drones

MDPI

Article

Enhancing Turbidity Predictions in Coastal Environments by Removing Obstructions from Unmanned Aerial Vehicle Multispectral Imagery Using Inpainting Techniques

Hieu Trung Kieu [1], Yoong Sze Yeong [2], Ha Linh Trinh [1] and Adrian Wing-Keung Law [1,2,*]

[1] Environmental Process Modelling Centre, Nanyang Environment and Water Research Institute, Nanyang Technological University, Singapore 637141, Singapore; trunghieu.kieu@ntu.edu.sg (H.T.K.); halinh.trinh@ntu.edu.sg (H.L.T.)

[2] School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798, Singapore; yyeong001@e.ntu.edu.sg

* Correspondence: cwklaw@ntu.edu.sg

Abstract: High-resolution remote sensing of turbidity in the coastal environment with unmanned aerial vehicles (UAVs) can be adversely affected by the presence of obstructions of vessels and marine objects in images, which can introduce significant errors in turbidity modeling and predictions. This study evaluates the use of two deep-learning-based inpainting methods, namely, Decoupled Spatial–Temporal Transformer (DSTT) and Deep Image Prior (DIP), to recover the obstructed information. Aerial images of turbidity plumes in the coastal environment were first acquired using a UAV system with a multispectral sensor that included obstructions on the water surface at various obstruction percentages. The performance of the two inpainting models was then assessed through both qualitative and quantitative analyses of the inpainted data, focusing on the accuracy of turbidity retrieval. The results show that the DIP model performs well across a wide range of obstruction percentages from 10 to 70%. In comparison, the DSTT model produces good accuracy only with low percentages of less than 20% and performs poorly when the obstruction percentage increases.

Keywords: remote sensing; UAV; coastal monitoring; turbidity; image inpainting

check for updates

Citation: Kieu, H.T.; Yeong, Y.S.; Trinh, H.L.; Law, A.W.-K. Enhancing Turbidity Predictions in Coastal Environments by Removing Obstructions from Unmanned Aerial Vehicle Multispectral Imagery Using Inpainting Techniques. *Drones* **2024**, *8*, 555. https://doi.org/10.3390/drones8100555

Academic Editor: Bo Li

Received: 28 August 2024
Revised: 30 September 2024
Accepted: 4 October 2024
Published: 7 October 2024

1. Introduction

The regulatory monitoring of coastal water turbidity is crucial to safeguard shoreline and marine ecology during construction activities, especially in the context of climate change. Remote sensing, with its expansive spatial coverage, can significantly aid the monitoring in this endeavor [1]. In recent years, the increasing utilization of UAV imagery has brought unique advantages such as high flexibility and on-demand deployment to further expand the monitoring capabilities [2]. UAV imagery can achieve configurable resolutions down to centimeters, contingent on flight altitude [3,4]. However, an emerging challenge associated with the high-resolution remote sensing is the presence of obstructions, such as vessels and marine objects in the imagery of the coastal environment, particularly in bustling port areas. Such obstructions impede data completeness by concealing valuable information beneath the obstructions [5,6]. In particular, for the monitoring of coastal turbidity plumes, these obstructions in images can introduce significant errors to turbidity predictions. Thus, the proper handling of obstructed information due to vessels and marine objects in imagery can be crucial in high-resolution remote sensing applications for the coastal environment.

In the field of computer vision, the techniques used to restore occluded regions caused by the obstruction of specific objects in the imagery are termed “image inpainting”, which have seen extensive development over recent decades [7]. Image inpainting methods aim to reconstruct the missing or damaged regions while maintaining visual plausibility.

They vary in terms of data types, input formats, referencing methods, and processing systems. Recent advancements in deep learning methods have also revolutionized this domain [8]. State-of-the-art image inpainting techniques can now be primarily categorized into two groups: (1) non-generative methods, which utilize “copy and paste” techniques, drawing information from the neighboring pixels within the image [9,10]; and (2) generative methods, which employ generative data-driven models to produce realistic and content-aware completion for occluded regions [11,12]. Noteworthy generative methods include the Context Encoder [13], Partial Convolutional [14], and DeepFill v2 [15], each designed to address specific issues such as blurriness, overall and local consistency, realism, etc. Table 1 presents an overview of the loss functions utilized in these different generative inpainting methods, along with their descriptions.

