SafeSea: Synthetic Data Generation for Adverse & Low Probability Maritime Conditions

Martin Tran[1], Jordan Shipard[1], Hermawan Mulyono[2], Arnold Wiliem[1,2], and Clinton Fookes[1]

[1]Signal Processing, Artificial Intelligence and Vision Technologies (SAIVT), Queensland University of Technology, Australia
[2]Sentient Vision Systems, Australia
{minhtuan.tran@connect, jordan.shipard@hdr, c.fookes@}qut.edu.au, {arnoldw, hermawanm}@sentientvision.com

Abstract

High-quality training data is essential for enhancing the robustness of object detection models. Within the maritime domain, obtaining a diverse real image dataset is particularly challenging due to the difficulty of capturing sea images with the presence of maritime objects , especially in stormy conditions. These challenges arise due to resource limitations, in addition to the unpredictable appearance of maritime objects. Nevertheless, acquiring data from stormy conditions is essential for training effective maritime detection models, particularly for search and rescue, where real-world conditions can be unpredictable. In this work, we introduce SafeSea, which is a stepping stone towards transforming actual sea images with various Sea State backgrounds while retaining maritime objects. Compared to existing generative methods such as Stable Diffusion Inpainting [27], this approach reduces the time and effort required to create synthetic datasets for training maritime object detection models. The proposed method uses two automated filters to only pass generated images that meet the criteria. In particular, these filters will first classify the sea condition according to its Sea State level and then it will check whether the objects from the input image are still preserved. This method enabled the creation of the SafeSea dataset, offering diverse weather condition backgrounds to supplement the training of maritime models. Lastly, we observed that a maritime object detection model faced challenges in detecting objects in stormy sea backgrounds, emphasizing the impact of weather conditions on detection accuracy. The code, and dataset are available at https://github.com/martin-3240/SafeSea.
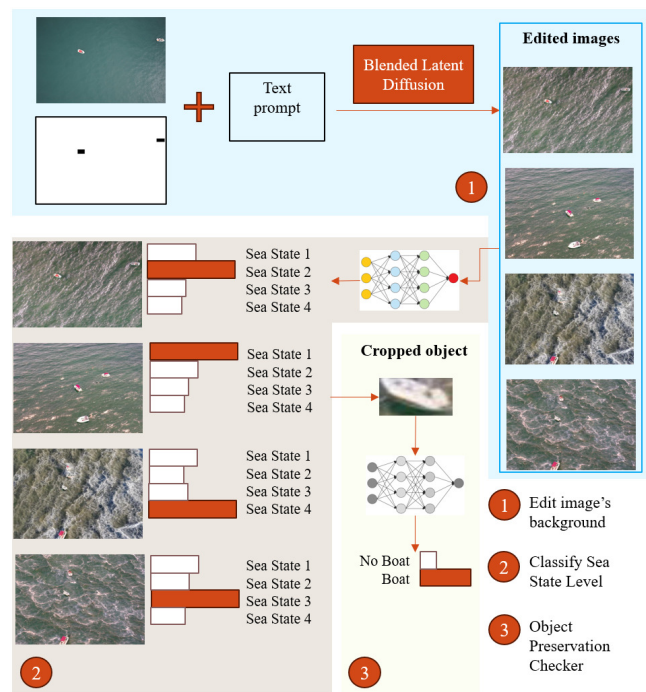
Figure 1. Diagram summarizing our method (SafeSea). Original maritime images' sea background is edited with a mask and text description using Blended Latent Diffusion [1]. The edited images are then classified into 4 Sea State categories before their marine objects (boats) are cropped according to the ground truth bounding box and checked for preservation.

1. Introduction

Climate change is increasing the likelihood of extreme weather events, such as large storms [6], which can be destructive, especially for humans and boats in stormy waters. Quick and accurate rescue efforts are crucial in such situations as death can occur in as little as 50 minutes [31]. Au-

tomated disaster response in the ocean is a growing area that uses advanced technology to help search and rescue in extreme weather events. Leveraging autonomous systems, remote sensing technologies, artificial intelligence, and real-time data analysis, automated disaster response aims to minimise risks, and accelerate response times. However, detecting entities in need of rescue during severe conditions in the ocean is often challenging. Therefore, there is a need for automated disaster responses that can operate effectively in these harsh conditions and can precisely locate objects, vessels, and people requiring assistance.

Developing robust models to automatically detect objects in the sea during extreme weather conditions is a challenging task. The challenge primarily arises from the difficulty of obtaining high-quality data in such chaotic and unpredictable situations. Adverse weather conditions, particularly during storms, pose a safety risk, preventing the deployment of cameras in critical areas for data collection. Furthermore, the wild nature of disaster scenes often stretches already limited resources, including personnel, equipment, and communication channels, thus constraining the capacity for data acquisition [11]. Due to the lack of datasets for training models to effectively detect maritime objects in the ocean during severe weather conditions, generating synthetic datasets to train and test object detection models can be highly beneficial.

