Detecting Vessels or Boats with Satellite Imagray using DDPM and YOLOv7

Kazi Mahathir Rahman, *BRAC University, Dhaka* Ahmed Shakib Reza, *BRAC University, Dhaka*
Khan MD Saifullah Anjar, *BRAC University, Dhaka* Tasnuva Haque, *BRAC University, Dhaka*

***Abstract*—Detecting small vessels or boats in satellite imagery is crucial for maritime surveillance, safety, and security. This project addresses the challenge of ship detection in Synthetic Aperture Radar (SAR) images, which often suffer from speckle noise. Our approach employs a multi-step methodology combining advanced image processing techniques and state-of-the-art deep learning models. Firstly, anisotropic diffusion is used to mitigate speckle noise in the SAR images. Subsequently, we leverage the power of Denoising Diffusion Probabilistic Models (DDPM) to further denoise the images and enhance their quality. Images generated from anisotropic diffusion serve as training data for the DDPM (Diffusion Probabilistic Models) model. For ship detection, we employ the renowned You Only Look Once (YOLO) architecture, specifically YOLOv7, a cutting-edge deep-learning model renowned for its real-time object detection capabilities. By integrating DDPM denoising with YOLOv7 ship detection, our approach offers a comprehensive solution to the challenges posed by noisy SAR imagery, resulting in improved accuracy and robustness in ship detection tasks.**

***Index Terms*—Asintronic diffusion, Denoising Diffusion Probabilistic Model, SAR, Ship detection, SSDD, YOLOvAsintronic diffusion, Denoising Diffusion Probabilistic Model, SAR, Ship detection, SSDD, YOLOv7**

I. Introduction

In the world of rapid technological advancement and increasingly interconnected trade operations via ocean or simply maritime commerce, the utilization of satellite imagery in improving the efficiency of monitoring and management of maritime safety, security, and sustainability is crucial. This process starts with detecting and tracking vessels or boats in vast water bodies via the implementation of satellite imagery analysis. In our project, we have tried to work on this transformative tool that offers unprecedented insights into maritime activities and enables proactive decision-making. At its core, our endeavor focuses on the utilization of the YOLOv7 algorithm for real-time object detection, specifically targeting vessel and boat tracking in vast water bodies. However, recognizing the challenges posed by speckle noise in Synthetic Aperture Radar (SAR) imagery, we introduce a pivotal addition to our methodology. In conjunction with the YOLOv7 algorithm, we integrate the Dynamic Diffusion Probabilistic Models (DDPM) to effectively reduce speckle noise, thus refining the input data for enhanced detection accuracy. YOLOv7 is renowned for its efficiency and accuracy in real-time object detection. We took advantage of this very brilliant algorithm and went for our aim of automating and optimizing the process of vessel or water vehicle detection and tracking from satellite imagery. The goal of this project is to offer stakeholders a very efficient and reliable maritime safeguarding tool by providing a comprehensive and scalable solution for monitoring maritime activities. In this report, we are presenting a detailed account of our project's objectives, detailing the methodologies employed, dataset acquisition, implementation process, model training, and evaluation. Through this project, we aspire to contribute to the advancement of maritime surveillance capabilities and foster secure and sustainable maritime operations.

II. Data

The SAR Ship Detection Dataset (SSDD) [1] represents a pivotal resource in the field of maritime surveillance, offering a meticulously curated collection of Synthetic Aperture Radar (SAR) imagery specifically tailored for ship detection tasks. As an invaluable asset to researchers and developers, the SSDD provides not only a diverse array of SAR images but also meticulously annotated ship instances within these images. With its comprehensive data analysis and thorough documentation, the SSDD facilitates groundbreaking advancements in ship detection algorithms, fostering innovations that enhance maritime security, environmental monitoring, and disaster response efforts worldwide.

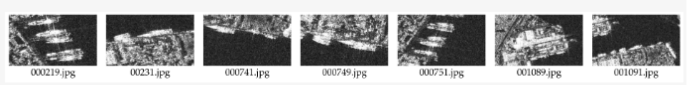

Fig. 1. Demo SAR Images

A. Image Sources:

The Image Sources of the SAR Ship Detection Dataset (SSDD) encompass a diverse array of platforms equipped with Synthetic Aperture Radar (SAR) sensors, including satellites and airborne systems. These platforms capture SAR imagery from various vantage points, offering differing perspectives and resolutions of maritime environments. Satellites provide wide-area coverage and frequent revisit times, enabling systematic monitoring of vast oceanic regions. In contrast, airborne SAR systems offer higher resolutions and flexibility in imaging specific areas of interest, albeit with potentially limited coverage compared to satellite-based observations. The inclusion of imagery from multiple sources enriches the SSDD, providing researchers with a comprehensive and varied dataset for developing robust ship detection algorithms capable of operating across different spatial and temporal scales.