Video inpainting techniques have also advanced significantly in recent years by considering the temporal dimension with the spatial structure and motion coherence, addressing the additional complexities of handling image sequences. Examples of video inpainting models include the “copy and paste” techniques that rely on the optical flow approach for pixel tracking and the Video Inpainting Network (VINet) techniques that prioritize temporal consistency through recurrent feedback, flow guidance, and a temporal memory layer [16]. To better capture the relationship between frames and, several inpainting models, including DSTT [17] and Fuse Former [18], further incorporate either 3D convolutional neural networks (CNNs) or self-attention-based vision transformers. Ulyanov et al. [19] introduced another alternative, the Deep Image Prior (DIP) model, which initializes a deep neural network with random noise and optimizes it to achieve the desired properties. In doing so, the DIP model offers a novel approach to address the challenges of intricate textures, dynamic backgrounds, and computational resource constraints while eliminating the necessity for model training.

Table 1. Existing loss functions in generative inpainting models.

Loss Terms	Description	References
Reconstruction Loss	Compute the pixel-wise distance between network prediction and ground truth images.	[13,20,21]
Adversarial Loss	Encourage closer data distributions between the real and filled images.	[17,18,22]
Perceptual Loss	Penalize the feature-wise dissimilarity between the reconstructed images from pre-trained Visual Geometry Group (VGG) network and ground-truth images.	[23–25]
Markov Random Fields Loss	Compute the distance between each patch in missing regions and the nearest neighbor.	[26,27]
Total Variation Loss	Compute the difference between adjacent pixels in the missing regions. Ensures the smoothness of the completed image.	[28–30]

Existing general inpainting models based on conventional image and video formats can be potentially further refined for remote sensing applications to address the obstruction issue. A recent study by [31] employed the DIP model to remove traces of sensitive objects from synthetic-aperture radar (SAR) images across various land surface classifications, which featured an abundance of distinctive characteristics. Long et al. [32] proposed a bishift network (BSN) model with multiscale feature connectivity, including shift connections and depth-wise separable convolution, to remove and reconstruct missing information obstructed by thick clouds from Sentinel-2 images. Park et al. [33] proposed a deep-learning-based image segmentation and inpainting method to remove unwanted vehicles on road UAV-generated orthomosaics. However, retrieving the water surface information obstructed by vessels and marine objects is more challenging. Firstly, the water surface is typically homogeneous, providing less information for the generative model to use in reconstruction [34,35]. Secondly, the dynamic nature of the coastal environment can introduce spatial and temporal noise into the inpainting process [36]. This effect is particularly noticeable in UAV data, where images are captured sequentially along a specific flight path. During this period, variations in the water surface and the movement of marine objects can introduce inconsistent information, complicating reconstruction for the inpainting model. In addition, the evaluation of model performance in removing obstructions from coastal remote sensing images requires a comprehensive assessment beyond individual image-wise evaluations. Therefore, as far as we are aware, inpainting models have not been implemented for reconstructing water surfaces in UAV imagery, highlighting a gap in the current literature.

In this study, state-of-the-art deep-learning-based inpainting models, namely, the DSTT and DIP models, are investigated to recover missing information obstructed by vessels and marine objects in the sequential UAV multispectral imagery of the coastal environment. Their performances are examined qualitatively and quantitatively using a dataset of UAV multispectral images acquired during this study for monitoring turbidity plumes in the coastal environment. In the following section, we will first describe the UAV survey before presenting the two models and discussing the results obtained.

2. Materials and Methods

2.1. UAV Imagery

A UAV survey for aerial image acquisition was conducted on 19 August 2022 at a coastal area of Singapore where intensive engineering operations of dredging and dumping activities for land reclamation were ongoing, creating turbidity plumes that were visible on the water surface. The integrated UAV-borne multispectral imagery system is shown in Figure 1, as reported in [37]. The system consisted of a DJI Matrice 300 RTK UAV (SZ DJI Technology Co., Ltd., Shenzhen, China), carrying a MicaSense RedEdge-MX Dual multispectral camera (MicaSense, Inc., Seattle, WA, USA). The camera had ten discrete light bands within the visible-to-near-infrared (VIS-NIR) spectrum. A Downwelling Light Sensor (DLS) was mounted onboard the UAV system to capture the light intensity variation during the UAV flight.