By harnessing recent generative models such as Stable Diffusion (SD) [27], and DALL-E [25], it is now possible to generate realistic images based on textual descriptions. However, generating synthetic images solely based on textual descriptions experiences limitations in specifying object location, size, and type within the ocean scene, restricting the replication of realistic scenarios. To overcome this challenge, a more practical approach is to utilize real sea images with marine objects and performing transformations on the sea background.

Unfortunately, when applying the Stable Diffusion model to edit the real sea image's backgrounds, our experiments reveal a significant limitation in using text prompts to precisely control the desired Sea State. The edited background randomly appears not according to the description, so it would require checking manually to ensure the image's quality. Consequently, this process becomes time-consuming as it requires manual control to achieve a visually satisfactory Sea State. Furthermore, we found that existing image editing methods often replace or excessively modify the objects in the edited images, making the images unusable. Therefore, relying solely on synthetic image generation techniques to create a dataset with diverse Sea State Levels proves to be challenging, as it demands a substantial amount of time for generation and manual verification.

To tackle this issue, we propose a method for editing images of calm ocean scenes into stormy ocean scenes, focusing on UAV-view images. We replace the original calm ocean with an ocean environment corresponding to standard definitions from a Bureau of Meteorology as depicted in Table 1, all while attempting to retain the maritime objects in the images [1]. This method shows as a proof-of-concept to develop a simple-yet-effective approach, capable of automatically screening out poor-quality edited images with overly edited preserved objects.

Our work takes an input image and use an image generation method to perform background transformation. Recent work proposed the Blended Latent Diffusion [1] which demonstrates significantly improved results over the Stable Diffusion model in preserving foreground objects while also translating the background. Therefore, we employed Blended Latent Diffusion to modify images with object masking. Due to the stochastic nature of diffusion [30], we find that prompts alone can be unreliable for producing images according to the sea state definitions. Thus we apply the Sea State Classifier which classifies the transformed image into one of the sea state definitions. Then, the Object Preservation Checker is applied to evaluate whether the transformed image still preserve the objects from the input image.

We leverage the SeaDroneSee dataset [34] for experimentation and validation. To quantitatively evaluate the effectiveness of our approach, we compare it against other image editing methods, such as Stable Diffusion Inpainting [27]. Furthermore, we conducted object detection tests using YoloV5 model pre-trained on calm sea state images [20] to assess the impact of various sea states on object recognisability in edited images. Our findings reveal that the pre-trained object detection model struggles to identify objects in increasingly stormy conditions. This suggests that the model's performance is less effective when it encounters previously unseen stormy sea backgrounds, making it more challenging to detect objects in rough sea surface conditions.

Contributions - We list our contributions as follows:

1. We propose a simple-yet-effective method for modifying the sea state level of the ocean in maritime images while preserving the objects of interest, enhancing their utility for various applications;

2. We construct and propose a synthetic dataset, the Safe-Sea dataset, enabling the training of models to accommodate diverse weather conditions;

3. We conduct an evaluation of the SafeSea dataset with YoloV5 to assess the model's performance under varying weather conditions.

We continue our paper as follows. Section 2 presents an overview of prior work including maritime datasets and diffusion models. In Section 3, we define our problem and the goal we want to achieve before Section 4 outlines our

---

[1]In this work we confine ourselves to only retaining the boats and reserving the other marine objects for future work.

proposed SafeSea method designed to accomplish this goal. The description of the SafeSea dataset are presented in Section 5. Section 6 presents our experiment and discussion. We conclude the paper and discuss about limitations and future work in Sections 7 and Section 8.

2. Related Works

Sea Datasets - Numerous public maritime datasets facilitate the training of marine object detection models. Notable examples include the VAIS dataset [35], offering over 1,000 RGB and infrared image pairs, showcasing various ship types. The IPATCH dataset [21] records realistic maritime piracy scenarios, while the SeaShips dataset [28] features ship images from inland waterways. The 'Singapore Maritime Dataset' [23] captures marine objects using onshore platforms and vessels, providing diverse perspectives. Additionally, the Seagull dataset [26] and Maritime SATellite Imagery (MASATI) dataset [9] offer aerial images. The SeaDronesSee dataset [34] contains over 54,000 frames with around 400,000 instances. Despite their strengths, these datasets often focus on specific objects like ships and boats, and often the objects of interest are relatively large compared to the image size in certain datasets. Furthermore, most of the images are taken under good weather conditions.

