S-Band SAR Target Classification via 2D and 3D Deep Learning Methods

Tyler Rust, Michael Pergeorelis, Chandra Kambhamettu
University of Delaware
18 Amstel Ave.
{trust, pergem, chandrak}@udel.edu

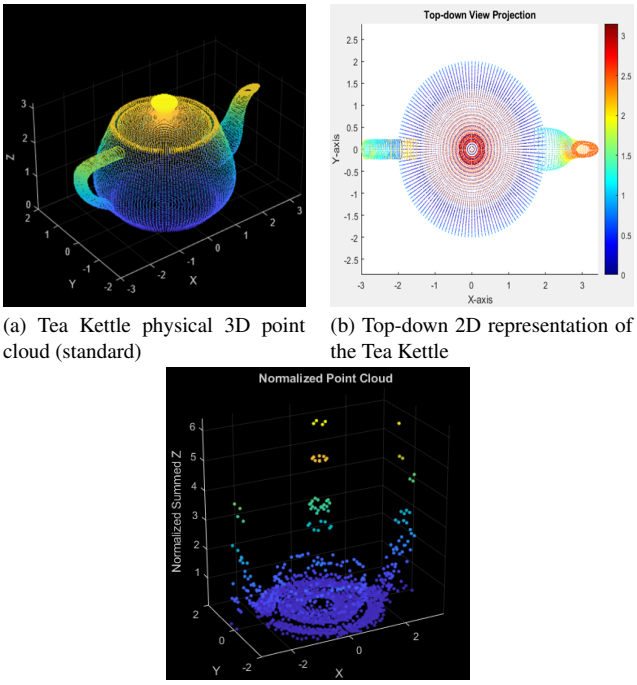
Colin Kelly
DEVCOM Army Research Laboratory
2800 Powder Mill Rd.
colin.d.kelly6.civ@army.mil

Abstract

S-band Synthetic Aperture Radar (SAR) offers advantages in foliage, ground penetration, and weather tolerance. However, it is comparatively underutilized for object classification due to the preference for higher band frequencies. This paper presents the first application of deep learning to S-band Synthetic Aperture Radar (SAR) data for local object classification. Our method of extrapolating the 2D SAR image to a 3D Radar Cross Section (RCS) response differs from previous work that uses targets physical 3D point clouds. We also present a novel preliminary 2D-3D fusion method for S-band SAR to demonstrate the effectiveness of ensemble methods on this type of data. We combine a lightweight custom convolutional neural network (CNN) with a PointNet-based network, enhancing feature extraction from image and point cloud domains. Our method is more precise and robust to clutter compared to single-modality techniques.

1. Introduction

Synthetic Aperture Radar (SAR) enables high-resolution imaging through a single moving antenna, making it invaluable for remote sensing applications such as vegetation monitoring [18], change detection in deforestation [15], flood impacts, moisture content in soil, wildfire damage assessments and astronomically has been used by satellites to map entire planets [17, 20]. Stripmap SAR, in particular, generates 2D images of 3D scenes by acquiring data along a linear trajectory, creating a "synthetic aperture" through coherent signal processing methods such as backprojection [5, 26]. The spatial resolution in the 2D plane depends on factors such as frequency, bandwidth, observation time, and aperture size. Higher-frequency bands (e.g., X- and K-band) are often preferred for small-scale tasks such as object classification [9, 24] due to their finer resolution, while lower-frequency bands (e.g., L- and S-band) are preferred for large-scale tasks such as global change detection because the waves travel significantly farther and provide superior penetration through foliage, soil, and buildings. Despite these advantages, S-band SAR remains underutilized in object classification due to its lower resolution and increased clutter (for information regarding frequency bands, see Table 1).



(a) Tea Kettle physical 3D point cloud (standard)

(b) Top-down 2D representation of the Tea Kettle

(c) Tea Kettle point cloud summed along the Z-axis (our method)

Figure 1. Our method of extrapolating 3D point clouds from 2D SAR imagery effectively sums values along the target's Z-axis (b). The tea kettle example effectively demonstrates the difference between physical (a) and summed point clouds (c)

Our interest in S-band SAR for object classification stems from a desire to detect heavily obscured objects at a distance, such as objects that have been buried, hidden in thick bushes, under branches, leaves, rubble, and more. This area of hidden object detection is a complex problem and a largely unexplored area of Computer Vision. Modern

research in object classification utilizes datasets that provide an assortment of modalities: electro-optical (EO), infrared (IR), laser (LiDAR), and radar. However, it is not easy to find datasets in any of these modalities with a primary focus on detecting heavily obscured objects. One recently published dataset similar to what we want to accomplish, PIDray, published by Zhang et al. in 2022, focuses on detecting prohibited objects using X-ray technology [27]. However, PIDray requires close contact via a dangerously high-frequency X-ray machine instead of long-range low-frequency S-band SAR.