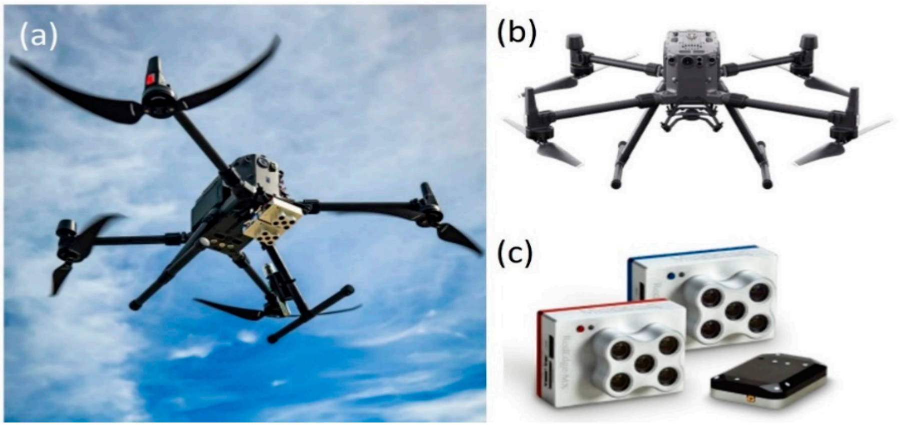

Figure 1. (**a**) The UAV-borne multispectral imagery system for water turbidity image acquisition and its components of (**b**) DJI M300 RTK UAV and (**c**) MicaSense RedEdge-MX Dual Camera. Reprinted with permission from [37].

The UAV survey captured the aerial images following a lawn-mowing scanning pattern as described in [38] with ten parallel scanning lines. Hence, the imagery covered the dredging vessel and generated turbidity plume. The flight was conducted at an altitude of 60 m above mean sea level (AMSL) with a cruising speed of 3.5 m/s to achieve a frontal overlap ratio of 85% and a ground resolution of 3.5 cm/pixel. A sequence of 791 images (each having a dimension of 1245 × 900 pixels) was captured during the UAV flight.

During the UAV survey, in situ field measurements of water turbidity were carried out synchronously using a turbidity probe with Global Positioning System (GPS) capability, namely YSI ProDSS Multiparameter Digital Water Quality (Xylem Inc., Washington, DC, USA). The probe was mounted at a depth of approximately 0.5 m below the water surface and on one side near the front of the sampling vessel to minimize potential turbulence caused by the vessel's motion [38]. It automatically logged the turbidity measurement (in Formazin Nephelometric Units, FNU) and its corresponding location coordinates at one-second intervals as the vessel traversed the survey area. A total of 744 in situ sampling data points were recorded during the UAV flight.

2.2. Data Pre-Processing

Since the sizes of vessels and marine objects vary between images, the obstruction percentage is introduced and calculated for each image as the percentage of non-water pixels relative to the total number of pixels. Based on these obstruction percentages, images from Line 1 and Line 10 of the UAV flight were chosen for the present analysis. Table 2 summarizes the obstruction percentages for both datasets.

Table 2. Obstruction percentage of UAV datasets.

Obstruction Percentage (%)	Line 1	Line 10
Min	4.15	0.91
Max	20.1	73.9
Average	11.6	33.2

Figure 2 shows sample images from the two datasets. The Line 1 dataset contains small obstructions, averaging 11.6%, in 24 out of 34 images, mainly featuring boats or floating pipes, whereas the Line 10 dataset has a large obstruction, averaging 33.2%, in 23 out of 34 images due to the dredging vessel. This difference in obstruction size between Line 1 and Line 10 enables the comparative evaluation on the performance of the inpainting models with respect to varying obstruction percentages.

