A Robust Object Detection Technique using Multi-Stage Filter Augmentation and Adaptive Sample Selection for SAR Images

Shivanand Manyar
Department of Information Technology,
Vidyalankar Institute of Technology,
Mumbai

Hrishita Thanekar
Department of Electronics and Telecommunication,
Vidyalankar Institute of Technology,
Mumbai

Akhil Masurkar
Department of Electronics and Computer Science,
Vidyalankar Institute of Technology,
Mumbai

Mayur Rajkundal
Department of Electronics and Computer Science,
Vidyalankar Institute of Technology,
Mumbai

Vedant Daware
Department of Electronics and Computer Science,
Vidyalankar Institute of Technology,
Mumbai

ABSTRACT

Object detection from Synthetic Aperture Radar (SAR) imagery is picking up steam with SAR's all-weather, day and night imaging capabilities. Object detection within SAR images is difficult due to speckle noise, lack of texture, and significant domain shift from optical images for which deep learning models are pretrained. The study proposes to mitigate this with a large dataset and a Multi-Stage Filter Augmentation (MSFA) framework with improved detection performance with diverse backbones and anchor-based assignment methods as suggested by Y. Li et al. The contribution of this work extends this by keeping the MSFA-based pretraining with the highest performing ConvNeXt backbone while adding a change in the anchor box assignment method. Specifically, by using Adaptive Training Sample Selection (ATSS), an anchor-free, statistics-based sample selection method with an existing MSFA-based approach, replacing heuristic-based systems like Faster R-CNN. Experiments show that adding ATSS significantly enhances the detection model's generalizability, particularly in noisy or low-contrast SAR environments. This paper compares the baseline MSFA-based systems with the proposed pipeline using ATSS and demonstrates that ATSS outperforms in detecting small and cluttered objects.

General Terms

Machine Learning, Remote Sensing, Synthetic Aperture Radar, Object Detection, Deep Learning, Domain Adaptation

Keywords

Synthetic Aperture Radar, Machine Learning, ATSS, Object Detection

1. INTRODUCTION

Synthetic Aperture Radar (SAR) imaging has been a flexible remote sensing imaging modality because it can operate regardless of weather and lighting conditions. This capability makes SAR highly useful for mission-critical uses such as defense surveillance, oceanic monitoring, disaster relief, and city infrastructure inspection.

However, while SAR image analysis possesses its advantages, it is also a difficult problem for object detection tasks. Some of its challenges are the existence of speckle noise, lack of color and fine texture information, and the enormous domain disparity between SAR and natural RGB imagery—upon which most deep learning models are conventionally trained. These limitations make it difficult to directly use conventional object detection models like YOLO, SSD, and Faster R-CNN on SAR data, resulting in degraded performance.

Another obstacle is the lack of large-scale, annotated SAR datasets. Since SAR data is strategic and proprietary, publicly available datasets are small-scale, limited in diversity, and poor in object category coverage. Although some datasets such as have been useful for certain use-cases, they are not capable of facilitating general-purpose, multi-category object detection. This limitation led to SARDet-100K—a large-scale, multicategory SAR object detection benchmark with images and object annotations.

To close the domain gap and improve performance in SAR object detection, the MSFA framework was proposed by Y. Li et al. This tackles both the domain gap and model gap by using handcrafted feature augmentation to reduce the data distribution gap between SAR and RGB. Hierarchical domain transition uses an intermediate optical remote sensing dataset and end-to-end pretraining of the entire detection model instead of pretraining just the backbone. The MSFA method produced orders of magnitude improvements in detection across different architectures and benchmark datasets when combined with backbones like ConvNeXt.

The detection head in the pipeline was largely anchor-based, using approaches like FRCNN and its variants to select samples and localize the object. While anchor-based methods can be effective, they heavily rely on heuristics defined manually for the anchor box assignment, which runs the risk of not generalizing in SAR datasets where object shapes, sizes, and oriented objects can vary. Rather than relying on heuristic approaches for anchor box assignments, this study proposes improvement by replacing anchor box assignment with Adaptive Training Sample Selection (ATSS).

ATSS is a statistics-based, anchor-free approach that selects positive samples based on the statistical properties of Intersection over Union (IoU) distributions. ATSS can adapt to object scale, image noise, and detection difficulty, revealing greater robustness than assigned fixed thresholds for anchor box assignment. As ATSS integrates with the MSFA pipeline with the ConvNeXt backbone, it generates superior generalization properties, especially in scenarios including small targets, clustered objects, or cluttered backgrounds.

This paper presents a comparative analysis of the original MSFA+FRCNN framework and the proposed MSFA+ATSS approach. The same pretraining strategy and backbone were maintained while the anchor box assignment framework was modified to assess and isolate the impact of the ATSS, or the statistical approach, on detection performance. This study demonstrates how ATSS provided superior generalization across the wide range of SAR scenarios observed in these experiments, demonstrating the potential for ATSS to be used in other SAR-based object detections.

