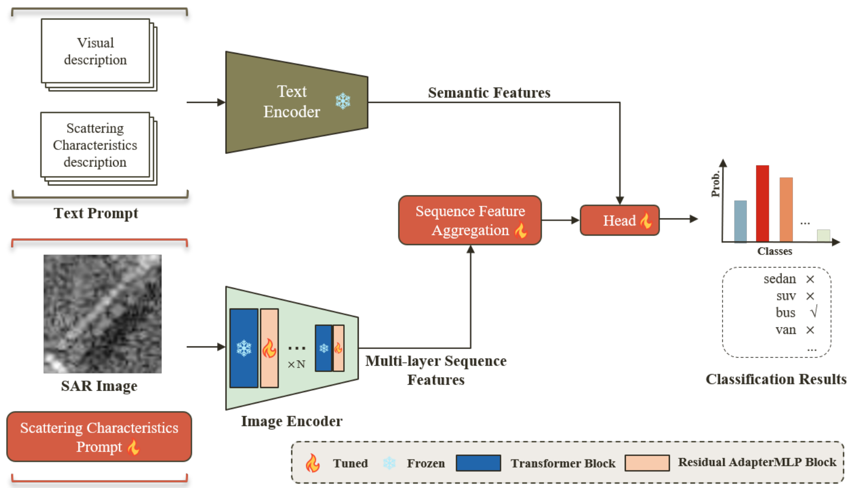

Figure 2. The framework of SPT.

3. They combine and design novel modules like Residual AdapterMLP and Sequential Feature Aggregation with a Dynamic Distributional Contrast Loss, significantly enhancing semantic feature learning in SAR images.

4.1.1 Scattering information extraction

The scattering information of SAR images is the phenomenon of radar wave interaction with ground objects [6]. Different objects have different scattering characteristics. The scattering information describes the characteristics of the object surfaces, which is important for understanding their edges and structures in SAR images. Embedding the scattering information into the prompt encoder may help to improve the object recognition performance of the model. They consider the scattering points extracted from the SAR image as its scattering information.

4.1.2 Scattering Characteristics Prompt

For a plain vision transformer (ViT) with N layers, an input image is divided into m fixed-sized patches. Each patch is then first embedded into d-dimensional latent space with positional encoding. Together with an extra learnable classification token ([CLS]), the patch is fed into the transformer for feature learning. In their SPT, as shown in Fig. 3, they add extra learnable vectors as the scattering characteristics prompt (SCP) in the input space to help the pre-trained model learn the scattering information of unseen SAR images. During the fine-tuning process, the parameters of scattering characteristics prompt is updated.

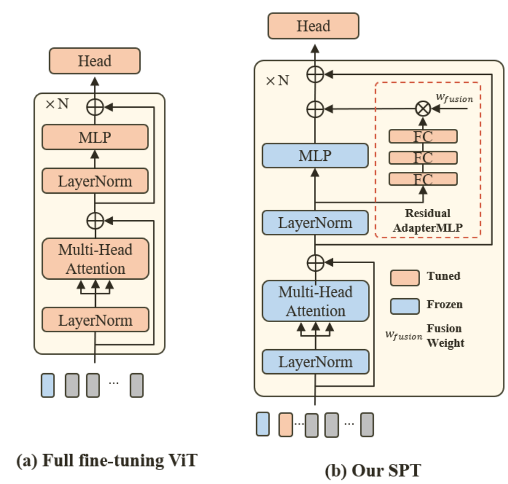

Figure 3. Comparison of full fine-tuning ViT and SPT. (a) Structure of full fine-tuning ViT. (b) Structure of SPT.

4.1.3 Residual AdapterMLP

ViTs are usually trained on large web-scale datasets and may lack exposure to SAR image information. To bridge this gap, as shown in Fig. 3, they introduce the Residual AdapterMLP (RAMLP) module in each transformer block, helping the model effectively capture accurate details from SAR images. With fully-connected layers, nonlinear activation functions, and residual feature fusion weights, RAMLP dynamically adjusts the ViT weights during fine-tuning.