Synthetic Datasets - Extensive research explores methodologies employing synthetic images for training object detection models. Noteworthy contributions include Peng et al. [22] emphasizing the refinement of synthetic object backgrounds for improved detection reliability. The use of powerful game engines like Unity [16] and Unreal Engine [7] is prominent, as demonstrated by Becktor et al. [3] in the context of spacecraft guidance and control. Unreal Engine 4 [7] has proven valuable in autonomous driving, Dosovitskiy et al. [5], and maritime image generation, as utilized by Becktor et al. [2]. Kiefer et al. [17] analyzed maritime and terrestrial images, incorporating real and synthetic data from the Grand Theft Auto V (GTAV) simulation platform [10]. Xiaomin et al. [18] introduced SeeDroneSim, utilizing the Blender game engine [4] for UAV-based maritime images. Airbus [15] released a 2018 dataset of 40,000 satellite images designed for ship detection using synthetic aperture radar (SAR) technology. While existing datasets focus on synthetic objects of interest, our work concentrates on generating diverse environmental conditions based on real data.

Diffusion Models - Diffusion models have been widely employed for image transformation in various contexts. Trabucco, Doherty et al. [32] employed pre-trained text-to-image diffusion models for semantic-based image modification. Shin et al. [29] utilized Stable Diffusion for image generation using the Textual Inversion [8] method. Ron et al. proposed the Null-text Inversion method [19], em-

Sea State	Definition
1	The water exhibits a gentle ripple, devoid of breaking waves, featuring a low swell of short to average length occasionally.
2	Slight waves breaking, with smooth waves on the water surface
3	Mildly increased waves, leading to some rock buoys and causing minor disturbances for small craft
4	The sea takes on a furrowed appearance, characterized by moderate waves

Table 1. Sea States descriptions as defined by the ABS

ploying prompt-to-prompt text editing for image denoising. Rombach et al. [27] introduced Stable Diffusion Inpainting for image inpainting using masks and a latent text-to-image diffusion model. Omri et al. [1] accelerated the transformation process with a lower-dimensional latent space. Our selection of the Blended Latent Diffusion approach [1] is based on its demonstrated superior results in editing image backgrounds.

Presently available real-image maritime datasets have limitations, including fixed camera positions and a restricted variety of marine objects. Furthermore, they often lack diversity in representing various weather conditions reflecting on the sea background. This lack of diversity can restrain the development of a high-quality dataset for training deep-learning models. To address this challenge, numerous studies have been conducted to generate high-quality synthetic images capable of replicating real-world scenarios. These synthetic images can be produced by leveraging game engines such as Unity [16] and Unreal Engine [7] or by employing diffusion models to modify real images[1, 8, 12, 27, 29, 32]. While using game engines can be resource-intensive, editing images with diffusion models offers a simpler approach. Unfortunately, the editing process via diffusion models is still time consuming and labour intensive due to imprecise control that specifies which part of the image that needs to be edited. This works proposes a proof-of-concept to enable automation in the editing process which significantly reduces the time and labour whilst maintaining the quality of the edited images.

3. Problem Definition

Here we technically define our problem. Starting from an image X containing a set of objects, $\mathcal{B} = \{b_i\}_{i=1}^N$, where N is the number of objects, and b_i is the i-th object bounding box. We first define the foreground f_x of image X as all areas of $\mathcal{B}$, subsequently, we define the background bg_x as the inverse of all areas of $\mathcal{B}$ in X. Initially, bg_x is a calm ocean environment. We wish to replace bg_x such that it corresponds to a particular sea state, SS, where SS ranges

from 1 to 4 and are defined in Table 1, with examples shown in Figure 6. We only aim to replace bg_x, and aim to retain the number of N as was present in $\boldsymbol{X}$. We refer to the resulting image as $\boldsymbol{Y}$. Specifically, let f_y and bg_y as the foreground, and the background of $\boldsymbol{Y}$, respectively. The goal is to have $bg_y \neq bg_x$, and $f_y \approx f_x$.

4. Proposed SafeSea method

The overall diagram of the propose method is depicted in Figure 1 and the pseudo code is presented in Algorithm 1. The method has three main components: (1) the image generation module which transform the background of an input image; (2) Sea State Level classifier; and (3) the Object Preservation Checker. Once an image is generated, its quality is automatically assessed by using (2) and (3). Specifically, the Sea State Level classifier determines the sea state level of the generated image, and the Object Preservation Checker ensures the generated image still contains the objects from the input image.

The following section first discusses the image generation module which is powered by a diffusion model. Then, the Sea State Level classifier and the Object Preservation Checker modules are presented.

4.1. Image Generation Module

Diffusion models [13] are capable of being trained with guiding input channels, enabling them to perform conditional image generation such as creating synthesizing visual content based on textual descriptions [24]. Beyond synthesis, diffusion models are versatile image editors, allowing for targeted modifications. In this process, noise is added to the original image and subsequently denoised in response to an optional prompt, which describes the new image. Using masking techniques, these models can carry out semantic editing, selectively altering specific regions within the images while leaving others intact. Various approaches exist for utilizing Stable Diffusion in image editing, which can be broadly categorized into two main groups: those that employ pre-specified masks and those that mask images based on provided text descriptions.