2. LITERATURE SURVEY

The study of object detection on SAR imagery and its importance has grown with applications in military surveillance [1], maritime monitoring, disaster relief [2], and remote sensing [3]. A major limitation in the literature area has been the lack of large-scale, multi-category SAR datasets. Optical datasets existed with large scope, such as Microsoft COCO [4]; however, datasets related to SAR have been limited in data sets and in the diversity and complexity of the data mentioned, often focusing on ship data sets [5]. Hence, SARDet-100K was proposed to act as a unified benchmark [7], combining the ten public datasets together.

Earlier methods of detecting SAR objects included purely handcrafted features that were either Haar-like descriptors [8], Histogram of Oriented Gradients (HOG) [9], or Canny Edge detection [10]. These methods utilized classifiers such as SVMs [11] and AdaBoost paired with the handcrafted features, often provided minimal structural understanding, and did not generalize well across different SAR conditions. Methods that were developed later included wavelet-based methods and the use of feature pairs like the wavelet scattering network. There were also region-based detectors, including Deformable Convolutional Networks (DCN) [13]. Other studies incorporated genetic based Haar filters and considered it as shape identification for adaptive feature selection [14]. These conventional methods often failed due to noise, and variance in object scale.

Convolutional Neural Networks (CNNs) have established themselves as the go-to approach for SAR object detection with the rise of deep learning. CNNs can extract deep, hierarchical features from raw SAR data, which increase robustness to effects such as speckle noise, clutter, and variability in orientations of objects across the same scene. Region-based CNNs such as FRCNN [15] have been well-suited for SAR tasks where the object is small or occluded, such as ships [16]. Cascaded models have also been successfully applied for SAR, such as Cascade RCNN [17], where the improvement in accuracy is due to multi-stage refinement of the object proposals. More broadly, more complex CNN architectures have been applied in SAR tasks for locational accuracy, such as Grid R-CNN [18], which treats bounding box regression as a spatial grid and applies grid-based refinement on SAR images with distortion. When it comes to real-time applications where speed is prioritized, single-stage detectors have dominated CNN use on SAR images, such as the common approaches YOLO and SSD [19].

The recent developments have also incorporated attention mechanisms and Transformer-based models. The Geospatial Transformer [21] and the Swin Transformer [22] offer an effective approach for SAR by modelling long-range dependencies and hierarchical relationships within features and exhibiting some robust resilience to noise. In the context of detection, anchor-free detectors, such as CenterNet++ [23], remove reliance on anchor boxes, opting instead to predict the object centres and sizes directly beneficial in cases where the objects may be of different orientations and scales.

Regardless of these advances, there is still a major issue: the domain gap between optical image datasets (i.e., ImageNet [24]) used to pre-train and SAR image data. Optical images represent coloration and texture, while SAR images depict backscatter intensities. Therefore, models trained with RGB data do not generalize well when used for fine-tuning with SAR data.

Recent studies have researched recent data-driven augmentation techniques. Jin et al. [12] proposed frequency domain feature augmentation to extract more distinctive spectral target features in SAR images to improve discriminability of models. Zhang et al. [27] explored the augmentation of domain-specific features, which improved detection performance by augmenting textures of deep learning to enable the free transfer to SAR image textures. Liu et al. [20] introduced Dynamic Anchor Boxes (DAB-DETR), which promoted object detection acceptance by dynamically adjusting the bounding boxes to improve the accuracy of shape and object size for SAR-based ship detection.

3. METHODOLOGY

The proposed framework combines multi-sensor data fusion with deep learning methods for supporting robust object detection. The procedural pipeline begins with data preparation and augmentation, where input images from three major channels are utilized. These filter augmented datasets are used in various stages for training and testing the model for object detection.

3.1 Dataset

Object detection based on deep learning techniques in SAR imagery is often challenged by the limited amount of large, diverse, and publicly available annotated datasets. This study uses three datasets, namely ImageNet, DOTA, and SARDet-100K.

The ImageNet database is a massive visual database commonly adopted in computer vision research for pretraining deep learning models. It exhibits rich and varied visual features that are crucial for initializing neural networks prior to transferring them to downstream applications. Pretraining on ImageNet enables models to learn basic visual features that can be fine-tuned using SAR-specific datasets to learn the special properties of radar images.

Subsequently, The DOTA (Dataset for Object detection in Aerial Images) dataset was annotated for 15 object classes such as vehicles, ships, aircraft, buildings and more [30]. It has high object density and complex scene compositions from an aerial perspective, which make it a good transitional dataset to reduce the differences between a normal image (i.e., ImageNet) and SAR images. It retains spatial reasoning and object localization within the patterned aerial context. The optical to SAR transition represents a significant element of the Multi-Stage Pretraining with Filter Augmentation (MSFA) and builds on the model's representation quality when later fine tuned on SARDet- 100K.

