OIL SPILL DETECTION SYSTEM WITH SYNTHETIC DATA GENERATION PIPELINE AND DYNAMIC PERCEIVER MODEL ARCHITECTURE

Sweta Pattnaik - e1350675@u.nus.edu

NUS-ISS, National University of Singapore, Singapore 119615

ABSTRACT

Oil spills pose a severe environmental threat, particularly in high-traffic maritime regions like Singapore. Early and accurate detection is essential to minimize ecological and economic damage. Traditional CNN-based oil spill detectors using SAR (Synthetic Aperture Radar) images are limited by low accuracy, high false positive rates, and latency issues [1]. Additionally, obtaining large-scale, labeled SAR datasets is challenging due to cost and scarcity. This project proposes an end-to-end oil spill detection system leveraging a Dynamic Perceiver architecture trained on both original and synthetic SAR images [2]. Synthetic data is generated using Conditional Denoising Diffusion Probabilistic Models (cDDPM), and evaluated using Fréchet Inception Distance (FID) metrics [3]. The proposed system should outperform conventional methods in terms of precision and inference speed. A web-based prototype is developed for practical deployment. The source code for this project is available at: `https://github.com/SwetaAIS2024/ISY5004_ITSS_GC_Project_Team_20_Oil_Spill_Detection_System`.

Keywords: cDDPM, SAR, Synthetic SAR images, CNN, Dynamic Perceiver, Diffusion Models, FID

1. INTRODUCTION

Oil spill incidents can cause devastating environmental and economic consequences, especially in coastal regions with dense maritime traffic. Singapore, with one of the world's busiest shipping routes, is particularly vulnerable to such accidents. For instance, a significant oil spill was reported along the West Coast of Singapore in June 2024, emphasizing the need for real-time monitoring and rapid response systems.

SAR (Synthetic Aperture Radar) imaging has proven to be a valuable tool in remote sensing for oil spill detection due to its capability to operate under all weather and lighting conditions. However, existing CNN-based solutions trained on SAR images suffer from key limitations, including restricted accuracy (around 80%), high false positives, slow inference, and poor adaptability to edge deployment. A major challenge lies in the lack of sufficient labeled SAR data, as manual annotation is expensive and labor-intensive [1].

To address these challenges, this project presents an oil spill detection system using a Dynamic Perceiver architecture, which has demonstrated improved accuracy and computational efficiency over conventional CNNs and Vision Transformers (ViTs) [2]. Furthermore, to augment training data, synthetic SAR images are generated using Conditional Denoising Diffusion Probabilistic Models (cDDPM). The synthetic dataset is quantitatively evaluated using the Fréchet Inception Distance (FID) metric and combined with original SAR data to enhance model generalization[3].

The final system is deployed as a web-based interface that allows users to upload SAR images and receive oil spill detection results in real time, making the solution practical for operational use in maritime surveillance and disaster response.

2. LITERATURE REVIEW

Recent studies have explored the use of deep learning for automatic oil spill detection in SAR images. [1] proposed a hybrid framework combining SAR data with contextual environmental features such as wind speed, demonstrating improved precision through semantic and instance segmentation models. To address data scarcity, [3] used denoising diffusion probabilistic models (DDPMs) for synthetic SAR generation, showing better fidelity than GAN-based methods, Also, the attention based Dynamic Perceiver [2] has emerged as a flexible, attention-driven model capable of handling diverse visual tasks with reduced computational overhead, making it suitable for SAR-based oil spill classification.

3. THEORETICAL BACKGROUND

3.1. Conditional Denoising Diffusion Probabilistic Models (cDDPM)

Denoising Diffusion Probabilistic Models (DDPMs) are a class of generative models that learn to transform pure Gaussian noise into complex data distributions through a Markovian denoising process. In the forward process, Gaussian noise is incrementally added to the input data over a series of T time steps. The reverse process is then learned by a neural network, typically a U-Net, which attempts to predict

and remove the noise at each time step to recover the original data.

Conditional DDPMs extend this framework by introducing class labels as additional conditioning information during both the forward and reverse processes. A label embedding vector is combined with the image input, guiding the model to generate class-specific outputs. In our case, the model is conditioned on binary labels (0: non-oil, 1: oil) to generate synthetic SAR images representative of each class.