Deep learning has significantly advanced SAR-based object recognition, CNNs largely dominate 2D image classification, and Graph Convolutional Networks (GCNs) and PointNet-based networks excel in 3D point cloud processing. However, existing methods primarily focus on high-frequency SAR [9, 24], leaving a gap in deep learning-based classification for S-band data. To bridge this gap, we propose a novel 2D-3D fusion approach that integrates a custom lightweight CNN with a PointNet-based network, enabling robust feature extraction from image and point cloud domains. This paper marks the first application of PointNet to S-band SAR data, demonstrating the effectiveness of multimodal fusion for cluttered low-frequency SAR environments.

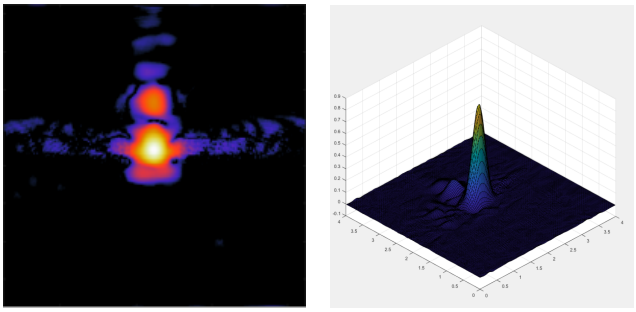
Our contributions are as follows:

- First, we introduce a 2D-3D fusion framework for S-band SAR classification, leveraging complementary image and point cloud features.
- Second, we present the first application of PointNet-based architectures to S-band SAR, improving classification accuracy in low-resolution, high-clutter conditions.
- Third, we validate our approach on a real-world dataset and demonstrate that multimodal fusion surpasses single-modality methods in accuracy and robustness.

2. Related Work

2.1. High vs Low Frequency

Due to the nature of wavelengths, using low-frequency radar for global and astronomical tasks has always been preferable. The low-band waves travel very far, very quickly, and the attenuation of signal strength is significantly lower than that of high-band frequencies, able to effectively travel through space and ultimately map entire planets [17, 20]. Due to this relationship, higher frequency data tends to make sense for small-scale tasks because the waves are not required to travel across the vast emptiness of space or penetrate forest canopies, buildings, or the ground. However, when trying to identify obscured and hidden objects at a distance, seeing them with low-frequency radar is significantly more possible. The attenuation effect on higher-frequency wavelengths by buildings, leaves, branches, and rubble is too strong to have a meaningful target response. However, through-the-wall sensing is a technology that utilizes low-frequency wifi radar signals, typically in the S-band (2.4GHz) for search and rescue and military operations [3]. These penetrative capabilities make radar an excellent sensor for obscured and buried targets that optical sensors cannot detect.



(a) Trihedral SAR image formed between 2.2-3.8GHz. Demonstrates a strong target response; defined sidelobes make this a unique and distinguishable target.

(b) The surface meshed point cloud. This is created using the same data used to create the SAR image pictured to the left.

Figure 2. Processed radar data for use in neural networks. Trihedral SAR image (a) and Trihedral point cloud (b).

2.2. Synthetic Aperture Radar

SAR was first invented in 1951 by Carl Wiley at Goodyear, now Lockheed Martin, and has been used for remote sensing ever since. NASA has used SAR to map the topography of the Earth using satellites and their Disaster Program to show earthquake fault slips. Thousands of SAR systems have been installed on various aircraft and ground vehicles [10]. SAR penetration has been an ongoing research topic for decades; in 1994, airborne SAR was performed at various low frequencies to penetrate foliage in the Panamanian rain forest and to detect buried objects at a desert site in Arizona [6]. Through-the-wall personnel sensing [3] has also been an area of interest for search and rescue missions and military operations.

As for research in object classification, MSTAR is the de facto standard for public SAR datasets. MSTAR is a military dataset of two Air Force Research Laboratory data collections from 1995 to 1996. It is an X-band SAR dataset containing SAR imagery; the first collection consists of 13 target types (20 actual targets) plus minor examples of articulation, obscuration, and camouflage. The second collection consists of 15 target types (27 actual targets). 1998 Diemunsch et al. [4] utilized a "hypothesis and test" approach with a Bayesian scoring framework for confidence scores. The Bayesian objectives were three key features: peak locations, semantically labeled regions, and the SAR image itself to produce a maximum likelihood

estimate. Fast forward to the present, modern CNN and transformer architectures dominate the space. Li et al. [12] utilized adversarial autoencoders (AAEs) to generate additional SAR chips for the MSTAR dataset and trained/tested with a YOLOv3 CNN classifier. Wang et al. [24] use the original PointNet architecture to classify high-frequency K-band point clouds of airplane targets such as F35s and F16s with great success. It is worth noting that our approach to utilizing PointView-GCN differs from the methodology used by Wang et al. because we are strictly extrapolating 3D information from 2D SAR imagery as opposed to directly collecting point clouds from radar. It isIncluding both methods could be beneficial.