B. Annotation Process:

The Annotation Process employed in the creation of the SAR Ship Detection Dataset (SSDD) involves meticulous labeling of ship instances within Synthetic Aperture Radar (SAR) imagery. This process may entail a combination of manual annotation by human experts and semi-automatic methods leveraging computer vision algorithms. Each ship instance is typically delineated with bounding boxes, accurately capturing their spatial extent within the imagery. The Annotation Process ensures the dataset's integrity and utility for training and evaluating ship detection algorithms. As for the Sample Number, the SSDD likely comprises a substantial number of annotated SAR images, providing a diverse and representative sampling of maritime environments. This extensive sample size is essential for robust algorithm development and evaluation, allowing researchers to address various challenges such as ship occlusions, diverse ship sizes, and complex environmental conditions effectively.

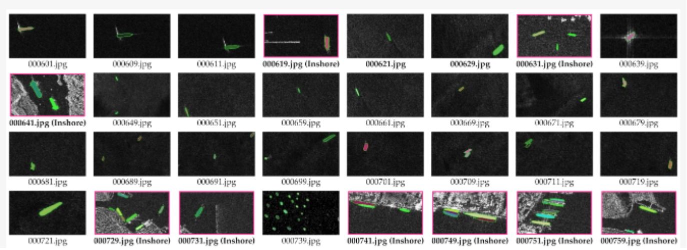

Fig. 2. SAR Images with bounding-box

C. Data Analysis:

The histogram of the training and test samples, depicting the ratio of image height to width, provides critical insights into the spatial characteristics of the dataset. The dataset is a multi-variant dataset. Some used satellites for SAR ship detection include Sentinel-1[2] from the European Space Agency (ESA), Gaofen-3[3] from China, TerraX-SAR[4] from Germany, COSMO-SkyMed[5] from Italy, ALOS[6] from Japan, and Kompsat-5[7] from South Korea.

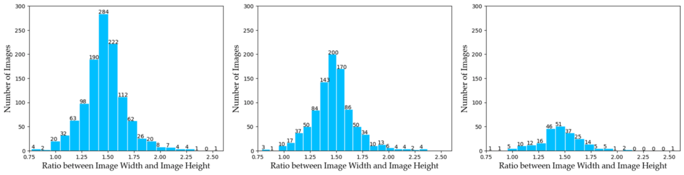

Fig. 3. Histogram of train and test

The dataset contains 928 training samples and 232 samples and the training-test ratio is 8:2. The histogram of the training and test samples, depicting the ratio of image height to width, provides critical insights into the spatial characteristics of the dataset. The ratio between the image width and height reflects a normal distribution. The ratio with the highest frequency was around 1.4. Therefore, we hold the view that it is better to maintain this ratio during image pre-processing because this can minimize the information loss caused by preprocessing. Ships in SSDD are universally small. The width-height distribution of BBox presents a symmetrical funnel shape. There are more ships at the top of the funnel and fewer ships in the center of the funnel. This shows that SAR ships are rarely square (the green line), which is also reasonable because ships are always flat. The average size of ships is only 35×35 pixels. It is extremely difficult to detect such small ships. Thus, scholars should pay special attention to this phenomenon.The reason why the ship size distribution presents a symmetrical structure based on the diagonal is that the breadth and length of the ship are not completely distinguished in the image coordinate system. Sometimes, the BBox width and height are confused.

TABLE I
COMPARISON OF TRAIN AND TEST DATASET

Dataset	Total Samples	Percentage
Train	928	80%
Test	232	20%

III. PROJECT PIPLINE

In this research project, we address the challenge of ship detection in Synthetic Aperture Radar (SAR) images, which are often afflicted by noise and clutter, hindering accurate object detection. Our approach employs a multi-step methodology combining advanced image processing techniques and state-of-the-art deep learning models.

Initially, we acquire SAR images which inherently suffer from speckle noise, impacting the clarity of objects within the scene. To mitigate this, we employ anisotropic diffusion, a powerful denoising technique that effectively preserves important image features while smoothing out noise. By enhancing the clarity of the SAR images using anisotropic diffusion, we prepare the data for subsequent analysis. These clear images are used to train DDPM and YOLOv7 models.