Recent studies, like AdapterFormer, highlight the importance of the MLP in fine-tuned ViTs for general image/video recognition. RAMLP not only prevents certain issues such as output degradation in ViTs but also improves its performance. Their version of the RAMLP module in the fine-tuned ViT differs from AdapterFormer by including more nonlinear layers and deeper embedding dimensions for better learning

4.1.4 Sequential Feature Aggregation

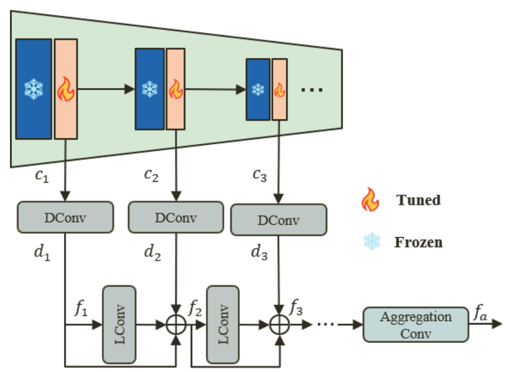

Figure 4. Details of sequential feature aggregation module.

The pre-trained ViT model produces diverse features in its different blocks, each containing distinct levels of semantic information from the SAR image. Considering these features as a sequential sequence, they design a sequential feature aggregation (SFA) module to selectively filter and merge the most relevant information for SAR object recognition. The details are shown in Fig. 4.

For the output features $\{c_1, \cdots, c_N\}$ from the N transformer blocks, IQSKJSP initially apply a channel downsampling convolution (DConv) to get a uniform channel count for the feature sequence $\{d_1, \cdots, d_N\}$. Subsequently, short and long-term sequence feature screening is performed to derive features f_N, encompassing information from various positions within the sequence. Lastly, an aggregated convolution, formed by stacking multiple 1×1 convolutional layers, adapts to fuse the screened features effectively.

$$f_N = LConv(f_{N-1}) + f_{N-1} \tag{2}$$

4.1.5 Loss Function

The loss function $\mathcal{L}$ in IQSKJSP's SPT model comprises two key components: L_A (Logit-Adjusted loss) and L_{DC} (Designed Dynamic Distributional Contrast Loss, DCLoss). L_A primarily addresses the challenges posed by the long-tailed distribution of data during training. On the other hand, L_{DC} is tailored to tackle the difficulties arising from the extreme inter-class similarity and intra-class differences in SAR image object recognition. The loss is shown in Eq. 3.

$$\mathcal{L} = L_A + L_{DC} \tag{3}$$

The primary objective of designing L_{DC} is to ensure that features from different object classes are widely separated, while features from the same object class are brought closer together, aligning with the actual data distribution. To address the local clustering issue within classes (intra-class differences), they incorporate a dynamic distance threshold screening strategy. This strategy allows for variations in distances between objects of the same class while ensuring they are kept distinct from features of other classes. For the features extracted from n SAR image objects, they measure their similarity using the Euclidean distance:

$$dist_{ij} = \|\|o_i - o_j\|\|_2^2 = \sum_{k=1}^{C} (o_{ik} - o_{jk})^2 \tag{4}$$

where i and j denote the index of the sar image, $\{1 \leq i, j \leq n\}$ and $\{i \neq j\}$. C is the channel number. To guide the learning of the similarity measure matrix $dist$, we build its ground truth, G_{sim}, based on the actual labels of each object. The value is set to 1 when the labels of two objects are the same and 0 otherwise. The calculation of L_{DC} can be expressed using the following equation:

$$L_{DC} = G_{sim} * (dist - \delta_p)^2 + (1 - G_{sim}) * max(\delta_q - dist, 0)^2 \tag{5}$$

δ_p and δ_q represent the dynamic thresholds for distances between objects of the same class and different classes, adapting to various data distributions. In IQSKJSP's specific implementation, they use the average values of the distances between objects of the same class and different classes as δ_p and δ_q, respectively.