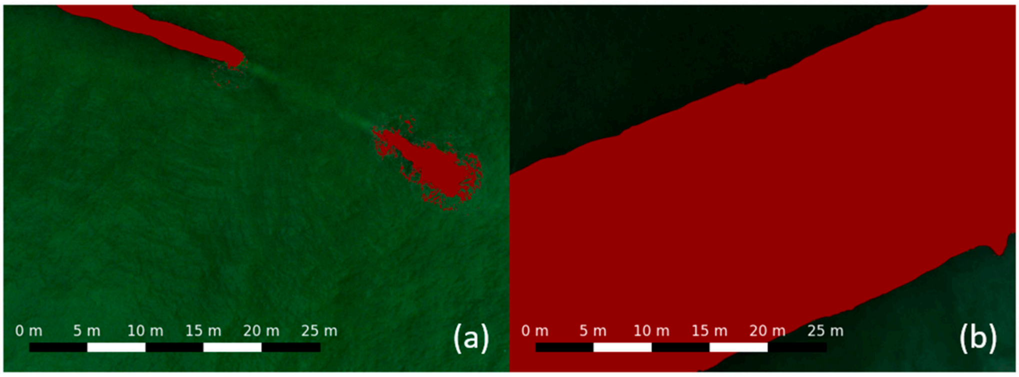

Figure 2. Sample images captured at (**a**) Line 1 and (**b**) Line 10 of the UAV flight. The vessel and marine objects are masked in red for confidentiality.

Adjustments in image alignment were first performed among the band images for both Lines 1 and 10 using the motion homography transform method [39]. Subsequently, the aligned images were calibrated using the light intensity information from the DLS to generate normalized reflectance images with pixel values ranging from 0 to 1. The red, green, and blue channels of the multispectral images were extracted and assembled into an RGB format to meet the input specifications of the inpainting models. The normalized images required cropping and resizing to satisfy the torch size requirements of the inpainting models. For the DSTT model, images were cropped to 1245 × 692 pixels and then resized to the stipulated dimension of 432 × 240 pixels. For the DIP model, a dimension divisible by 64 was required by the model architecture. Hence, images were cropped and resized to a final dimension of 1216 × 896 pixels. Figure 3 shows the image cropping and resizing of the two models.

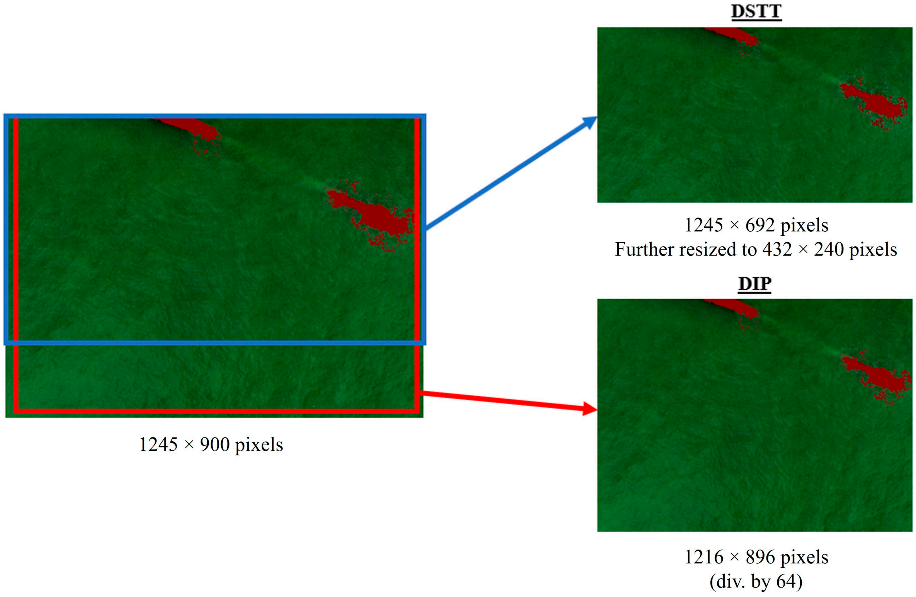

Figure 3. Image resizing for the two models.