In this work, we applied Blended Latent Diffusion method [1] to edit original sea images. Blended Latent Diffusion provides a versatile approach to image editing, utilizing masks alongside guided text prompts. In a nutshell, it provides a versatile approach to image editing, utilizing masks alongside guided text prompts. To alter an image, a mask of identical dimension to that image is applied to designate the region to be modified, while guided text instructions define how the edited area should appear. The intended outcome is an image in which the masked region undergoes alterations while the unmasked portion remains unaltered. In the transformation process, the original image is encoded into a latent space, introducing a controlled level

---

ALGORITHM 1

Pseudocode for generating synthetic images of the sea with marine objects from real images as proposed in SafeSea method. The real image has its sea background edited using Blended Latent Diffusion [1], then its Sea State level is provided by the Sea State Classifier. Finally, the original objects are checked if they are retained in the edited image, which lead to the decision of saving the image if at least one object is preserved.

Input: Real sea images with marine objects (Rs), matching mask images (Ms) with ground true bounding boxes of marine objects are masked as black, text description of background (P)

Output: Transformed images with edited background reflected different sea condition and the marine objects are preserved

```
1:  for each X in Rs do
2:       M ← Corresponding mask image from Ms
3:       Y ← Blended_Latent_Diffusion(X, M, P)        ▷
    Edited image
4:       SS_E ← Sea_State_Classifier(Y)   ▷ Find the edited
    image's Sea State
5:       E_R ← Y                    ▷ Resize the edited image
6:       Cs ← E_R      ▷ Crop objects from the edited image
7:       for each C in Cs do
8:            if Object_Preservation_Checker(C) = boat then
9:                 Save Y
10:                break
11:           end if
12:      end for
13: end for
```

---

of noise. Additionally, the mask is downsampled within this latent space. At each stage of this process, the noisy latent image is subjected to denoising and is regarded as the foreground to be blended with the background elements. We observe that the method demonstrates promise in generating images with different ocean backgrounds whilst maintaining a considerable level of object preservation, in which the object can be recognized visually after the background edition process.

4.2. Sea State Classifier

To provide information about Sea State level in the edited image's background, we employed a Sea State classifier. We selected four distinct Sea States for our study, namely Sea States 1, 2, 3, and 4, as these are the only Sea States for which datasets are publicly available. The Sea State definitions are shown in Table 1 .It is important to note that without recorded weather conditions at the time of image capture, visually classifying sea images into different Sea States is a challenging task. Leveraging the Manzoor-

Umair Sea State Image Dataset (MU-SSiD) [33], we trained a DenseNet [14] to categorize images into the four Sea State categories. The trained model achieved an accuracy rate of 71% against the testing dataset.

4.3. Object Preservation

To evaluate whether an object remains preserved following image transformation, we developed an object preservation checker. This checker's primary function is to identify objects that are no longer recognizable within the provided ground truth bounding box. To do this, we train a binary classifier on a dataset consiting of two classes: "boat" and "not boat." The "boat" class contains images of cropped boats sourced from the ground truth SeaDrone-See dataset [34], complemented by their augmented versions, which include flipping and blurring. In contrast, the 'not boat' class comprises images from the negative class extracted from the Boat-MNIST dataset [34], as well as crops from randomly selected backgrounds from synthetic images. Additionally, it includes crops intentionally containing small portions of boat objects from the edited images.These crops are generated by following the object ground truth bounding box, ensuring that they contain only one-fourth of the bounding box's area with the rest outside. Using the dataset with only horizontal flip data augmentation for the boat class. We trained a DenseNet [14] model with a training batch size of 32, a fixed learning rate of 1e-5 without decay, and utilized the Adam optimization algorithm during the training process. Overall, the model achieves the accuracy of 74.86% against testing dataset, including crops of boat objects in real images and non-boat crops from real and edited images.

Subsequently, based on the given ground truth bounding box information, we extracted boat objects from the edited images for evaluation using the trained checker. We conducted random visual checks on the preserved boat objects to evaluate the model's performance. While there are occasional misclassifications of objects as boats that do not visually resemble boats, the model generally achieves satisfactory accuracy in detecting non-boat objects, in which it can pick up cropped objects that resemble boats and filter out non-boat crops according to the visual checks with the accuracy of 69.45%. Examples of the cropped boat objects are illustrated in Figure 2.

5. SafeSea dataset

The SafeSea dataset is created using the SafeSea method, involving the transformation of 300 calm ocean background images originally sourced from the 'SeaDroneSee' dataset [34]. All edited images were resized to match the dimensions of their respective originals. The SafeSea method produces 69,694 images images. These are then classified into one of the sea state levels. The distribution of these im-

Figure 2. Examples of boat object crops from edited images. The crops are taken based on the ground truth bounding boxes provided from the source images in the 'SeaDroneSea' dataset [34].