2.2.1. MSTAR Dataset

As for research in object classification, MSTAR is the de facto standard for public SAR datasets. MSTAR is a military dataset of two Air Force Research Laboratory data collections from 1995 to 1996. It is an X-band SAR dataset containing SAR imagery; the first collection consists of 13 target types (20 actual targets) plus minor examples of articulation, obscuration, and camouflage. The second collection consists of 15 target types (27 actual targets). 1998 Diemunsch et al. [4] utilized a "hypothesis and test" approach with a Bayesian scoring framework for confidence scores. The Bayesian objectives were three key features: peak locations, semantically labeled regions, and the SAR image itself to produce a maximum likelihood estimate. Fast forward to the present, modern CNN and transformer architectures dominate the space. Li et al. [12] utilized adversarial autoencoders (AAEs) to generate additional SAR chips for the MSTAR dataset and trained/tested with a YOLOv3 CNN classifier. Wang et al. [24] use the original PointNet architecture to classify high-frequency K-band point clouds of airplane targets such as F35s and F16s with great success. It is worth noting that our approach to utilizing PointView-GCN differs from the methodology used by Wang et al. because we are strictly extrapolating 3D information from 2D SAR imagery as opposed to directly collecting point clouds from radar. Including both methods in the ensemble network could be beneficial.

2.3. Robustness

SAR has many distinct benefits for object detection; data is invariant to the time of day because radar provides its illumination [1]. Relatively long wavelengths can penetrate through weather phenomena such as clouds and fog [1]. Radar provides a ground resolution that does not degrade with distance (target response degrades, and noise can increase as the signal attenuates over distance). Azimuth and range resolutions are independent of the distance between the sensor and the ground [1]. Radar measurements are naturally precise due to signal coherence; radar imaging systems can be configured to have outstanding geometric accuracy [5, 10]. At lower radar frequencies, the system becomes more susceptible to errors. However, SAR is tolerant of positioning errors due to its gathering information at many positions along the aperture. Radar illumination is coherent; radar antennas emit consistent and controlled energy pulses. Natural coherence enables the creation of digital elevation models and sensitive measurements of topographic changes over time and is the foundation for SAR images [1].

2.4. Point Cloud Processing

A typical disadvantage of point-based representation is that surface information is implied by point density and orientation [23]. Point clouds are commonly oversampled, containing more data than necessary for object representation. Furthermore, when point clouds are converted into 3D voxels or image grids before being fed to a deep net architecture, the data can become unnecessarily large and contain artifacts. Since voxels represent only a single point on a grid, the space between each voxel is not represented and would have to be reconstructed or approximated via interpolation. Despite this, we can voxelize our data due to the low-frequency radar data we operate. Our point density is already sparse due to the low frequency; refer to Figure 2b. We are classifying on rather unique "blobs" within our images representing the target's response on a logarithmic scale; refer to Figure 2a.

2.5. PointNets

PointNet, introduced by Qi et al. [21] in 2016, is a neural network designed to process raw point clouds for classification and segmentation tasks directly. PointNets use only a single symmetric function, max pooling, to learn the optimization functions and criteria to select only the most "important" points, downsampling point clouds for efficiency. The original PointNet architecture is not capable of capturing local structures. This limits the network's ability to recognize fine-grained patterns and generalizability to complex scenes. In 2017, Qi et al. [22] introduced PointNet++ to improve upon PointNet by recursively using PointNets on nested partitions of the input point clouds, allowing the network to learn local features and their descriptors and individual pixels to be assigned labels for semantic segmentation. PointNet++ also introduced a convolution operation as the kernel to learn local spatial features, introduced a multi-scale and multi-resolution grouping to deal with the variation in different areas, and introduced farthest point sampling (FPS) to sample points more efficiently. PointView-GCN is a PointNet network published by Mohamaddi et al. [16] in 2021. They use multi-level Graph Convolutional Networks (GCNs) to hierarchically aggregate shape features of single-view point clouds to encode geometrical object cues and their multi-view relations, producing a more

descriptive global shape feature to improve classification accuracy. PointView-GCN is the architecture that we will be using for our results because it has the highest accuracy results on the ModelNet-40 dataset with a public codebase [25] (at the time of writing, PointView-GCN is temporarily unavailable on the ModelNet40 Benchmark leaderboard for reasons unknown to us). Multi-view processing is beneficial for our task because synthetic aperture data allows us to split our single large aperture into as many apertures as we see fit. In our case, the left half, right half, and whole aperture are processed as separate views refer to Figure 3.