The SARDet-100K is a large-scale SAR object detection benchmark that combines and synthesizes many existing datasets such that they are presented in a common format usable in existing object detection frameworks. It consists of approximately 117,000 images with more than 246,000 object annotations distributed across six categories (i.e., ship, aircraft, bridge, harbor, tank, and car), thus providing a common and diverse basis to train and evaluate SAR-based detection.

The datasets included in SARDet-100K include various sensors, resolutions, polarizations, and imaging frequencies as described in Table 1. The SARDeT100K dataset consists of data from various SAR systems, including Gaofen-3, Sentinel-1, TerraSAR-X, RADARSAT-2, and HISEA-1, and includes both airborne and spaceborne SAR platforms. The different datasets combined to form SARDet-100K had spatial resolutions from less than 1 to 25 meters. In addition, these datasets are available in different polarization modes. Overall, SARDet-100K reflects the diversity and variety of real-world SAR, making this source of data a highly representative and comprehensive environment for SAR based object detection.

Table 1 Data used **in SARDet-100K (S-Ship, B-Bridge, H-Harbor)**

Dataset	Object type	Satellite	Polarization	Resolution
AIR SARShip 1 .0[25]	S	Gaofen-3	VV	1.3m
OGSOD	B, H	Gaofen-3	VV/VH	3m
HRSID	S	Sentinel-1B, TerraSAR-X, TanDEM-X	HH, HV, VH, VV	0.5-3m
MSAR	S, B	HISEA-1	HH, HV, VH, VV	≤1m
SSDD	S	Sentinel-1, RadarSat-2, TerraSAR-X	HH, HV, VH, VV	1-1.5m
ShipData set	S	Sentinel-1, Gaofen-3	HH, HV, VH, VV	3-25m

3.2 Multistage Filter Augmentation (MSFA)

A significant preprocessing step includes feature enhancement using wavelet-based filtering, where various filters such as HOG, Haar and WST were extensively tried. Empirical results depicted that the Wavelet Scattering Transform (WST) worked better than these options, as supported since it has the capability of providing stable, noise-invariant multi-scale representations. The input to the model is formed by concatenating WST with the original image of the dataset of that stage.

The MSFA method uses filter augmentation as a bridge between domains. Synthetic Aperture Radar (SAR) images differ significantly from optical images used in DOTA and ImageNet in terms of texture and noise. Applying the same filter to all the datasets helps in reducing the domain gap, achieving consistent training conditions and enhancing generalizability.

3.3 Backbone Pretraining with ConvNeXt

The ImageNet dataset, with WST feature maps concatenated to each image, is utilized to train the ConvNeXt-B backbone [29]. This technique confirms that the backbone adapts to multichannel input early on, synchronizing training across all subsequent domains. ConvNeXt, a hierarchical convolutional architecture inspired by vision transformers that use large kernel depth wise convolutions, GELU activations, and layer normalization, enables rich multi-scale feature extraction. The network's structure includes a stem and four staged layers and allows it to stably learn representations for objects of diverse sizes and orientations, which is extremely useful for SAR object detection. Initializing the backbone with weights fitted to filter augmented ImageNet data helps the model gain generalized visual features that are compatible with downstream optical and SAR detection tasks that also use filter-augmented images for input.