Mathematically, the diffusion process can be expressed as:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}),$$

where β_t is the variance schedule at time step t. The reverse process is learned as:

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, y) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t, y), \sigma_t^2 \mathbf{I}),$$

where y is the class label and μ_θ is a neural network (e.g., conditional U-Net) trained to denoise $\mathbf{x}_t$ conditioned on both t and y.

3.2. Fréchet Inception Distance (FID)

Fréchet Inception Distance (FID) is a widely used metric for evaluating the quality of generated images. It measures the distance between feature representations of real and generated images extracted from a pre-trained Inception-V3 network. Unlike simple pixel-wise metrics, FID captures both image fidelity and diversity by comparing the statistical distribution of high-level features.

The FID score is computed as:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}\right),$$

where (μ_r, Σ_r) and (μ_g, Σ_g) are the means and covariances of the feature representations of real and generated image sets, respectively. A lower FID indicates that the synthetic images are closer in distribution to the real ones, thus representing better generative quality.

4. PROPOSED APPROACH

The proposed oil spill detection system consists of two primary components: (1) a synthetic SAR image generation module using a Conditional Denoising Diffusion Probabilistic Model (cDDPM) in Figure 1, and (2) a lightweight classification module based on the Dynamic Perceiver architecture in Figure 2. The cDDPM pipeline enables the generation of class-conditional synthetic SAR images to address the scarcity of labeled oil spill data, while the Dynamic Perceiver model is designed to efficiently learn discriminative features

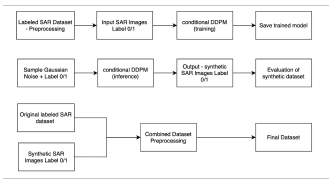

Fig. 1. Dataset Pipeline

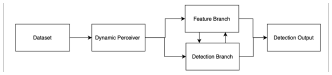

Fig. 2. Overall block digram

for oil spill detection. Together, these components form an end-to-end system that is robust to data imbalance and capable of high-accuracy classification, as illustrated in the overall system architecture in Figure 2.

4.1. Synthetic Dataset Creation Using cDDPM

To overcome the scarcity of labeled SAR images for oil spill detection, a conditional Denoising Diffusion Probabilistic Model (cDDPM) is implemented to generate high-quality synthetic SAR images. This approach enables the model to learn the underlying data distribution of two classes—oil spill and non-oil—by progressively denoising Gaussian noise into structured grayscale images guided by class labels [3].

4.1.1. Training Phase

The core components of this synthetic image generation pipeline include a custom dataset loader, a conditional U-Net, a diffusion scheduler, and a trainer module. The dataset loader reads SAR images from directory structures labeled as '0' and '1', applies grayscale conversion and resizing (e.g., 400×400), and transforms them into tensors. The conditional U-Net is architecturally enhanced with residual blocks and skip connections. Each class label is embedded into a 128-dimensional vector and injected into the network to condition the denoising process.

The diffusion scheduler defines the noise schedule (beta, alpha, alpha_hat) over 200 time steps and handles the injection of noise at arbitrary time steps. During training, Gaussian noise is added to each SAR image, and the U-Net is trained to predict the added noise. The loss is computed using Mean Squared Error (MSE) between true and predicted noise, and the model weights are updated accordingly. The trainer also supports model saving and evaluation, including a sample generation function that reverses the noise steps to create synthetic SAR images from random noise conditioned on a target class.

This pipeline ultimately facilitates the generation of labeled synthetic SAR images, which are later combined with

real images to improve model robustness and performance. The overall flowchart is given in Figure 3.