In the next step, image annotation was conducted to identify the obstructing objects in the images, as shown in Figure 4. The precise annotation mask of the objects was created by manipulating the luminance values in specific color channels to selectively control the visibility of the pixels. Tailored thresholds were then established for each image to ensure accurate vessel delineation from the background, and undesired background elements were removed based on their brightness levels. In addition, to augment the data for evaluation, the mask was enlarged to include the surrounding pixels, creating synthetic object areas with ground-truth information about the water surface surrounding the object, which can be used to evaluate the inpainting model performance.

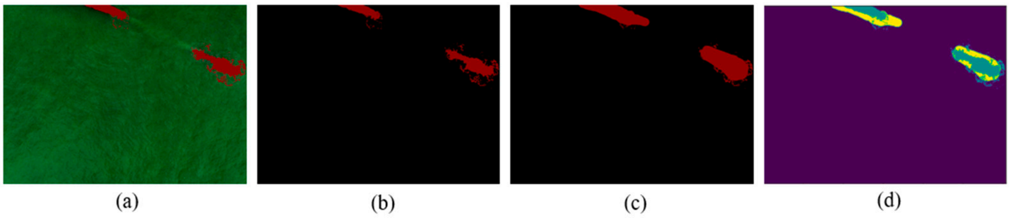

Figure 4. (**a**) Original UAV image, (**b**) precise annotation mask, (**c**) synthetic annotation mask, and (**d**) the difference between (**b**) and (**c**). The yellow area was used as ground-truth information for model evaluation.

From the calibrated reflectance image, we employed a GPS-based stitching algorithm [35] to generate the mosaic image for each flight line. The mosaic image was then georeferenced using the GPS information embedded in the raw images. Subsequently, spectral information was obtained by extracting the red, green, and blue values from the pixel corresponding to the coordinates of in situ turbidity measurements. Cross-referencing against known sampling locations and ground-truth measurements was conducted to enhance the credibility of the results obtained.

2.3. Vessel Removal Method

As discussed earlier, two deep inpainting models, namely, DSTT and DIP, were investigated to retrieve the obstructed information from the aerial images. The DSTT model separates spatial–temporal attention learning into two distinct sub-tasks. It utilizes temporally decoupled Transformer blocks for object movements across frames and spatially decoupled Transformer blocks for similar background textures within frames [17]. The combination of these two blocks enables precise attention to background textures and moving objects, facilitating visually plausible and temporally coherent inpainting appearances. Furthermore, the DSTT model also incorporates a hierarchical encoder for robust feature learning, maintaining a multi-level local spatial structure. This innovative design yields a more efficient spatial–temporal attention scheme. As highlighted earlier, the DSTT model has not been examined for coastal remote sensing as far as we are aware. Hence, its adoption necessitates an assessment of its suitability with various considerations for coastal applications. The parameters of the DSTT model adopted in this study are shown in Table 3.

Table 3. Parameters of DSTT model.

Parameter	Value
Dimension of outputs	432 × 240
Output format	mp4
No. of reference frames to be selected	−1 (all possible frames)
Output video fps	5
Neighbor stride	5

Comparatively, the DIP model leverages the inherent structure of a randomly initialized neural network for image processing [19]. Unlike traditional methods that rely on labeled training data, the DIP model initializes a neural network with random noise and iteratively refines its parameters to enhance or reconstruct the input images. The architecture of the network serves as a prior component, capturing essential fundamental features of the natural images. Through optimization, the DIP model achieves tasks such as denoising or inpainting without the need for explicit training. To elaborate further on the prior component and randomization used in the DIP model, the initial weight matrix θ_0 is first randomly generated and iteratively updated to minimize the data term. At every iteration t, the weight matrix θ is mapped to an image $x = f_\theta\ (z)$, where z is a fixed tensor, and the mapping f is a neural network. The mapped image x is then used to compute the task-dependent loss $E(x, x_0)$. The gradient of the loss with respect to θ is used to update the parameters. The success of the DIP model depends on the careful choice of architecture, hyperparameters, and loss functions tailored to the specific image processing task. In this study, the hyperparameters of the DIP model were refined through a trial-and-error approach to determine the optimal hyperparameter set, as shown in Table 4.

Table 4. Parameters of DIP model.