4.1.6 Training Strategy

To gradually adapt the pre-trained ViT model to SAR object recognition, IQSKJSP's SPT model undergoes four stages of training.

In the first stage, they fine-tune the SPT initially on a mixed dataset of EO and SAR images, followed by another fine-tuning specifically on SAR images. This process yields an initial foundation model tailored to the SAR image domain.

Moving to the second stage, they create a balanced dataset by sampling images from various classes in the SAR image training dataset for further fine-tuning.

In the third stage, their focus is on understanding the differences between the training data and the validation(development phase) / test(test phase) data. They capture features, prediction scores, and prediction categories from the model's last layer for all validation/test data. By applying a threshold, they select high-scoring samples and employ Gaussian Mixture and K-means++ algorithms for feature clustering. The overlap of these results serves as the final clustering outcome. If a high-scoring sample is within a cluster, IQSKJSP assign the prediction categories of the high-scoring sample to all cluster members. This process generates reliable predictions, and they construct a balanced dataset by incorporating some training images for further fine-tuning.

In the fourth stage, they exclusively use high-scoring samples as reliable predictions to build a balanced dataset for additional fine-tuning. Since they find that the strategy in the third stage becomes less effective as the model's prediction accuracy improves. Consequently, they iterate through the fourth stage until the model's prediction accuracy reaches a plateau.

4.2. Rank 2: MITHF

The 2023 PBVS SAR Classification results [8] showed that the accuracy of classifying electro-optical (EO) images is much higher than that of SAR images under similar conditions. This led MITHF to explore ways to transfer the EO image network's capabilities to SAR images. In the 2024 PBVS SAR classification challenge, training data includes pairs of SAR and EO images, but only SAR images are used in the test phase. Based on this, they developed an method that uses EO image generation and integration. The framework of MITHF's approach is illustrated in the following Fig. 5:

The pipeline is as follow: (1) First, MITHF train their model so it can convert SAR images into EO images. They start with a ViT that has been pre-trained to act as their image encoder. This encoder takes the features from a SAR image and works with an image decoder to rebuild these features into an EO image, matching the original SAR image's scale. During training, they use the actual EO images provided by the organizers as a supervision to guide the model. Once this stage is complete, they end up with corresponding pairs of SAR and EO images for use as test data.

Then, they train separate models for both SAR and EO images using AdapterFormer [2]. This training approach is inspired by the top SAR and EO solutions of PBVS2023 challenge, where they first train on SAR and EO classifications separately. They utilize various networks, including ResNet101, VGG, and MobileNet, for both SAR and EO classifications, aiming for a combined approach through multi-model integration.

During testing, they discovered that the SAR and EO models each excelled at identifying different categories. Therefore, they combined the strengths of both models to achieve their final results.

4.3. Rank 3: GuanYu

In this challenge, GuanYu employs a two stage model. In Stage 1 they trained multiple 10-class classifiers and use them to vote for each sample. They then take the most-voted class as the final label and the average confidence of these models as the final confidence score. They observe that some classes such as: class 0: sedan, 1: suv, 2: pickup truck, and 3: van, are visually similar. They exploit this and train a two-stage classification module to optimize their performance.

In Stage 2, they added a classification model and a clustering module to re-identify the samples that are easily confused (class $0-3$). Specifically, they used the Perceptual hash algorithm to cluster the samples that were classified as class $0-3$, and then classified the samples within each cluster using the 4-class classifier. The most frequently occurring category was taken as the label for all samples in the cluster.

To fuse SAR and EO, they first extract features using a backbone model and then test transformer and traditional concatenation feature fusion strategies, as well as a concatenation module based on channel self-attention. Ultimately, they used a concatenation feature fusion strategy based on channel attention. They used the output results of multiple models to vote for the final result. The confidence of the sample is calculated from the average of multiple voting models. In the experiment, they found that the distribution of class $0-3$ categories is similar, so they trained a two-stage module to further optimize.