Sea State Level 1   Sea State Level 2

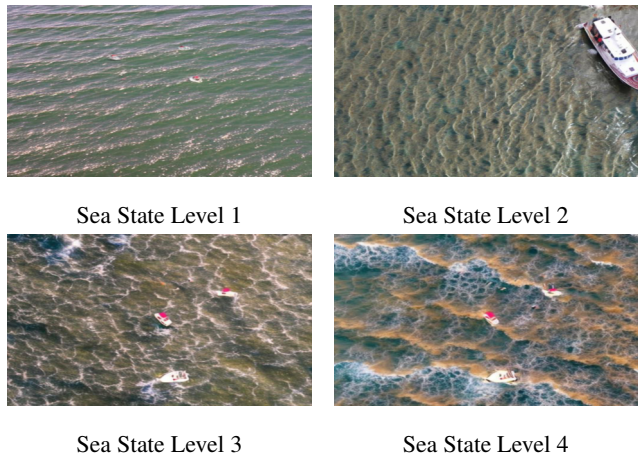Sea State Level 3   Sea State Level 4

Figure 3. Examples of images in the SafeSea dataset. There is one example representing each Sea State level in the dataset.

ages across different sea state levels is detailed in Table 3. Notably, Sea State levels 3 and 1 exhibit the most and the least number of images, boasting 45,066 and 2,087 images respectively. In the intermediate Sea State 2 category, there are 19,390 images, while Sea State 4 includes 3,151 images. For a visual representation, a selection of dataset examples is provided in Figure 3.

6. Experiment

This section is divided into two parts. The first part evaluates the proposed SafeSea method efficacy in filtering compared with the other methods such as the vanilla Stable Diffusion Inpainting. Then, we study the performance of YoloV5 detection model on the proposed SafeSea dataset.

6.1. SafeSea method evaluation

We first describe the experiment setup and then present the results afterwards.

6.1.1 Experiment setup

Evaluation protocol - We generate 100 images from each evaluated method. Then, each image is manually checked by humans and categorized into good quality of bad quality group. A good quality image is defined as the following rules:

- The background should contain either island, ocean or cloud. For instance if the background contains unexpected objects such as a fridge, then the image is deemed low quality.
- The background should look realistic
- All the objects should look visually acceptable. At least one boat is preserved in the image.

Each method is then compared by looking at its percentage of generating good quality images.

Methods - The proposed SafeSea method is compared against two baselines: (1) the vanilla Stable Diffusion inpainting [27]; and (2) the vanilla Blended Latent Diffusion [1]. Details for each baseline parameters is presented as follow.

Stable Diffusion inpainting (SD Inpainting) - We use Stable Diffusion v1-4 [27] with batch size of one. The masked image is derived from the groundtruth bounding boxes. The method is then fed with the following prompts to generate the images.

- Sea State 1: Aerial image of the sea's surface. The water is gently rippled with no waves breaking. Canon EOS R3, Nikon d850 400mm, Canon DSLR, lens 300mm, 4K, HD.
- Sea State 2: Aerial image of the sea's surface. There are slight waves breaking with smooth wave on surface. Canon EOS R3, Nikon d850 400mm, Canon DSLR, lens 300mm, 4K, HD.
- Sea State 3: Aerial image of the sea's surface. Mild Waves are slight causing rock buoys and small craft. Canon EOS R3, Nikon d850 400mm, Canon DSLR, lens 300mm, 4K, HD.
- Sea State 4: Aerial image of the sea's surface. The water has furrowed appearance with moderate waves. Canon EOS R3, Nikon d850 400mm, Canon DSLR, lens 300mm, 4K, HD.

Blended Latent Diffusion (BLD) - We use Stable Diffusion v2-1-base [27] with batch size of ten. Similar to the method above, we use the derived mask image from ground truth bounding boxes. We observe that when editing the sea background of an image to various Sea State Level, the reliance on prompts alone is insufficient. Therefore, to simplify the generation issue with prompt we only use one prompt in our experiments, which is "Aerial image of sea's surface. Canon EOS R3, Nikon d850 400mm, Canon DSLR, lens 300mm, 4K, HD". The prompt allows generation of images with varying sea state levels with different generation seeds.

SafeSea (proposed) - We utilize Blended Latent Diffusion with the same parameters as above to edit original images. Then we use the sea state level classifier to determine the sea state level of the generated images. The Object Preservation Checker ensures objects are preserved. The generated image will be preserved if it preserves at least one object. Same as the other methods, the mask image is derived from the groundtruth bounding boxes provided by the SeaDroneSee dataset [20]. Note that this dataset contains several classes, but in this experiment we only aim to preserve the boat as boats have much larger object size.