2.6. Attention and Fairness

The field of explainable AI seeks to characterize model accuracy, fairness, transparency, and outcomes. Neural networks utilize attention mechanisms to weight importance of regions within data. It has been established that neural networks tend to take shortcuts and converge on incorrect features commonly associated with the intended target class [8] instead of the target features themselves. PointNet [21] introduced the concept of "Critical Points", a subset of points deemed important enough to remain as individual points after the final pooling layer. Levi et al. [11] concluded that PointNet neural networks have a habit of attributing outliers and coincidental features as influential or critical to the target; this is a real problem in Synthetic Aperture Radar as lower frequencies tend to have higher noise. It is possible for the ground response to possibly drown out weaker target responses and lead to our PointNet classifying on background noise. By defining the "focus" of a point cloud as the normalized entropy, they then "refocus" the influence from the most influential critical points to the least influential points and see a similar but more generalizable, robust performance. Refocusing our points would be difficult because our point clouds are generally very noisy due to the low frequency data and strong ground response

3. Methodology

Our side-looking aperture radar (SLAR) is collected in a controlled environment with a motorized planar scanner and a rotating table. The planar scanner has two translational degrees of freedom and can move 9 meters along the X axis and 2 meters along the Z axis, refer to Figure 4. Our UXO dataset has six explosive targets and four clutter targets split between two distinct categories, explosive and non-explosive. The primary task of our network is binary classification, differentiating a target into the respective category and whether or not it is an explosive object. The secondary task of multi-class classification is identifying the exact object within each category.

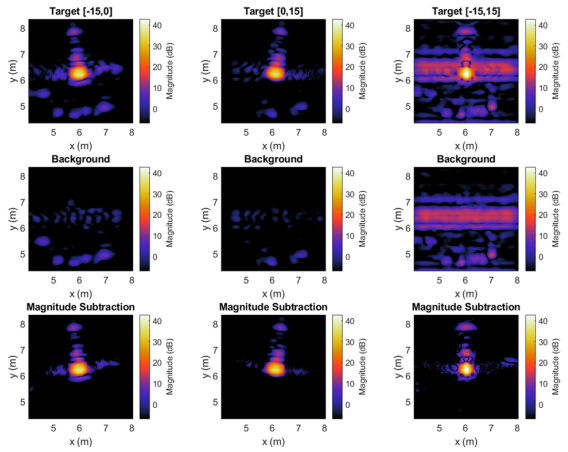


Figure 3. SAR images for each independent aperture angle: Left column (right aperture), Center column (left aperture), Right column(whole aperture)

3.1. Hardware

All data is taken using custom-built hardware. The Radio Frequency System On Chip (RFSoC) is built and produced by Huntington Ingalls Industries (HII). The RFSoC is connected to four antennas, two transmitters, and two receivers to capture all four polarizations: HH, HV, VH, and VV (e.g., A VH polarization means a vertical transmitter and horizontal receiver) simultaneously. The RFSoC and antenna are securely mounted on our planar scanner to ensure the accuracy of motor units for accurate spatial SAR image formation (Refer to Figure 4).