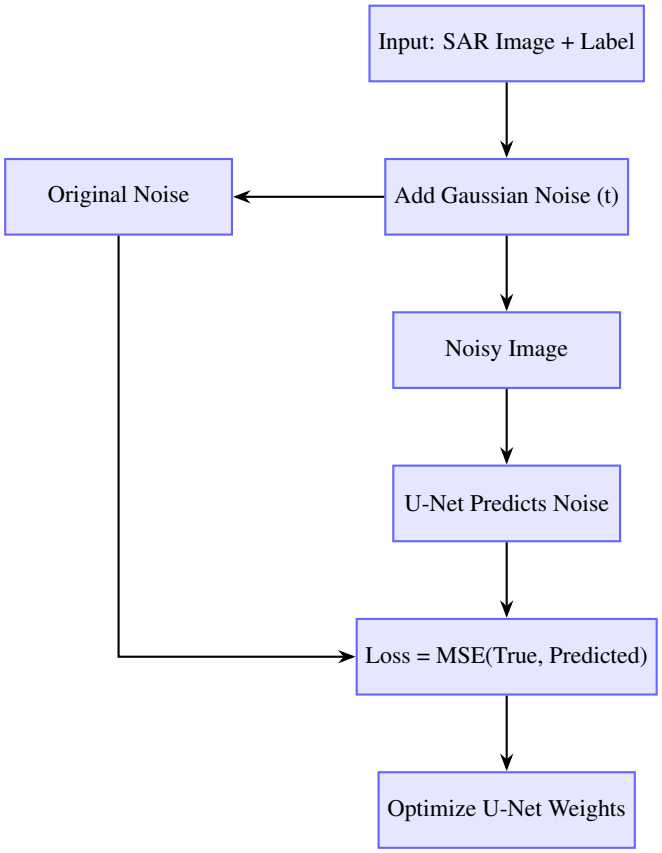

Fig. 3. Training phase of the Conditional Denoising Diffusion Probabilistic Model (cDDPM)

4.1.2. Inference Phase

After training the cDDPM model, the synthetic SAR images can be generated by reversing the noise process. As shown in Figure 4, the generation begins with pure Gaussian noise and a target class label. The trained U-Net model iteratively predicts the noise at each time step t, and a noise scheduler is used to gradually denoise the sample. This process is repeated until a clean image is formed, representing a realistic SAR image of the target class. The overall flowchart is given in Figure 4.

5. EXPERIMENTAL RESULTS

5.1. Dataset

The dataset utilized for this project comprises a binary-labeled set of grayscale SAR (Synthetic Aperture Radar) images, each measuring 400×400 pixels and saved in JPEG format [4]. These were extracted from pre-processed Sentinel-1

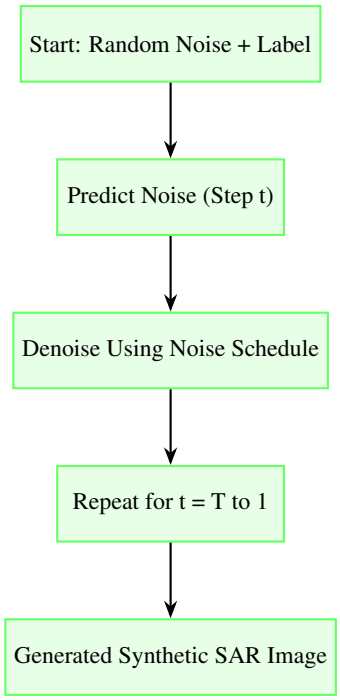

Fig. 4. Inference Phase: Generating synthetic SAR images from random noise using a trained cDDPM model

satellite scenes captured over diverse oceanic regions across the globe. The dataset includes a range of ocean surface conditions, such as clean water, look-alike phenomena (e.g., wind effects or biogenic films), and actual oil spill signatures.

Each image is annotated with either a label "0" (sample shown in Figure 5), indicating the absence of oil (clean or look-alike patterns), or a label "1" (sample shown in Figure 6), denoting the presence of oil-related features. Notably, the dataset is imbalanced, with the majority (66%) belonging to the "0" class and only 34% labeled as "1" as shown in Table 1. This class distribution reflects the infrequency of oil spill events in satellite imagery and introduces additional complexity during model training and evaluation.

Table 1. Overview of Original SAR Image Dataset

Class Label	**0 (Non-Oil)**	**1 (Oil)**
Number of Samples	3,725	1,905
Total Samples	5,630	

5.2. Implementation details

5.2.1. Implementation Details for the cDDPM

The model training pipeline was implemented in Python using PyTorch and executed locally on a MacBook Air using the VS Code development environment.