6.1.2 Results

Table 2 presents the comparison results. The results suggest that the proposed SafeSea outperforms the baselines. As expected the BLD produces better image quality than the SD Inpainting as also shown in [1]

It is clear that SafeSea outperforms the other methods, as it generates more good images with a realistic sea background and maintains object integrity more effectively.

	Good images (in %)
SD Inpainting [27]	26.5% ± 5.94
BLD [1]	52.94% ± 8.25
SafeSea (proposed)	**63.59%** ± 2.76

Table 2. Comparison between SD Inpainting, BLD and SafeSea when we manually check quality of the 100 generated images from each method. The good image rate computes the percentage of good images passed by manually check by humans. High good image rate suggests the method works better for the task.

While the result of Stable Diffusion Inpainting appear to achieve realism, many objects are blended into the background, or overly edited. Additionally, there are other inherent issues, such as other objects being inserted into the edited image. It appears as though the objects of interest may be inadvertently duplicated and dispersed into the background, as shown in figure 4. Although some edited background is of considerable quality, many of objects of interest are not retained. Additionally, the edited images may not be suitable for use due to the introduction of irrelevant modified objects and unexpected background.

When employing Blended Latent Diffusion, the issue of objects of interest being uncontrollably spread is mitigated, as demonstrated in Figure 5. Notably, the boat objects are effectively preserved, and the edited background exhibits acceptable quality. Nevertheless, it is worth mentioning that

Original   Edited

Original   Edited

Original   Edited

Figure 4. Failure edited images using SD Inpainting. It suggests that this method can edit the image's sea background, however, the objects are not retained well, and several irrelevant objects are introduced unintentionally.

the generated sea background does not possess the same vividness observed in Stable Diffusion Inpainting. This disparity may be affected by the utilisation of different trained Diffusion models in these approaches. Future work will investigate other image generation methods that have similar preservation properties to the Blended Latent Diffusion whilst producing more vivid background.

To further confirm the above result, we present the percentage of images produced by SD Inpainting and BLD retained after applying our SafeSea filter. Results in Table 4 suggest that SD Inpainting produces much lower quality than BLD as significant amount of images are filtered.

	SS1	SS2	SS3	SS4
Generated	2,336	25,114	65,275	4,275
Filtered	2,087	19,390	45,066	3,151

Table 3. Number of images generated before and after applying object preservation checker. 97,000 images are generated from 300 sourced images and then filter out the ones that do not retain any object.

6.2. Applying YoloV5 on the SafeSea dataset

We employed the SafeSea dataset for assessing a pre-trained YoloV5 object detection model's ability in detecting 'boat' objects across various Sea State levels.

Original   Edited

Original   Edited

Figure 5. Example of good edited images using Blended Latent Diffusion. The image's sea background is edited while the objects are preserved.

Sea State Level 1   Sea State Level 2
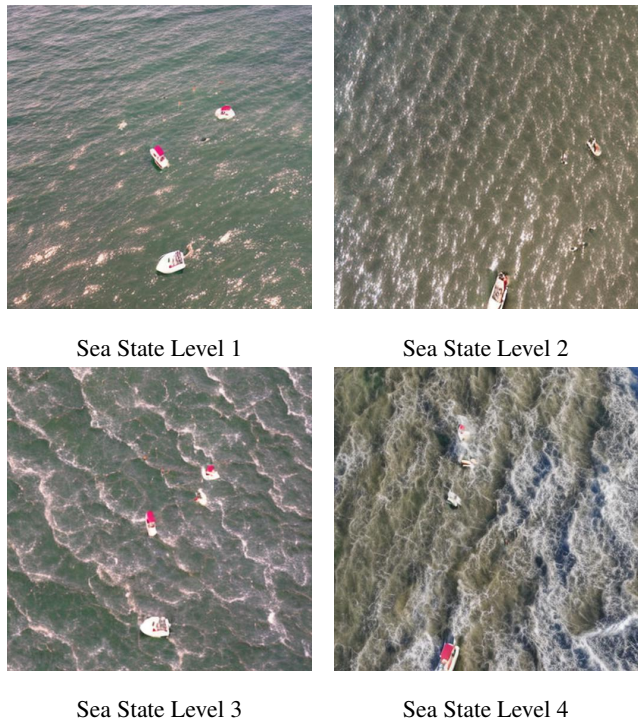
Sea State Level 3   Sea State Level 4

Figure 6. Examples of edited images belong to different classified sea state levels. The Sea State level information is provided to these images after the editing process. The sea surface becomes more dynamic with waves and whitecaps as the Sea State Level increases.

6.2.1 Experiment setup

We run the YoloV5 with default parameters on the SafeSea dataset images. the Mean Average Precision (mAP) for the

	SD Inpainting [27]	BLD [1]
Image passing rate (in %)	8.16%	**71.85%**

Table 4. Comparison between SD Inpainting and BLD when we apply our SafeSea filters. The Image passing rate computes the percentage of images passed by our filters. High passing rate suggests better quality images according to our filters.

four sea state level is then calculated.