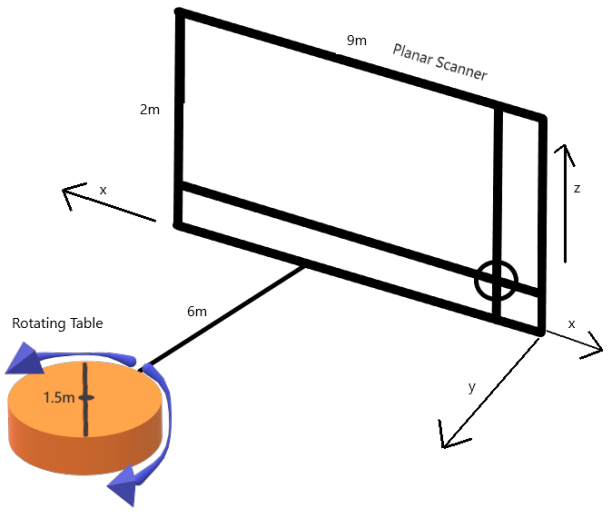


Figure 4. Planar Scanner and Rotating Table Diagram

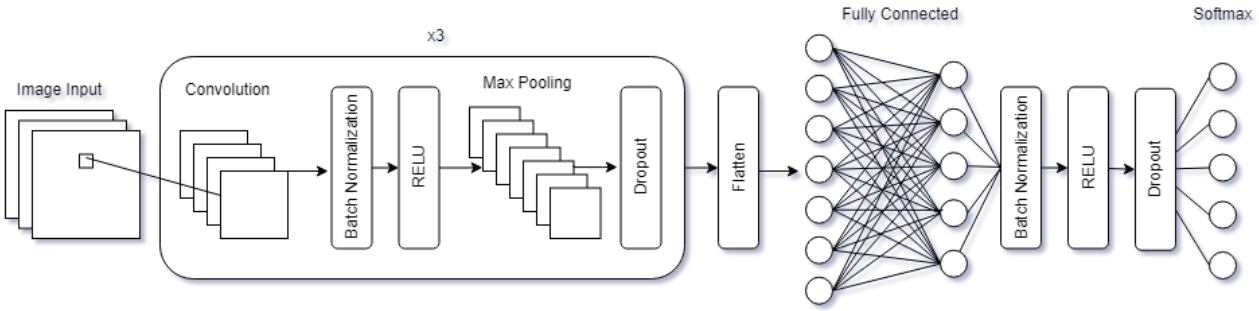


Figure 5. RadarCNNv1 used to classify SAR images.

3.2. Data Collection

Each target has data collected at three heights on the Z axis, 1, 1.5, and 2 meters. At each respective height, the rotating table does a full 360-degree rotation in 20 degree increments to account for different angle responses. Every time the table rotates 20 degrees, the scanner does a single linear pass, creating an aperture for a SAR image. The target is placed in the center of the rotating table, directly facing the linear scanner at a ground range of 6 meters. The table starts rotated at -180 degrees and stops at 180 degrees. It is worth noting that some targets within our dataset are asymmetrical and have angles where the response is significant, but other angles where the response is buried deep within the noise (Refer to Figure 6).

By defining each linear pass as a separate aperture, we can generate distinct SAR images at each table azimuth angle. Then, each of those linear apertures can be divided further into nearly infinite sub-apertures. For our purposes, we will subdivide each of our linear apertures into a left and right half to simulate stereo vision. Therefore, each pass will generate three unique SAR images by segmenting the aperture into the left half, right half, and whole aperture; right-squinted, left-squinted, and zero-squint images, respectively (Refer to Figure 3). Background data is collected on each respective day of data collection for use in background subtraction.

We are collecting data for four polarizations (HH, HV, VH, VV), and each 360 degree table rotation produces 19 target apertures from the 20 degree rotation increments, with two identical images being generated at -180 and 180 degrees. For each of these 19 target apertures, we produce our multi-view left-squinted, right-squinted, and whole aperture, totaling 57 for images. Since we are repeating the above collection at three distinct heights, each polarization will have $57 * 3 = 171$ images. Accounting for all four polarizations, each target will have $171 * 4 = 684$ unique SAR images. After performing background subtraction, each target class will have $684 * 2 = 1368$ SAR images. We then re-use the 2D SAR image to create 3D point clouds by extrapolating the RCS response at each pixel and assigning it as the z value for each (x,y) image coordinate, producing a voxelized point cloud for each SAR image. Thus, we can run our 2D CNN on the SAR images and use the same SAR data to create 3D point clouds that the PointView-GCN network can process.

3.3. Networks

The two networks, PointView-GCN [16] and our custom RadarCNNv1 in Figure 5, are trained independently with the same training/test split of $50/50$. We chose this abnormal training/test split due to the similarity in data; many targets do not drastically change with each 20-degree rotation. Predictions are calculated using the naive approach of a weighted average between the two network predictions. The 2D SAR data for the CNN is augmented using affine transformations, rotation, translation, reflection, and shearing. The PointView-GCN network augments data using similar affine transformations, rotation, scaling, and translations and introduces Gaussian jitter to the 3D points.

3.4. PointNet Training

Given a point set $\{p_i\}_{i=1}^{n}$ where $\{p_i\} \in \mathbb{R}^3$ is represented by a matrix $x_0 \in \mathbb{R}^{n\times 3}$ we use a training/test split of $50/50$, and the PointNet training data is augmented with affine transformations such as scale, reflection, rotation, and shearing. Gaussian jitter is also applied to perturb the original data.

$$x_k = x_0 + \lambda z_k \tag{1}$$

where $z_k \sim \mathcal{N}(0, I)$ is a random noise vector from a normal distribution and λ is a scalar value to control the noise magnitude [24]. This network also supports multi-view processing, allowing us to pass our left, right, and whole aperture point clouds as multi-channel data.

3.5. CNN Training

The RadarCNNv1 network is capable of accepting 2D SAR imagery as well as raw, complex 2D radar data. For our

purposes, we trained exclusively on image data. We experimented with training on complex matrix data; however, due to the scene of our data collection, there is significant multipath noise. Each target had distinguishable electromagnetic interactions with the background, and we were achieving 100% accuracy due to the network converging on background multipath features instead of target features.