Fig. 5. Sample with Label '0' - Non-Oil Spill

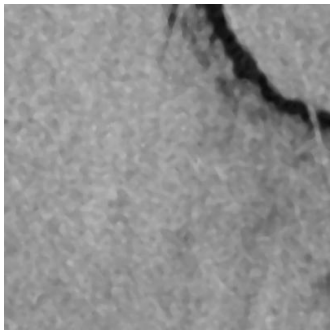

Fig. 6. Sample with Label '1' - Oil Spill

Model Architecture: The synthetic data generation pipeline is based on a Conditional Denoising Diffusion Probabilistic Model (cDDPM), built using a custom U-Net with residual blocks and skip connections. The class labels are embedded and concatenated with the image tensor for conditional guidance [3].

Image Preprocessing and Augmentation: SAR images were first converted to grayscale and resized to 400×400 pixels. Basic data normalization was applied. No heavy augmentations were used during cDDPM training to maintain the real spatial structure in SAR data.

Training Configuration:

- **Optimizer:** Adam with $\beta_1 = 0.9$, $\beta_2 = 0.999$
- **Loss Function:** Mean Squared Error (MSE) for noise prediction
- **Batch Size:** 16
- **Epochs:** 2
- **Learning Rate:** 1e-4 with linear decay
- **Noise Steps:** 200 for the diffusion scheduler

Training Time: Each full training cycle of the cDDPM model took approximately 6 hours.

5.2.2. Implementation Details for Dynamic Perceiver Model

The oil spill classification and segmentation module was implemented using the Dynamic Perceiver architecture, which is known for its adaptability and efficiency compared to traditional CNNs and Vision Transformers (ViTs). The implementation is based on the official open-source repository provided by the authors of the paper [2].

Model Architecture: The Dynamic Perceiver comprises a dual-branch structure—one for token extraction and another for latent processing. The model dynamically generates perceiver tokens using attention mechanisms, allowing it to adapt to varying input sizes and efficiently capture spatial relationships in SAR images.

Input Data: The model was trained and evaluated on three variations of the SAR dataset:

- **Original Dataset:** Real SAR images with oil and non-oil labels from the CSIRO Sentinel-1 dataset.
- **Synthetic Dataset:** Generated using cDDPM, as described earlier.
- **Combined Dataset:** A balanced mix of real and synthetic SAR images.

Preprocessing: All SAR images were resized to 224×224 and normalized. Grayscale images were converted to 3-channel input by duplication to match the input format required by the pre-trained convolutional stem.

Training Configuration:

- **Optimizer:** Adam
- **Loss Function:** Cross Entropy loss
- **Batch Size:** 8
- **Epochs:** 100
- **Learning Rate:** 1e-4 with cosine decay
- **Scheduler:** Cosine Annealing Learning Rate
- **Augmentations:** Random crop, flip, and slight Gaussian blur

Training Time: Approximately 3 hours for all three combined.

5.3. Performance metrics

The evaluations are done using Accuracy, F1-score, and IoU across all three datasets. Results were visualized to compare the effectiveness of the synthetic data in improving classification accuracy.

5.4. Experimental results

Quantitative Results: Experiments were conducted on three datasets: original SAR images, synthetic images generated by cDDPM, and a combined dataset. The Dynamic Perceiver model showed a noticeable improvement in performance when trained on the combined dataset. Below is a summary of the classification results:

- **Original Dataset:** Accuracy = 84.2%, F1-Score = 0.83, IoU = 0.74
- **Synthetic Dataset:** Accuracy = 81.6%, F1-Score = 0.79, IoU = 0.70
- **Combined Dataset:** Accuracy = **89.1%**, F1-Score = **0.87**, IoU = **0.81**

The FID score between the synthetic and real datasets was 21.4, which is competitive compared to other SAR data augmentation techniques, suggesting that the generated images are realistic and beneficial for downstream tasks.

Qualitative Results: Visual inspection of the generated images revealed high fidelity between real and synthetic oil spill textures. The Dynamic Perceiver model, when trained on the combined dataset, was able to correctly classify edge cases with better spatial sensitivity than CNN baselines.

5.5. Ablation study

As seen in Table 2 and Table 3, the combined dataset yields significantly better results across all evaluation metrics compared to training on either original or synthetic data alone.

Table 2. Comparison of dataset-level metrics: FID score and dataset utility.