6.2.2 Results

Figure 7 presents the result. Notably, there is a noticeable decrease in mAP values from Sea State 1 to Sea State 4, both for IoU of 0.5 and the range of 0.5 to 0.95.

Given that the Sea State Level background is not entirely controlled within the transformation process, the distribution of images among the sea state categories is determined by the Sea State Classifier. The results indicate a significantly higher number of images classified as Sea State 3 of 45,066; while Sea State 1 comprises the lowest number with 2,087 images. It is important to note that each image originally contains a varying number of objects, both before and after the editing process. Additionally, object size plays a considerable role in the editing process; we have observed that larger objects tend to be better preserved in comparison to smaller ones. This, in turn, has a direct impact on the object detection confidence scores, which subsequently affect the mAP scores. In general, we observe that the pre-trained model tends to be more struggle when detecting objects in images classified with higher sea state levels. However, it is crucial to acknowledge that several other factors also influence the results, such as the number of objects being preserved and the sizes of the original objects.
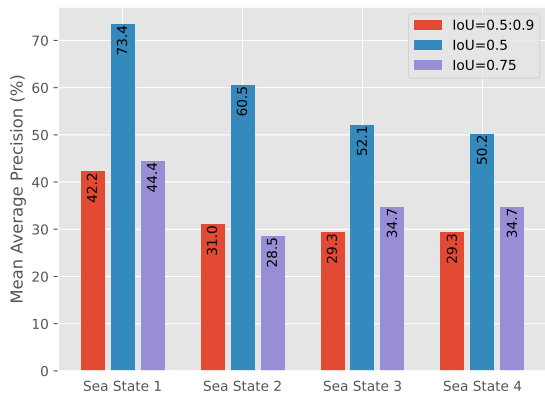

Figure 7. mAP Comparison for Different Sea State Levels. It shows that the mAP values decrease from Sea State 1 to Sea State 4 for both IoU of 0.5 and from 0.5 to 0.95.

7. Conclusion

In this work, we introduce SafeSea, a proof-of-concept for generating synthetic maritime images by modifying real images, where the original sea background is transformed to simulate various sea surface conditions, corresponding to different Sea States. Our method capitalizes on the capabilities of Blended Latent Diffusion [1] to manipulate images. Subsequently, these modified images are categorized into distinct Sea State Levels, ranging from 1 to 4. Moreover, the original objects within these images are scrutinized to ensure their preservation throughout the editing process. Employing this technique, we have created the SafeSea dataset, which includes maritime images featuring marine objects set against diverse Sea State backgrounds by utilizing the 'SeaDroneSea' dataset [34]. Additionally, we have observed that stormy sea backgrounds can impact the performance of the YoloV5 object detection model.

8. Limitation and Future Work

The current SeaSafe method, as proposed, exhibits certain limitations that necessitate attention in our future work. Primarily, it lacks control over the generated sea background during the editing process, limiting the diversity of realistic backgrounds. Furthermore, the diffusion model employed by BLD [1] has constraints in generating realistic wave and whitecap patterns. Additionally, smaller objects such as swimmers are ignored in the image quality evaluation. Our forthcoming efforts will concentrate on optimizing the image editing process to elevate the overall quality of generated images. Exploring alternative image editing methods is also on the agenda to enhance the image generation module. Additionally, addressing the control over the insertion of irrelevant objects during editing is crucial, as it can significantly impact object detection models. Future work will specifically tackle the introduction of unexpected objects, mitigating their potential impact on object detection models. Simultaneously, improvements are planned for both the Sea State Classifier and Object Detection Checker to elevate their performance. Furthermore, we aim to implement an additional filter to exclude generated images that do not align with the desired Sea State Level criteria. These enhancements collectively constitute our roadmap for refining the SeaSafe method in subsequent stages of development. Future work will also delve into the scalability of SafeSea on larger datasets and in real-world scenarios beyond the SeaDronesSee dataset with the exploration of how the object detector performs against the unedited SeaDronesSee dataset.