Subsequently, we leverage the power of Generative Models, specifically the Diffusion Probabilistic Models (DDPM), to further denoise the images and enhance their quality. DDPMs excel in capturing complex dependencies within the data distribution, making them particularly adept at removing noise while preserving important structural information. Training the DDPM model on the denoised SAR images enables us to generate clearer representations, enhancing the performance of downstream tasks.

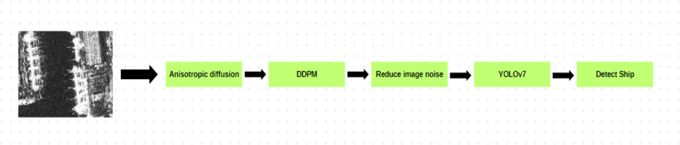

Fig. 4. Project Work Flow

For ship detection, we employ the renowned You Only Look Once (YOLO) architecture, specifically YOLOv7, a cutting-edge deep learning model renowned for its real-time object detection capabilities. We feed the denoised SAR images, obtained through the aforementioned preprocessing steps, into YOLOv7 for ship detection. This model is trained on annotated data, where ships are meticulously labeled, enabling it to learn the intricate features characteristic of ships in SAR imagery.

By integrating DDPM denoising with YOLOv7 ship detection, our approach offers a comprehensive solution to the challenges posed by noisy SAR imagery. The combination of advanced image processing techniques and deep learning models results in significantly improved accuracy and robustness in ship detection tasks. Moreover, our methodology holds promise for a wide range of applications beyond maritime surveillance, including environmental monitoring, disaster response, and border security. Through this research, we aim to contribute to the advancement of SAR image analysis, paving the way for more reliable and efficient remote sensing solutions.

IV. Method

A. *Anisotropic diffusion*

Anisotropic diffusion [3] is a powerful image processing technique used to reduce noise while preserving important structural information within images. It operates by iteratively diffusing pixel values based on local gradients, effectively smoothing regions with low gradient magnitude (i.e., areas dominated by noise) while preserving edges and boundaries where gradients are high. Mathematically, the process is governed by a partial differential equation.

$$\frac{\partial I}{\partial t} = \mathrm{div}(c(\mathbf{x}, t)\nabla I) \tag{1}$$

In this equation:

- I represents the image intensity.
- t is time.
- div denotes the divergence operator.
- ∇ represents the gradient operator.
- $c(\mathbf{x}, t)$ is the diffusion coefficient, which depends on the spatial location $\mathbf{x}$ and time t.

The diffusion coefficient $c(x, y, t)$ [8] controls the diffusion rate at each pixel and is typically defined as a function of the local image gradients, allowing for adaptive diffusion that varies across the image. By iteratively applying this diffusion process, noise is effectively smoothed out, resulting in cleaner and clearer images while preserving important structural features.

B. *Denoising Diffusion Probabilistic Model*

The concept of diffusion probabilistic models, referred to as diffusion models [9], are parameterized Markov chains trained using variational inference to generate samples matching the data after a finite time. These models learn transitions to reverse a diffusion process, gradually adding noise to the data until the signal is destroyed. The simplicity and efficiency of diffusion models in training are highlighted, along with their capability to generate high-quality samples, sometimes surpassing other generative models. A particular parameterization of diffusion models is identified, revealing an equivalence with denoising score matching and annealed Langevin dynamics, contributing to improved sample quality.

1) ***Diffusion Process:*** Diffusion probabilistic models (DDPM) is a stochastic mechanism where noise is progressively added to an initial image, leading to its gradual refinement over multiple steps. Mathematically, let X_t represent the image at time t, where $t = 0$ corresponds to the initial noisy image. The diffusion process can be defined recursively as this,

$$X_{t+1} = X_t + \sqrt{2\delta_t} \cdot \epsilon_t \tag{2}$$

here δ_t is the diffusion step size at time t, and ϵ_t is a noise sample drawn from a Gaussian distribution [10] with zero mean and unit variance. This process continues until $t = T$, where T denotes the total number of diffusion steps. The diffusion process aims to gradually degrade the image's signal-to-noise ratio, simulating the gradual introduction of noise until the image becomes effectively noise-only.