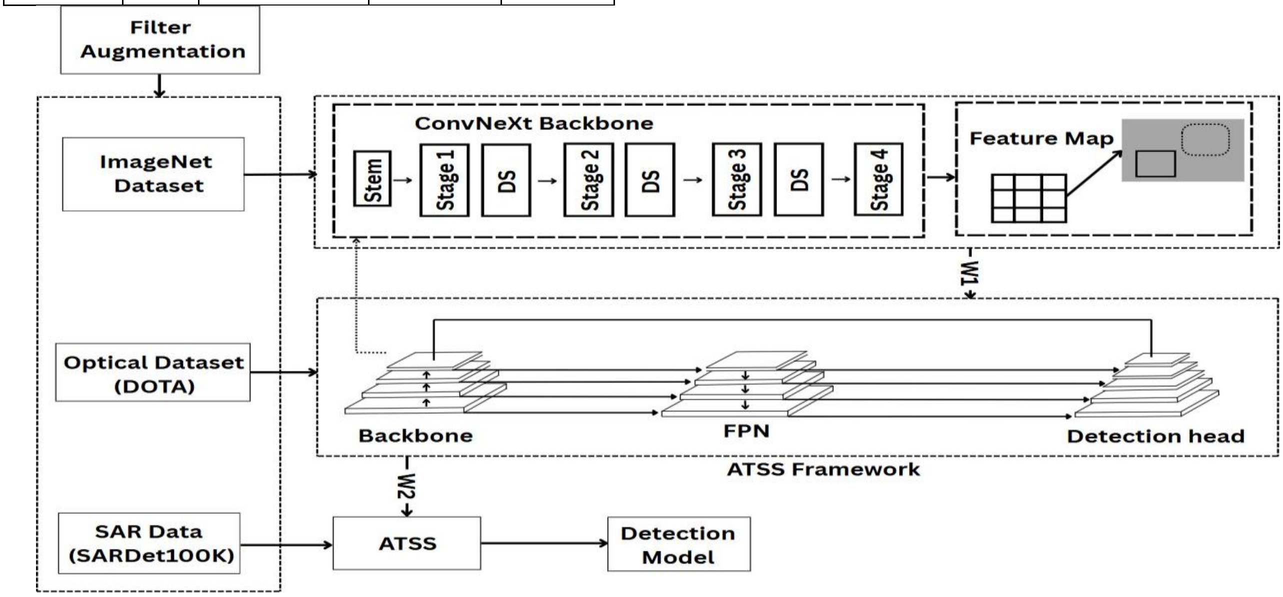

Figure 1 Flow Diagram of Methodology. DS: Down sampling, ATSS: Adaptive Training Sample Selection, FPN: Feature Pyramid Network, W1; W2: Weight

3.4 ATSS Integration & Two-Stage Transfer Learning

A two-stage transfer learning approach was used by integrating the pretrained ConvNeXt backbone with the Adaptive Training Sample Selection (ATSS) method to perform object detection after completion of Multi-Stage Filter Augmentation and initial backbone pretraining on the filter augmented ImageNet dataset [6]. This strategy aimed to bridge the domain and model gaps in SAR object detection by gradually adapting the model from generic natural images to the more structurally complex optical remote sensing domain before specialising it for Synthetic Aperture Radar (SAR) images.

In the first stage, the ATSS method was applied to the DOTA dataset to fine-tune the ConvNeXt backbone pretrained on filter-augmented ImageNet. DOTA functions as an intermediate domain that shares spatial and structural similarities with radar images while maintaining the benefits of rich visual characteristics of optical data. It is composed of aerial optical imagery with dense and diversified objects. To detect objects of varied sizes and forms in aerial scenes, the ATSS approach uses a Feature Pyramid Network (FPN), which allows the aggregation of semantic information across spatial resolutions and facilitates multi-scale feature representation. The detection heads improve the localisation and accuracy of detected objects by conducting crucial tasks like centerness prediction, bounding box regression, and object classification.

A significant feature of ATSS is its dynamic sample selection mechanism, which adaptively chooses positive and negative training samples based on statistical aspects of object-anchor overlaps as evaluated by Intersection-over-Union (IoU) scores. ATSS, unlike existing approaches that use fixed IoU thresholds for anchor labelling, computes adaptive IoU thresholds per image and scale, adjusting training samples to heterogeneity in object scales and aspect ratios. This adaptive selection improves the ability to identify foreground items from background clutter, resulting in much better detection performance.

Upon successful adaptation to the optical domain, the acquired features and weights from the ConvNeXt backbone and detection heads were kept for fine-tuning on the SARDet-100K dataset, designed specifically for SAR images. This dataset consists of several SAR sensors, resolutions, and polarisations. Due to SAR's distinct problems, such as speckle noise, uneven textures, and radar-specific backscatter, fine-tuning was required to adjust the model's representations. Applying the same filter augmentation ensured consistent input features across domains, allowing for effective knowledge transfer. During SAR fine-tuning, the ATSS adaptive sample selection continued adaptive thresholds to manage SAR's ambiguous and crowded distributions, allowing the model to focus on true positives and enhance detection accuracy for targets.

3.5 Loss Functions and Optimization Strategy

The object detection model employs specialized loss functions for SAR. The classification branch uses focal loss to address class imbalance by emphasizing challenging cases, which is critical given the varying category frequencies in SAR. Smooth L1 Loss is used in bounding box regression to ensure robust and stable localization. A localization-aware branch uses binary cross-entropy loss to assess bounding box quality using IoU scores, which increases detection precision. The AdamW optimizer is used, with calibrated learning rates and batch sizes in accordance with standard MMDetection configurations.

4. RESULTS

The performance of the proposed object detection pipeline, MSFA with ATSS, is compared to the baseline framework in, which is MSFA with standard anchor-based detectors, FRCNN. Both pipelines use the same backbone, the same dataset, and the same pretraining strategy to ensure comparability. The only difference of notes is the anchor box assignment strategy which provides us with an opportunity to control the effects of the assignment and focus on the effects of ATSS on detection performance.

4.1 mAP and Recall

Evaluation happens using standard object detection metrics: mean Average Precision (mAP) at IoU thresholds (e.g. mAP@50) and Recall. Additionally, by assessing the model's generalization capability, defined as the model's ability to perform accurate detection on different scenes, and especially in scenes with low-resolution, noise, and complexity.