Our CNN model is simple and implements standard, well-researched layers with a focus on preventing overfitting. Our main module involves convolution through our image input, which is our three SAR aperture images stacked on top of each other. Channel 1 is the left aperture, channel 2 is the right aperture, and channel 3 is the whole aperture. After convolving, we perform batch normalization on our data to normalize the activations within each layer. We then apply the RELU activation function to introduce non-linearity to our model. Max pooling allows us to reduce the dimensionality of our features, and finally, the dropout layer ensures that all output nodes will be relied upon for classification information. We repeat this convolution three times with an increasing number of filters to capture both local and global features.

4. Experiments

The initial images had a pixel resolution of 200×200 and a spatial resolution of $4m^2$, each pixel representing a space of $2cm^2$. However, the size of the target response never exceeded a single square meter; therefore, in our final results, we have chosen only to include 50×50 images covering a spatial resolution of a single square meter centered on the target response as we can segment our synthetic aperture into as many apertures as we see fit. We train two separate models, one combining the three target look angles (left, right, and full aperture) into a three-channel SAR image and the other training on every aperture image separately.

4.1. Results

In Tables 1 and 2, we see that combining the three apertures produces significantly stronger results in both multi-classification and binary classification tasks, respectively achieving F1-scores of 59.74% and 88.07%; however, it is less likely that data collected in the field will have all of these views available due to obscuration. Adding the PointView-GCN predictions to our CNN predictions always resulted in a higher accuracy and F1 score.

4.2. Ablation

4.2.1. Network Weighting

Initially, we applied an equal 50-50 weight to our CNN and PointNet networks. However, PointNet tends to perform better on targets with lower signal-to-noise ratios (SNR), such as in figure 6 due to the shape of the target response being nearly indistinguishable in 2D images as opposed to the 3D point clouds. The CNN and PointNet predictions each have a minimum weight of 20%, and the other 60% is distributed by taking the SNR values of each image and proportionately distributing weights between the maximum SNR values. We see in Table 3 that the improvements from SNR weighting, while minor, still see improvement.

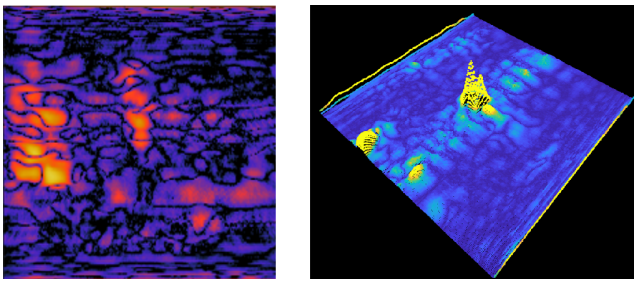


Figure 6. SAR images with weaker target responses (left) are more likely to be correctly identified by their point cloud (right)

4.2.2. MSTAR Dataset

As previously stated, MSTAR is the de facto standard of public SAR datasets. We use the MSTAR Public Mixed Target dataset, eight targets, seven of which are ground vehicles, and one is a calibration target (like our trihedral). This sub-dataset is used to compare the usefulness of our method in S-band radar with that of the higher X-band MSTAR radar dataset. Similar to our dataset, we take the intensity values at each pixel and assign them as the Z-value for the 3D points to create point clouds from 2D SAR imagery directly. We conclude that there is no notable increase to the RadarCNNv1 performance with or without the PointView-GCN; refer to Table 4. We also verify that both networks are independently performing exceptionally well on this test dataset, with all metrics coming in above 98%. We compare this to the benchmark results of Amrani et al. [2] in 2021. Unfortunately, finding a central leaderboard for this dataset is difficult, but our CNN results are competitive with the state-of-the-art results. When downloading the dataset directly from the official Air Force Research Laboratory (AFRL) website, there is no distinct train and test set, so it is possible that our training/test splits do not match that of other papers and their subsequent results. NUDT4MSTAR [14] by Liu et al. was recently accepted into CVPR 2025 with a SAR Automatic Target Recognition (ATR) dataset that is 10x larger than the highly impactful MSTAR vehicle SAR ATR dataset. In the future, NUDT4MSTAR could become the benchmarking standard that supersedes MSTAR.