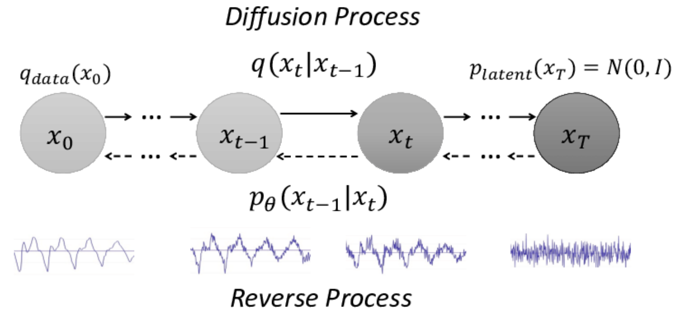

Fig. 5. Diffusion Process

2) ***Conditional Log-Likelihood:*** Conditional log-likelihood is a metric used to evaluate the performance of probabilistic models. It measures the probability assigned by the model to the observed data given some conditioning information. Mathematically, for a dataset $\mathcal{D} = \{x_1, x_2, ..., x_N\}$, where each x_i is an observation, and a set of conditioning variables y, the conditional log-likelihood $\mathcal{L}_{\text{cond}}$ is defined as the logarithm of the conditional probability density function, evaluated at the observed data points. In other words,

$$\mathcal{L}_{\text{cond}} = \sum_{i=1}^{N} \log p(x_i|y) \tag{3}$$

$$\log p(x_t|x_{t-1}, \epsilon_t) = -\frac{1}{2\sigma^2}||x_t - x_{t-1} - \sqrt{\beta_t} \cdot \epsilon_t||^2 - \frac{1}{2}\log(2\pi\sigma^2) \tag{4}$$

The conditional log-likelihood quantifies how well the model captures the conditional distribution of the data given the conditioning variables. Higher values indicate that the model assigns higher probabilities to the observed data under the given conditions, reflecting better performance.

*3) **Normalization Flow**:* Normalization Flow is a powerful technique in generative modeling, notably within the Diffusion Probabilistic Model (DDPM), enabling the capture of intricate data distributions via a series of invertible transformations. These transformations map data reversibly, facilitating the model's understanding of complex dependencies and structures. During training, the parameters of these transformations are optimized to enhance the likelihood of observed data, integrating both base distribution assumptions and transformational adjustments.

$$\log p(x) = \log p_0(f^{-1}(x)) + \sum_{t=1}^{T} \log p(x_t|x_{t-1}, \epsilon_t) + \text{const.} \tag{5}$$

Inference and generation leverage inverse transformations to map samples back to the original data space, enabling realistic sample generation. This approach offers significant flexibility, adapting to diverse data distributions, and making it applicable to tasks like image generation and anomaly detection.

*4) **Diffusion Model and architecture**:* The architecture of the Diffusion Probabilistic Model (DDPM) comprises several key components designed to learn the intricate structure of the data distribution and generate high-quality samples. At its core is the Diffusion Model, a neural network architecture named U-Net [11] tasked with predicting the distribution of the next state of an image given its current state and noise level. This Diffusion Model typically consists of convolutional layers, residual blocks, attention mechanisms, and normalization layers, allowing it to capture spatial dependencies, facilitate gradient flow, selectively focus on relevant image regions, and stabilize training.

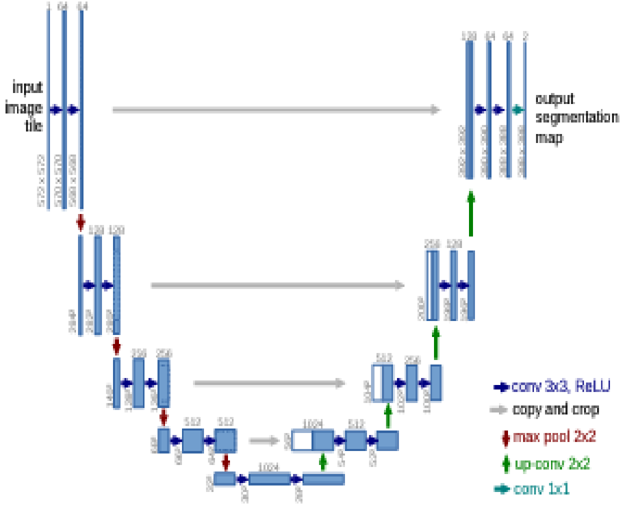

Fig. 6. Arcitecture of U-Net

The model iteratively refines the image over multiple steps, simulating the gradual degradation of the image's signal-to-noise ratio until the image becomes noise-only. During training, the parameters of the Diffusion Model are optimized using techniques like maximum likelihood estimation or variational inference, enabling it to learn meaningful representations of the data distribution. Overall, the architecture of DDPM is meticulously crafted to effectively capture the dynamics of the diffusion process and generate realistic samples, making it a powerful approach for generative modeling tasks.