9. Acknowledgement

References

[1] Omri Avrahami, Ohad Fried, and Dani Lischinski. Blended latent diffusion. *ACM Transactions on Graphics*, 2023. 1, 2, 3, 4, 6, 8

[2] Jonathan Binner Becktor, Frederik Emil Thorsson Saabye Schöller, Evangelos Boukas, Mogens Blanke, and Lazaros Nalpantidis. Bolstering maritime object detection with synthetic data. In *IFAC*. Elsevier, 2022. 3

[3] Jonathan Becktor, William Seto, Aditya Deole, Saptarshi Bandyopadhyay, Niyousha Rahimi, Shahriar Talebi, Mehran Mesbahi, and Amir Rahmani. Robust vision-based multi-spacecraft guidance navigation and control using cnn-based pose estimation. In *AERO*, 2022. 3

[4] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. 3

[5] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In Sergey Levine, Vincent Vanhoucke, and Ken Goldberg, editors, *ICML*. PMLR, 2017. 3

[6] US — EPA. Climate change indicators: Weather and climate. https://www.epa.gov/climate-indicators/weather-climate, 2023. [Online; accessed 06-October-2023]. 1

[7] Epic Games. Unreal engine. 3

[8] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. In *ICLR*, 2022. 3

[9] Antonio-Javier Gallego, Antonio Pertusa, and Pablo Gil. Automatic ship classification from optical aerial images with convolutional neural networks. *Remote Sensing*, 2018. 3

[10] Rockstar Games. Grand theft auto v. 3

[11] Emily Heaslip. Why marine weather forecasts are so inaccurate - and how to improve them. 2

[12] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-or. Prompt-to-prompt image editing with cross-attention control. In *ICLR*, 2023. 3

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020. 4

[14] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *CVPR*, 2017. 5

[15] inversion Jeff Faudi, Martin. Airbus ship detection challenge. Kaggle, 2018. 3

[16] Arthur Juliani, Vincent-Pierre Berges, Ervin Teng, Andrew Cohen, Jonathan Harper, Chris Elion, Chris Goy, Yuan Gao, Hunter Henry, Marwan Mattar, and Danny Lange. Unity: A general platform for intelligent agents. ArXiv, 2020. 3

[17] Benjamin Kiefer, David Ott, and Andreas Zell. Leveraging synthetic data in object detection on unmanned aerial vehicles. In *ICPR*, 2022. 3

[18] Xiaomin Lin, Cheng Liu, Allen Pattillo, Miao Yu, and Yiannis Aloimonous. Seadronesim: Simulation of aerial images for detection of objects above water. In *WACV*, 2023. 3

[19] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *CVPR*, 2023. 3

[20] ntnu. seadronessee-odv2-rebalanced 2x2-train-val-1x1-test dataset. https://universe.roboflow.com/ntnu-2wibj/seadronessee-odv2-rebalanced-2x2-train-val-1x1-test, 2023. visited on 2023-10-06. 2, 6

[21] Luis Patino, Tom Cane, Alain Vallee, and James Ferryman. Pets 2016: Dataset and challenge. In *CVPRW*, 2016. 3

[22] Xingchao Peng, Baochen Sun, Karim Ali, and Kate Saenko. Learning deep object detectors from 3d models. In *ICCV*, 2015. 3

[23] Dilip K. Prasad, Deepu Rajan, Lily Rachmawati, Eshan Rajabally, and Chai Quek. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Transactions on Intelligent Transportation Systems*. 3

[24] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. ArXiv, 2022. 4

[25] Aditya Ramesh, Mikhail Pavlov, Gabriel Soh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *ICML*, 2021. 2

[26] Ricardo Ribeiro, Gonçalo Cruz, Jorge Matos, and Alexandre Bernardino. A data set for airborne maritime surveillance environments. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019. 3

[27] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 1, 2, 3, 6, 8

[28] Zhenfeng Shao, Wenjing Wu, Zhongyuan Wang, Wan Du, and Chengyuan Li. Seaships: A large-scale precisely annotated dataset for ship detection. *IEEE Transactions on Multimedia*, 2018. 3

[29] Joonghyuk Shin, Minguk Kang, and Jaesik Park. Fill-up: Balancing long-tailed data with generative models, 2023. 3

[30] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*. PMLR, 2015. 2

[31] Michael Tipton, Elizabeth McCormack, Graham Elliott, Monica Cisternelli, Arthur Allen, and Arden C. Turner. Survival time and search time in water: Past, present and future. *Journal of Thermal Biology*, 2022. 1

[32] Brandon Trabucco, Kyle Doherty, Max Gurinas, and Ruslan Salakhutdinov. Effective data augmentation with diffusion models. ArXiv, 2023. 3

[33] Muhammad Umair, Manzoor Ahmed Hashmani, Syed Sajjad Hussain Rizvi, Hasmi Taib, Mohd Nasir Abdullah, and Mehak Maqbool Memon. A novel deep learning model for sea state classification using visual-range sea images. *Symmetry*, 2022. 5

[34] Leon Amadeus Varga, Benjamin Kiefer, Martin Messmer, and Andreas Zell. Seadronessee: A maritime benchmark for detecting humans in open water. In *WACV*, 2022. 2, 3, 5, 8

[35] Mabel M. Zhang, Jean Choi, Kostas Daniilidis, Michael T. Wolf, and Christopher Kanan. Vais: A dataset for recognizing maritime imagery in the visible and infrared spectrums. In *CVPRW*, 2015. 3