Table 1. Multiclass Classification Results

Data	Network	Accuracy	Precision	Recall	F1-Score
50x50x1	RadarCNNv1	47.63	48.34	48.17	48.25
	PointView-GCN	34.42	39.24	35.20	37.11
	PointView-GCN + RadarCNNv1	**48.07**	**51.27**	**48.80**	**50.01**
50x50x3	RadarCNNv1	56.93	58.92	57.24	58.07
	PointView-GCN	41.32	45.07	42.47	43.73
	PointView-GCN + RadarCNNv1	**57.89**	**61.17**	**58.37**	**59.74**

Table 2. Binary Classification Results

Data	Network	Accuracy	Precision	Recall	F1-Score
50x50x1	RadarCNNv1	79.33	78.55	**78.31**	78.43
	PointView-GCN	71.75	70.81	69.26	70.03
	PointView-GCN + RadarCNNv1	**79.82**	**79.33**	78.27	**78.80**
50x50x3	RadarCNNv1	86.58	86.28	87.45	86.86
	PointView-GCN	78.60	77.91	77.68	77.79
	PointView-GCN + RadarCNNv1	**87.98**	**87.53**	**88.61**	**88.07**

Table 3. SNR vs Equal Weighting (Combined Networks)

Data	Network	Accuracy	Precision	Recall	F1-Score
Equal Weight	50x50x1	48.07	50.24	48.75	49.48
	50x50x3	**57.98**	59.95	58.36	59.14
	50x50x1 Binary	79.71	79.10	**78.34**	78.72
	50x50x3 Binary	87.89	87.44	88.50	87.97
SNR Weight	50x50x1	48.07	**51.27**	**48.80**	**50.01**
	50x50x3	57.89	**61.17**	**58.37**	**59.74**
	50x50x1 Binary	**79.82**	**79.33**	78.27	**78.80**
	50x50x3 Binary	**87.98**	**87.53**	**88.61**	**88.07**

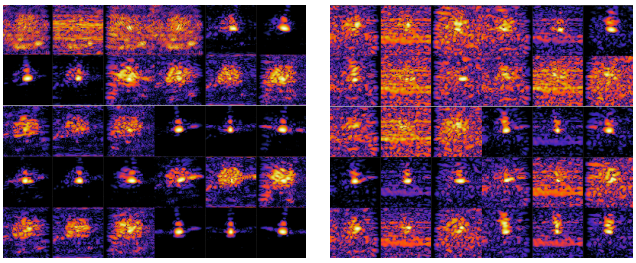
Table 4. MSTAR Ablation Classification Results

Network	Accuracy	Precision	Recall	F1-Score
RadarCNNv1	**99.88**	**99.89**	**99.91**	**99.90**
PointView-GCN	98.94	98.95	98.74	98.84
PointView-GCN + RadarCNNv1	**99.88**	**99.89**	**99.91**	**99.90**

5. Conclusion and Future Directions

In this paper, we have introduced the first application of deep learning to S-band Synthetic Aperture Radar. We have introduced a novel method of point-cloud generation using RCS values directly from SAR images as an alternative to physical object point clouds and established the first CNN+PointNet network for S-band Radar. Moreover, we demonstrated that by combining these two distinct networks, we could exploit our radar data for more robust and improved results.

In the future, it could be interesting to run separate tests for each of the polarizations as co-polarization responses HH and VV tend to have stronger RCS responses than cross-polarizations HV and VH (Figure 7). When we decided to use PointView-GCN for our testing, it was at the top of the ModelNet40 leaderboard. Now, there have been many advances since PointView-GCN [16] in the way of



(a) Co-polarization SAR Trihedral Images.

(b) Cross polarization SAR Trihedral Images.

Figure 7. Comparison of a. Co-polarization (HH: Horizontal transmitter and horizontal receiver) and b. Cross-Polarization (HV: Horizontal transmitter and vertical receiver) for SAR Trihedral Images.

Mamba3D [7], Point DeepONet [19], and PointGST [13].

6. Acknowledgments

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-24-2-0049. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] THOMAS P. AGER. An introduction to synthetic aperture radar imaging. *Oceanography*, 26(2):20–33, 2013. 3

[2] Moussa Amrani, Abdelatif Bey, and Abdenour Amamra. New sar target recognition based on yolo and very deep multi-canonical correlation analysis. *International Journal of Remote Sensing*, 43(15–16):5800–5819, 2021. 6

[3] Kevin Chetty, Graeme E. Smith, and Karl Woodbridge. Through-the-wall sensing of personnel using passive bistatic wifi radar at standoff distances. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1218–1226, 2012. 2

[4] Joseph R. Diemunsch and John Wissinger. Moving and stationary target acquisition and recognition (MSTAR) model-based automatic target recognition: search technology for a robust ATR. In *Algorithms for Synthetic Aperture Radar Imagery V*, pages 481 – 492. International Society for Optics and Photonics, SPIE, 1998. 2, 3

[5] NASA Earth Science Data Systems. What is synthetic aperture radar?, 2020. 1, 3

[6] Dominick A. Giglio. Overview of foliage/ground penetration and interferometric SAR experiments. In *Algorithms for Synthetic Aperture Radar Imagery*, pages 209 – 217. International Society for Optics and Photonics, SPIE, 1994. 2