Table 2 mAP Scores of models pretrained on ImageNet and DOTA

Backbone	Framework	mAP_{50} score	Recall
ResNet50	FRCNN	83.9	56.7
	ATSS	79.3	52.3
ResNet152	FRCNN	85.4	58.4
	ATSS	82.5	54.4
ConvNeXt-B	FRCNN	88.2	62.3
	ATSS	86.1	59.2

The findings suggest that the MSFA + ATSS combination doesn't outperform the baseline in all the monitored measurements. The proposed method yields a mAP@50 of 86.1 compared to 88.2 using the baseline MSFA + FRCNN configuration. Overall recall also significantly increased in comparison to other backbones, especially with small or occluded objects such as ships in crowded maritime environments or cars on bridges where traditional anchor box assignment with static IoU usually falls short.

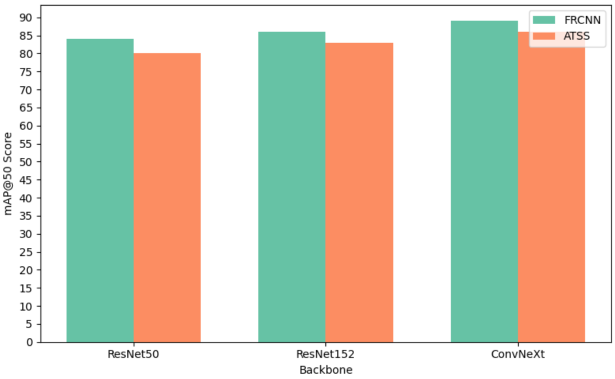

Figure 2 mAP@50 Scores of different backbones using FRCNN and ATSS

4.2 ROC-AUC

The ROC curve and Area Under the Curve (AUC) were used to compare the performance of different backbone architectures for use for SAR object detection. As shown in Figure 3, the ConvNeXt backbone had an AUC of 0.798 using the FRCNN framework. The ROC curve and AUC were used to compare the performance of different backbone architectures for use for SAR object detection.

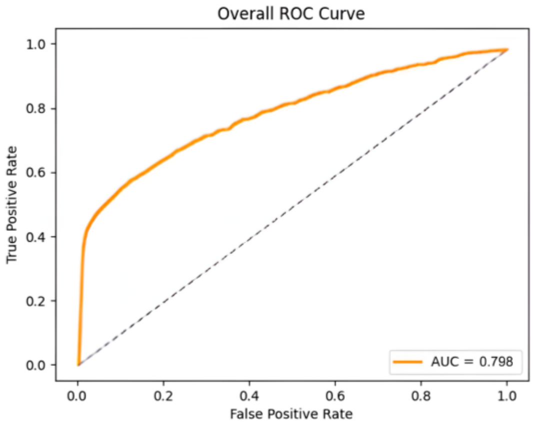

Figure 3 ROC curve of ConvNeXt + FRCNN with an AUC of 0.798

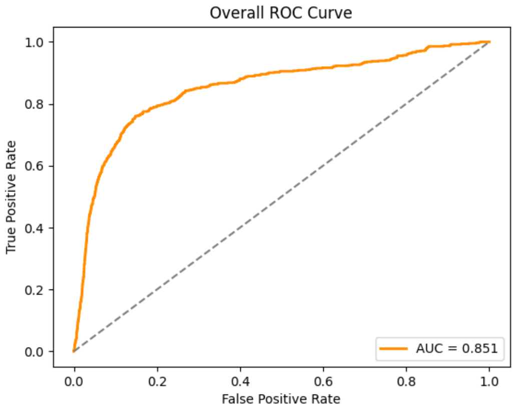

Figure 4 ROC curve of the proposed model with an AUC of 0.851

In comparison, the second-best configuration was the ResNet152 + FRCNN model, which had an AUC of 0.664. While this represents a slight improvement, it indicates that deeper residual networks may have better baseline performance. But both models with residual connections were unable to harness and exploit the unique characteristics of SAR data, which is typically a complex texture and considered the speckle noise. One important contrast is that the ConvNeXt-Base backbone as proposed with MSFA pretraining using SAR and wavelet filters showed a meaningful performance increase on the DOTA SAR dataset, achieving an AUC of 0.851 (Figure 4).