5. Conclusion

The 2024 MAVIC-C demonstrated innovated improvements to the multi-modal classification problem. Performance remained high, despite the more rigorous test split that more thoroughly test a model's generalizability. This year challenge has garnered significant attention, evidenced by the registration of 146 teams eager to participate, showing a growing interest in the multi-modal aerial view image classification topic. Results from this year will be a benchmark for next years' competition.

6. Acknowledgements

Acknowledgements

Funding for the project was provided through AFRL. We would like to acknowledge Peter Bruns, and Justice Wheelwright for their help in running the challenge. Acknowledgements The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Author Universities, Industries, or the U.S. Government.

References

[1] S. Chen, H. Wang, F. Xu, and Y. Jin. Target classification using the deep convolutional networks for sar images.

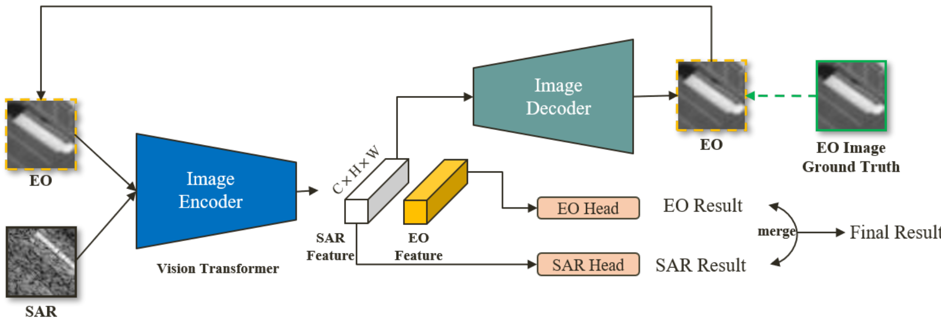

Figure 5. The pipeline of MITHF's method.

IEEE Transactions on Geoscience and Remote Sensing, 54 (8):4806–4817, 2016. 1

[2] Shoufa Chen, Chongjian GE, Zhan Tong, Jiangliu Wang, Yibing Song, Jue Wang, and Ping Luo. Adaptformer: Adapting vision transformers for scalable visual recognition. In *Advances in Neural Information Processing Systems*, pages 16664–16678. Curran Associates, Inc., 2022. 7

[3] 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

[4] 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. 1

[5] Nathan Inkawhich, Matthew Inkawhich, Eric Davis, Uttam Majumder, Erin Tripp, Chris Capraro, and Yiran Chen. Bridging a gap in sar-atr: Training on fully synthetic and testing on measured data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 2942–2955, 2021. 1

[6] Julie Jackson, Brian Rigling, and Randolph Moses. Canonical scattering feature models for 3d and bistatic sar. *IEEE Transactions on Aerospace and Electronic Systems*, 46:525–541, 2010. 5

[7] 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. 3

[8] 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*, pages 412–421, 2023. 2, 7

[9] Uttam Majumder, Erik Christiansen, Qing Wu, Nate Inkawhich, Erik Blasch, and John Nehrbass. High-performance computing for automatic target recognition in synthetic aperture radar imagery. Proc. SPIE 10185, 2017. 1

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

[11] 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. 2

[12] James Patrick, Ryan Brant, and Erik Blasch. Hyperspectral imagery throughput and fusion evaluation over compression and interpolation. In *2008 11th International Conference on Information Fusion*, pages 1–8, 2008. 2

[13] Yuanyuan Qing, Jiang Zhu, Feng Hongchuan, Weixian Liu, and Bihan Wen. Two-way generation of high-resolution eo and sar images via dual distortion-adaptive gans. *Remote Sensing*, 15, 2023. 1

[14] Timothy Ross, Stephen Worrell, Vincent Velten, John Mossing, and Michael Bryant. Standard sar atr evaluation experiments using the mstar public release data set. In *SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery V*, 1998. 1

[15] Timothy D. Ross, Jeff J. Bradley, Lannie J. Hudson, and Michael P. O'Connor. SAR ATR: so what's the problem? An MSTAR perspective. In *Algorithms for Synthetic Aperture Radar Imagery VI*, pages 662–672, 1999. 1