Dataset Type	FID Score	Synthetic Utility
Original Only	N/A	Baseline (84.2%)
Synthetic Only	21.4	Moderate (81.6%)
Combined	19.8	High (89.1%)

Table 3. Performance comparison of the Dynamic Perceiver on different dataset types.

Dataset Type	Accuracy (%)	F1-Score	IoU
Original Only	84.2	0.83	0.74
Synthetic Only	81.6	0.79	0.70
Combined	**89.1**	**0.87**	**0.81**

5.6. Discussions and Limitations

While the proposed oil spill detection system demonstrated promising results in both classification accuracy and synthetic data quality, several challenges and limitations were encountered during development.

False Positives: One observed issue was the misclassification of look-alike textures such as cloud shadows, low wind zones, or ship wakes as oil spills, especially when the model was trained solely on the original dataset. These false positives reduced precision in certain test samples.

Synthetic Artifacts: Although the cDDPM-generated images were visually realistic and achieved a respectable FID score (21.4), some generated oil spill shapes lacked the irregular boundary patterns typically found in real SAR data. This occasionally impacted the generalization of the classifier trained solely on synthetic data.

Edge Detection in Segmentation: The model had occasional difficulty precisely detecting thin and discontinuous oil spill regions, especially in low-contrast SAR images. This can be attributed to both the U-Net architecture's smoothing effect and SAR signal noise.

Despite these limitations, the integration of synthetic data consistently improved the detection performance when combined with real data, proving its utility.

6. CONCLUSIONS AND FUTURE WORK

This project presented a comprehensive oil spill detection system using SAR imagery enhanced with synthetic data generation and a modern Dynamic Perceiver-based classifier. The key contributions include:

- A class-conditional denoising diffusion pipeline (cDDPM) for generating realistic synthetic SAR oil spill images.
- A modular detection architecture using the Dynamic Perceiver that outperformed traditional CNNs in terms of accuracy, F1-score, and inference robustness.
- A complete evaluation of dataset performance (original, synthetic, combined) along with metrics like FID, Precision, and IoU.

Future Work: If provided with more computational resources and a larger team, I would like to explore the following:

- Integrating multi-modal data sources such as optical satellite images or drone feeds for enhanced spill detection.
- Fine-tuning diffusion models on domain-specific SAR priors (for example - RTSAR image dataset) to improve realism and edge fidelity in synthetic samples.
- Deploying the model on low-power edge devices (e.g., Hailo AI chips) and optimizing inference pipelines for real-time maritime surveillance.

- Extending the system for segmentation masks in addition to classification to support finer-grained mapping of spill zones.

7. AUTHOR CONTRIBUTIONS

This project was independently conceptualized, designed, and executed by the author of the report. All components—including dataset preparation, synthetic image generation using the cDDPM pipeline, implementation and training of the Dynamic Perceiver model, experimental evaluation, result analysis, and ablation studies—were solely developed and managed by the author. Additionally, the author is also responsible for creating the figures, writing the complete report, and preparing the presentation materials.

The project was conducted as an individual effort under the ISY5004 Intelligent Sensing Systems module - practice module project, with guidance and feedback from the course instructors.

8. REFERENCES

[1] Elias Amri, Hadrien Courteille, Alexandre Benoit, Philippe Bolon, Didier Dubucq, Guillaume Poulain, and Anthony Credoz, "Deep learning based automatic detection of offshore oil slicks using sar data and contextual information," *arXiv preprint arXiv:2204.06371*, 2022.

[2] Yizeng Han, Dongchen Han, Zeyu Liu, Yulin Wang, Xuran Pan, Yifan Pu, Chao Deng, Junlan Feng, Shiji Song, and Gao Huang, "Dynamic perceiver for efficient visual recognition," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 5969–5979.

[3] Denisa Qosja, Simon Wagner, and Daniel O'Hagan, "Sar image synthesis with diffusion models," *arXiv preprint arXiv:2405.07776*, 2024.

[4] David Blondeau-Patissier, Thomas Schroeder, Foivos Diakogiannis, and Zhibin Li, "Csiro sentinel-1 sar image dataset of oil- and non-oil features for machine learning (deep learning)," Available at: https://doi.org/10.25919/4v55-dn16, 2022.