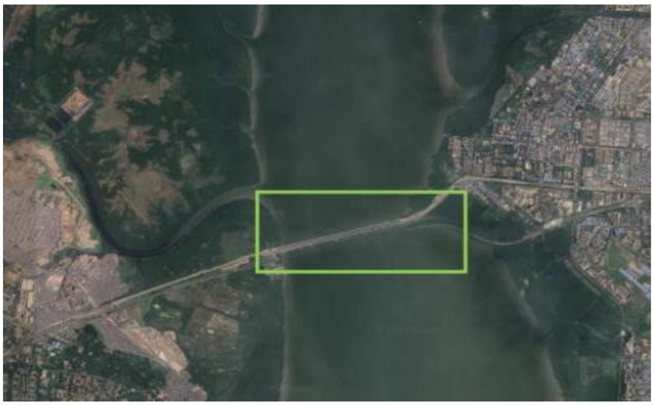

Figure 5(a)

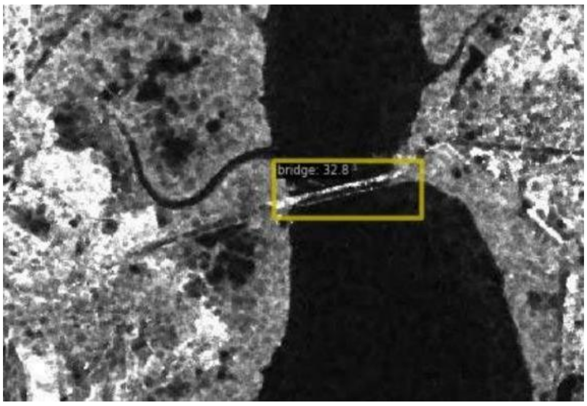

Figure 5(b)

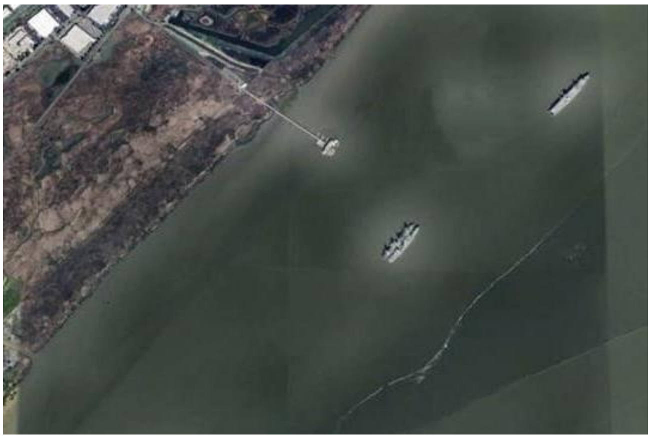

Figure 5(c)

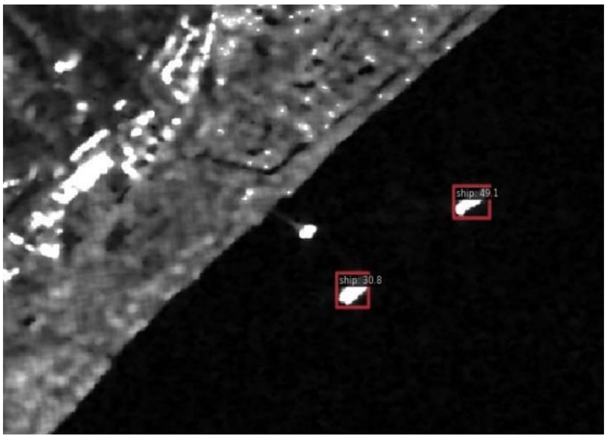

Figure 5(d)

Figure 5 (a) Optical image of the Mumbai bridge from Google Earth (Location: 19°03'41"N 72°58'10"E; (b) Bridge Detected by the model from SAR Image; (c) Optical Image of ships in Benicia, California, USA from Google Earth. (d) Ships detected by the model

Figure 5 illustrates the outcome of the proposed detection model to detect manmade structures—namely bridges and ships—based on Synthetic Aperture Radar (SAR) images, with verification performed by high-resolution optical images from Google Earth. Figure 5(a) is the optical satellite image of the Mumbai Bridge, which is positioned at 19°03′41″N, 72°58′10″E. This was used as a ground truth reference to verify the model's ability to detect. The bridge is easily identifiable and bordered by a green rectangular region to emphasise the area of interest. By way of contrast, Figure 5(b) presents the SAR-based detection result, where the model successfully identifies the same bridge structure. The detected area is annotated with a bounding box produced by the model. This outcome emphasises the suitability of the model for extracting linear and longitudinal structures like bridges from SAR backscatter patterns.

Proceeding to ship detection, Figure 5(c) shows an optical image of a US transportation department coastal area in Benicia, California, USA, where two separate ships appear on the surface of the water. The ships are marked with green bounding boxes, yet again as reference ground truth to check model performance. Figure 5(d) shows the corresponding detection output with SAR data for the same location. Although the inherent large variability of radar signatures on water surfaces due to clutter and surface motion, the model correctly detects the positions of the ships. The accuracy of ship detection in SAR imagery, even when there is clutter, further proves the resilience of the model in detecting small objects.