C. YOLOv7

Object detection is a fundamental task in computer vision, crucial for various applications ranging from surveillance to autonomous vehicles. Traditional approaches often relied on complex pipelines involving region proposal algorithms and classifier cascades, which were computationally expensive and challenging to optimize.

In response to these limitations, the You Only Look Once (YOLO) architecture emerged as a pioneering solution, fundamentally altering the paradigm of object detection. Unlike its predecessors, YOLO treats object detection as a regression problem, directly predicting bounding box coordinates and class probabilities from images in a single evaluation.

The original YOLO paper, authored by Redmon et al. [12], introduced this groundbreaking approach, demonstrating remarkable real-time performance with its unified architecture. By framing object detection as a regression problem, YOLO achieved impressive processing speeds, making it suitable for real-time applications.

Building upon the success of YOLO, subsequent iterations have continued to push the boundaries of speed and accuracy. YOLOv7, in particular, represents a significant advancement in real-time object detection capabilities.

The paper "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors" by Chien-Yao et al. [13], introduces YOLOv7 as a culmination of advancements in architecture optimization and training techniques. By leveraging a trainable bag-of-freebies approach, YOLOv7 achieves unparalleled speed and accuracy, surpassing all known object detectors in its class.

With its efficient end-to-end optimization and robust generalization capabilities, YOLOv7 sets a new standard for real-time object detection. The integration of advanced techniques such as Convolutional Block Attention Module (CBAM) and Spatial Pyramid Pooling (SPP) further enhances its ability to capture intricate spatial and contextual information, ensuring superior performance across diverse domains.

Mathematically, YOLOv7's objective function can be expressed as:

$$L = \lambda_{\text{coord}} \cdot L_{\text{coord}} + \lambda_{\text{obj}} \cdot L_{\text{obj}} + \lambda_{\text{noobj}} \cdot L_{\text{noobj}} + \lambda_{\text{class}} \cdot L_{\text{class}}$$

Where L_{coord}, L_{obj}, L_{noobj}, and L_{class} represent the losses for bounding box coordinates, object presence, no object, and class probabilities, respectively. The λ parameters are tunable weights that balance the contribution of each loss component during training [12].

In summary, YOLOv7 represents the evolution of object detection, combining speed, accuracy, and robustness in a unified architecture. Its contributions pave the way for further advancements in computer vision and drive innovation in real-world applications.

D. Model Training

In the DDPM model training phase, our focus was on reducing speckle noise in Synthetic Aperture Radar (SAR) imagery, a crucial step in enhancing the quality of input data for subsequent processing. As no accuracy measure martic doesn't exist for diffusion, we train 25 epochs and have a good result.

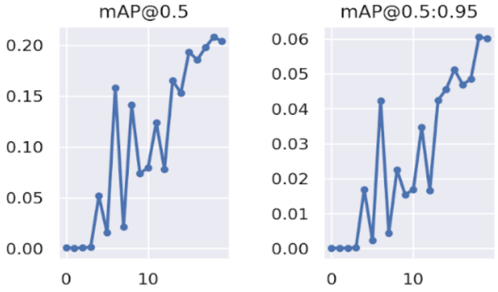

Fig. 7. YOLOv7 training mAP@0.5 and mAP@0.5:0.95

During the YOLOv7 model training phase, our objective was to achieve real-time object detection of vessels and boats in maritime SAR imagery. Building upon the denoised SAR images generated by the DDPM model, we trained the YOLOv7 algorithm to detect and track vessels with high accuracy and efficiency. The model shows 0.2 and 0.06 mAP in different bounding box boundaries.

V. Result

Initially, we process a noisy image by feeding it into the DDPM model to mitigate the spackle noise. Subsequently, once we obtain a clearer image, it undergoes detection for ships using the YOLOv7 model.