[7] Xu Han, Yuan Tang, Zhaoxuan Wang, and Xianzhi Li. Mamba3d: Enhancing local features for 3d point cloud analysis via state space model, 2024. 8

[8] Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial examples, 2021. 4

[9] Richard O. Lane, Adam J. Wragge, Wendy J. Holmes, Stuart J. Bertram, and Tim Lamont-Smith. Object detection in eo/ir and sar images using low-swap hardware. In *2021 Sensor Signal Processing for Defence Conference (SSPD)*, pages 1–5, 2021. 1, 2

[10] Stephen W Lasswell. History of sar at lockheed martin (previously goodyear aerospace). In *Radar Sensor Technology IX*, pages 1–12. SPIE, 2005. 2, 3

[11] Meir Yossef Levi and Guy Gilboa. Robustifying point cloud networks by refocusing, 2024. 4

[12] Siyan Li, Yue Xiao, Yuhang Zhang, Lei Chu, and Robert C. Qiu. Learning efficient representations for enhanced object detection on large-scene sar images, 2022. 3

[13] Dingkang Liang, Tianrui Feng, Xin Zhou, Yumeng Zhang, Zhikang Zou, and Xiang Bai. Parameter-efficient fine-tuning in spectral domain for point cloud learning. *arXiv preprint arXiv:2410.08114*, 2024. 8

[14] Yongxiang Liu, Weijie Li, Li Liu, Jie Zhou, Bowen Peng, Yafei Song, Xuying Xiong, Wei Yang, Tianpeng Liu, Zhen Liu, and Xiang Li. Atrnet-star: A large dataset and benchmark towards remote sensing object recognition in the wild, 2025. 6

[15] Richard Lucas, Daniel Clewley, Ake Rosenqvist, Josef Kellndorfer, Wayne Walker, D.H. Hoekman, Masanobu Shimada, and Humberto Navarro de Mesquita Junior. *Global Forest Monitoring with Synthetic Aperture Radar (SAR) Data*, pages 273–297. CRC Press, 2016. 1

[16] Seyed Saber Mohammadi, Yiming Wang, and Alessio Del Bue. Pointview-gcn: 3d shape classification with multi-view point clouds. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 3103–3107. IEEE, 2021. 3, 5, 7

[17] NASA. Planetary fact sheet. https://nssdc.gsfc.nasa.gov/planetary/factsheet.html, 2024. Accessed: 2024-12-21. 1, 2

[18] Ramesh K. Ningthoujam, Heiko Balzter, Kevin Tansey, Keith Morrison, Sarah C.M. Johnson, France Gerard, Charles George, Yadvinder Malhi, Geoff Burbidge, Sam Doody, Nick Veck, Gary M. Llewellyn, Thomas Blythe, Pedro Rodriguez-Veiga, Sybrand Van Beijma, Bernard Spies, Chloe Barnes, Marc Padilla-Parellada, James E.M. Wheeler, Valentin Louis, Tom Potter, Alexander Edwards-Smith, and Jaime Polo Bermejo. Airborne s-band sar for forest biophysical retrieval in temperate mixed forests of the uk. *Remote Sensing*, 8(7), 2016. 1

[19] Jangseop Park and Namwoo Kang. Point-deeponet: A deep operator network integrating pointnet for nonlinear analysis of non-parametric 3d geometries and load conditions, 2024. 8

[20] ESA Earth Observation Portal. Nisar - nasa-isro synthetic aperture radar mission. https://www.eoportal.org/satellite-missions/nisar#eop-quick-facts-section, 2024. Accessed: 2024-12-21. 1, 2

[21] Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *CoRR*, abs/1612.00593, 2016. 3, 4

[22] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *CoRR*, abs/1706.02413, 2017. 3

[23] Tuan-Anh Vu, Srinjay Sarkar, Zhiyuan Zhang, Binh-Son Hua, and Sai-Kit Yeung. Test-time augmentation for 3d point cloud classification and segmentation, 2023. 3

[24] Mou Wang, Shunjun Wei, Hao Su, Qizhe Qu, Min Yan, and Jun Shi. Object recognition of three-dimensional sar based on pointnet. In *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, pages 1–6, 2019. 1, 2, 3, 5

[25] Papers with Code. Modelnet40 benchmark, 2024. 4

[26] A.F. Yegulalp. Fast backprojection algorithm for synthetic aperture radar. In *Proceedings of the 1999 IEEE Radar Conference. Radar into the Next Millennium (Cat. No.99CH36249)*, pages 60–65, 1999. 1

[27] Libo Zhang, Lutao Jiang, Ruyi Ji, and Heng Fan. Pidray: A large-scale x-ray benchmark for real-world prohibited item detection, 2022. 2