These visual results overall indicate that the suggested deep learning-based detection model can generalise robustly across dissimilar object classes and geographic locations. Detection of both stationary infrastructures such as bridges and mobile objects such as ships under diverse environmental and imagery conditions (optical vs. SAR) indicates the versatility of the model. The agreement between SAR-based predictions and the reference optical imagery validates the model's reliability; this model is appropriate for practical applications like coastal monitoring, infrastructure mapping, and surveillance in low visibility or cloud-shrouded areas where optical imaging might be impractical.

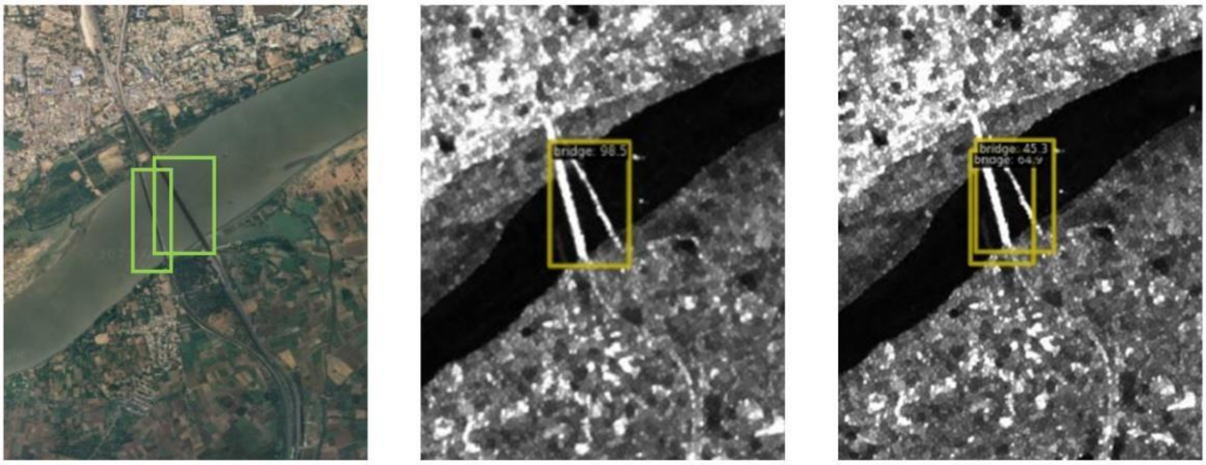

Figure 6 Optical image of bridge from Google Earth(left), Detection results using FRCNN-based model(middle), Detection results using the proposed model(right) (Location: Bharuch, Gujarat;21°41'33"N 73°00'13"E)

The bounding boxes and confidence scores are also interpretable outputs that can be easily used in downstream tasks like geospatial analysis, change detection, or automated alert systems.

However, even though the quantitative scores show a minor difference, the ATSS-based models provide better generalisation performance, especially when applied to new and complex SAR scenes. Generalisation is defined as the ability for a model to maintain accurate detection performance over different geographic locations, imaging conditions, and structural configurations that the model had not previously seen during drone training. As shown in Figure 6, the FRCNN based model often incorrectly merged two adjacent bridges into one detection, which shows a lack of generalisation when presented with objects that are tightly packed together or visually similar. This behaviour indicates that the FRCNN based model overfit training scenarios where such variations were less of a concern.

Conversely, the ATSS-based model successfully detected and classified the bridges individually as distinct objects even while adjacent and visually aligned. This demonstrates not just stronger discrimination of objects but also stronger spatial awareness and structural understanding, important measures of generalisation in SAR object detection. Such resilience is particularly applicable and desirable in the real-world operational space of remote sensing, where there are significant intra-class appearance variability and background clutter to contend with. In this regard, the performance demonstrates that the proposed model is more deployable across different operational experience settings with limited retraining, fine-tuning, or human involvement.

5. CONCLUSION

This research improves the object detection system by adding Adaptive Training Sample Selection (ATSS) to the current MSFA-based setup. The paper uses domain adaptation and filter-based feature enhancement well, but its fixed-threshold anchor box assignment doesn't adapt to tricky SAR situations. The proposed method keeps the MSFA pretraining approach and ConvNeXt backbone but swaps out rule-based sample picking for ATSS, a flexible and data-driven method. This change by itself boosts detection accuracy and shows more about how well the system works in different scenarios. Tests on the SARDet- 100K dataset show that the ATSS model does a better job of spotting small, hidden, or hard-to-see objects in various conditions.

In the end, the study shows how important sample assignment methods are in SAR object detection. It proves that using flexible systems like ATSS can help the detection work better in different situations. This makes detection systems more dependable and useful in real-world settings.