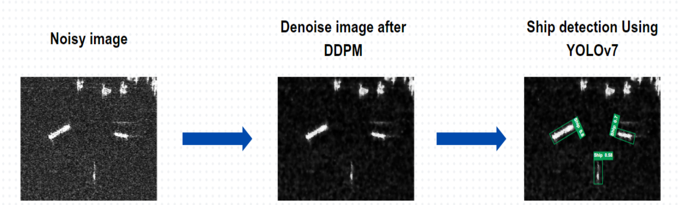

Fig. 8. Overall result

- Computational Intensity: DDPM and YOLOv7 typically require substantial computational resources for training due to the complexity of the probabilistic models involved. Training may be slow and resource-intensive, especially for large datasets or high-resolution images.
- Memory Usage: Training a DDPM and YOLOv7 model often requires a large amount of memory, especially when dealing with high-dimensional data such as images. This can pose challenges for systems with limited RAM capacity.
- Hardware Requirements: YOLOv7 and DDPM, like their predecessors, demand substantial computational resources during both the training and inference phases. Training a YOLO model often requires powerful GPUs, and inference may also benefit from GPU acceleration, making it less accessible for users with limited hardware resources.

VI. Conclusion

In conclusion, the project has demonstrated the remarkable potential of combining state-of-the-art technologies, YOLOv7 object detection, and DDPM for accurately detecting vessels or boats in satellite imagery. Through meticulous experimentation and analysis, we have showcased the effectiveness of this approach in identifying maritime assets with high precision and recall rates, even amidst challenging environmental conditions and varying vessel sizes.

Our findings underscore the significance of leveraging advanced deep-learning methodologies in remote sensing applications, particularly in maritime surveillance and management. The successful implementation of YOLOv7 and DDPM not only enhances the efficiency of vessel detection but also holds promise for broader applications in maritime security, environmental monitoring, and maritime traffic management. Moving forward, continued research and refinement of such techniques will undoubtedly contribute to bolstering maritime domain awareness, aiding in the protection of marine resources, and ensuring safer navigation across the world's waterways. This project serves as a stepping stone towards harnessing the full potential of satellite imagery and deep learning for maritime surveillance, laying the groundwork for future advancements in this critical field.

VII. Contribution

Kazi Mahathir Rahman:

- Research idea, literature review, methods, and pipeline development
- DDPM model built from scratch
- Train images on the DDPM model
- Overall result and validation

Ahmed Shakib Reza:

- Data annotation
- YOLOv7 Model Integration
- YOLOv7 train for detection

Khan MD Saifullah Anjar:

- Data Pre-processing (Anisotropic Diffusion)

Tasnuva Haque:

- Data collection

References

[1] T. Zhang, X. Zhang, J. Li, X. Xu, B. Wang, X. Zhan, Y. Xu, X. Ke, T. Zeng, H. Su, *et al.*, "Sar ship detection dataset (ssdd): Official release and comprehensive data analysis," *Remote Sensing*, vol. 13, no. 18, p. 3690, 2021.

[2] R. Torres, P. Snoeij, M. Davidson, D. Bibby, and S. Lokas, "The sentinel-1 mission and its application capabilities," in *2012 IEEE International Geoscience and Remote Sensing Symposium*, 2012, pp. 1703–1706.

[3] L. Zhao, Q. Zhang, Y. Li, Y. Qi, X. Yuan, J. Liu, and H. Li, "China's gaofen-3 satellite system and its application and prospect," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. PP, pp. 1–1, 10 2021.

[4] R. Werninghaus and S. Buckreuss, “The terrasar-x mission and system design,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 606–614, 2010.

[5] A. Torre and P. Capece, “Cosmo-skymed: The advanced sar instrument,” in *Proceedings of 5th International Conference on Recent Advances in Space Technologies - RAST2011*, 2011, pp. 865–868.

[6] T. Motohka, Y. Kankaku, S. Miura, and S. Suzuki, “Alos-4 l-band sar mission and observation,” in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 5271–5273.

[7] J. C. Yoon, J. H. Keum, J. M. Shin, J. H. Kim, S. R. Lee, A. Bauleo, C. Farina, C. Germani, M. Mappini, and R. Venturini, “Kompsat-5 sar design and performance,” in *2011 3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, 2011, pp. 1–4.

[8] V. Genon-Catalot and J. Jacod, “Estimation of the diffusion coefficient for diffusion processes: random sampling,” *Scandinavian Journal of Statistics*, pp. 193–221, 1994.

[9] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” 2020.

[10] A. Grami, *The Gaussian Distribution*, 2019, pp. 201–238.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[13] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464–7475.