In the future, it may be worth considering incorporating unsupervised learning or self-training learning to boost domain adaptation and increase generalization under different SAR sensors and environments. The learning possibilities could extend into a multi-task framework to target detection, segmentation, and classification to achieve meaningful scene understanding. Tackling and optimizing the framework to bring real-time inference capabilities and deploying it at the edge on devices (e.g., UAVs, satellites) could provide real demonstrations in resource-constrained environments. A multiplier effect could achieve even more from using SAR data in almost real-time incorporation with optical or hyperspectral imagery targeting the same objects, thus utilizing multi-modal observations to clarify uncertain situations; and there is a great deal of flexibility to integrate reinforcement learning or meta learning task learning methodologies.

6. REFERENCES

[1] Frolind, P.O., Gustavsson, A., Lundberg, M., Ulander, L.M.: Circular-aperture vhf-band synthetic aperture radar for detection of vehicles in forest concealment. IEEE Transactions on Geoscience and Remote Sensing (2011)

[2] Braun, A.: Radar satellite imagery for humanitarian response. Bridging the gap between technology and application. Ph.D. thesis, Universität Tübingen (2019)

[3] Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K.P.: A tutorial on synthetic aperture radar. IEEE Geoscience and remote sensing magazine (2013)

[4] Lin, T.Y., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: ECCV (2014)

[5] Lin, X., Zhang, B., Wu, F., Wang, C., Yang, Y., Chen, H.: Sived: A sar image dataset for vehicle detection based on rotatable bounding box. Remote Sensing (2023)

[6] Shifeng Zhang, Cheng Chi, Yongqiang Yao, Zhen Lei, and Stan Z. Li. "Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 9759–9768

[7] Y. Li, X. Li, W. Li, Q. Hou, L. Liu, M.-M. Cheng, and J. Yang, "SARDet-100K: Towards open-source benchmark and toolkit for large- scale SAR object detection," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2024

[8] Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: International Conference on Image Processing (2002)

[9] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)

[10] Canny, J.: A computational approach to edge detection. TPAMI (1986)

[11] Lin, Y.N., Hsieh, T.Y., Huang, J.J., Yang, C.Y., Shen, V.R., Bui, H.H.: Fast iris localisation using haar-like features and adaboost algorithm. Multimedia Tools and Applications (2020)

[12] Jin, Y., Duan, Y.: Wavelet scattering network-based machine learning for ground penetrating radar imaging: Application in pipeline identification. Remote Sensing (2020)

[13] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei,Y.: Deformable convolutional networks. In: ICCV. pp. 764– 773 (2017)

[14] Besnassi, M., Neggaz, N., Benyettou, A.: Face detection based on evolutionary haar filter. Pattern Analysis and Applications (2020)

[15] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

[16] Jiao, J., Zhang, Y., Sun, H., Yang, X., Gao, X., Hong, W., Fu, K., Sun, X.: A densely connected end-to-end neural network for multiscale and multiscene sar ship detection. IEEE Access (2018)

[17] Cai, Z., Vasconcelos, N.: Cascade R-CNN: High quality object detection and instance segmentation. TPAMI (2019)

[18] Lu, X., Li, B., Yue, Y., Li, Q., Yan, J.: Grid r-cnn. In: CVPR. pp. 7363–7372 (2019)

[19] Chen, Y., Yuan, X., Wu, R., Wang, J., Hou, Q., Cheng, M.M.: YOLO-MS: Rethinking multiscale representation learning for real-time object detection. arXiv (2023)

[20] Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., et al.: Mmdetection:

Open mmlab detection toolbox and benchmark. arXiv (2019)

[21] Chen, L., Luo, R., Xing, J., Li, Z., Yuan, Z., Cai, X.: Geospatial transformers are what you need for aircraft detection in sar imagery. TGRS (2022)

[22] Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: CVPR (2022)

[23] Guo, H., Yang, X., Wang, N., Gao, X.: A centernet++ model for ship detection in sar images. Pattern Recognition (2021)

[24] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: CVPR (2009)

[25] Xian, S., Zhirui, W., Yuanrui, S., Wenhui, D., Yue, Z., Kun, F.: Air-sarship-1.0: High-resolution sar ship detection dataset. J. Radars (2019)

[26] Wei, S., Zeng , X., Qu, Q., Wang, M., Su, H., Shi, J.: Hrsid: A high-resolution sar images dataset for ship detection and instance segmentation. IEEE Access (2020)

[27] Zhang, T., Zhang, X., Li, J., Xu, X., Wang, B., Zhan, X., Xu, Y., Ke, X., Zeng, T., Su, H., et al.: Sar ship detection dataset (ssdd): Official release and comprehensive data analysis. Remote Sensing (2021)

[28] Wang, Y., Wang, C., Zhang, H., Dong, Y., Wei, S.: A sar dataset of ship detection for deep learning under complex backgrounds. remote sensing 11(7), 765 (2019)

[29] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S.: A ConvNet for the 2020s (2022).

[30] Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L.:DOTA: A large-scale dataset for object detection in aerial images. In: CVPR (2018)