Parameter	Value
Dimension of outputs	432 × 240
Output format	mp4
No. of iteration	30,000
Network architecture	skip_depth_6
Learning rate	0.01

2.4. Evaluation Metrics

2.4.1. Assessment with Ground-Truth Information

The concept of "synthetic area" introduced in [40] was adopted. Following vessel annotation, we expanded the border of the mask to encompass the surrounding pixels of the vessel region, ensuring that the inpainting models reconstruct pixels with ground-truth information. Subsequently, the inpainted pixels were compared with the original pixels from the synthetic areas to evaluate the model performance. The evaluated metrics include the mean absolute error (MAE) and the coefficient of determination (R^2) as shown in the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{m} (y_i - \overline{y})^2} \tag{1}$$

$$MAE = \frac{\sum_{i=1}^{m} |\hat{y}_i - y_i|}{m} \tag{2}$$

where $\hat{y}_i$ and y_i are the inpainted and ground-truth values of pixel i, respectively, $\overline{y}$ is the mean pixel value of the ground-truth image, and m is the number of pixels in the synthetic area.

2.4.2. Temporal Consistency

The images were captured sequentially at fixed time intervals as the UAV executed its flight pattern, resulting in obstructions appearing in adjacent images. Therefore, it is essential that the inpainting model can ensure the consistency of the obstructions among adjacent images when processing the entire image sequence. To assess the temporal consistency, corresponding regions among adjacent images were compared. Given that the overlap ratio of 85% was configured during the UAV flight, each image shares 85% of its covered region with the preceding and succeeding images (Figure 5). These overlapping regions were extracted and utilized to calculate the variance score using Equation (3) as follows:

$$Variance = \frac{1}{N} \sum_{i=1}^{N} (y_i - \overline{y})^2 \tag{3}$$

where $\overline{y}$ is the mean value of pixel location i across the two frames, and N is the total number of pixels of a flattened frame. A lower variance among the sequential frames implies better inpainting performance as the overlapping regions in adjacent images have lesser variation from each other.

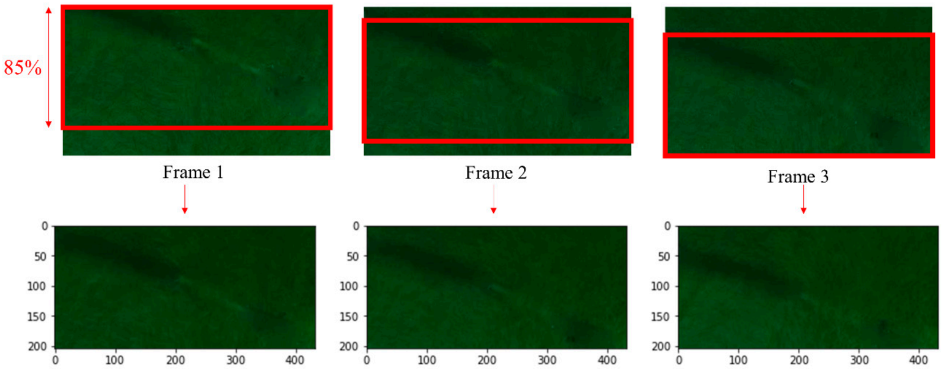

Figure 5. Illustration to compare variance among frames.

2.4.3. Assessment of Turbidity Retrieval Performance

Expanding beyond the visual impact assessment, we further examined the accuracy of the inpainted information in determining the turbidity on the water surface. The assessment procedures are illustrated in Figure 6. From the retrieved turbidity and RGB information, a filter is applied to remove the data points with a time gap between imaging and sampling that exceeded 5 min. Subsequently, the inpainted RGB information is used to calculate the percentage difference from the original values. A histogram of the percentage difference is then plotted, and a threshold is determined based on the majority of the values. Data points exceeding this threshold would imply a significant percentage difference and are then labeled as inpainted data; otherwise, they are labeled as background data. Background data, characterized by similar original and inpainted RGB values along with their corresponding turbidity values, are then utilized to establish a turbidity retrieval function though multivariate linear regression. Contrarily, the inpainted data are used as input in the multivariate linear regression model to predict its corresponding turbidity values.

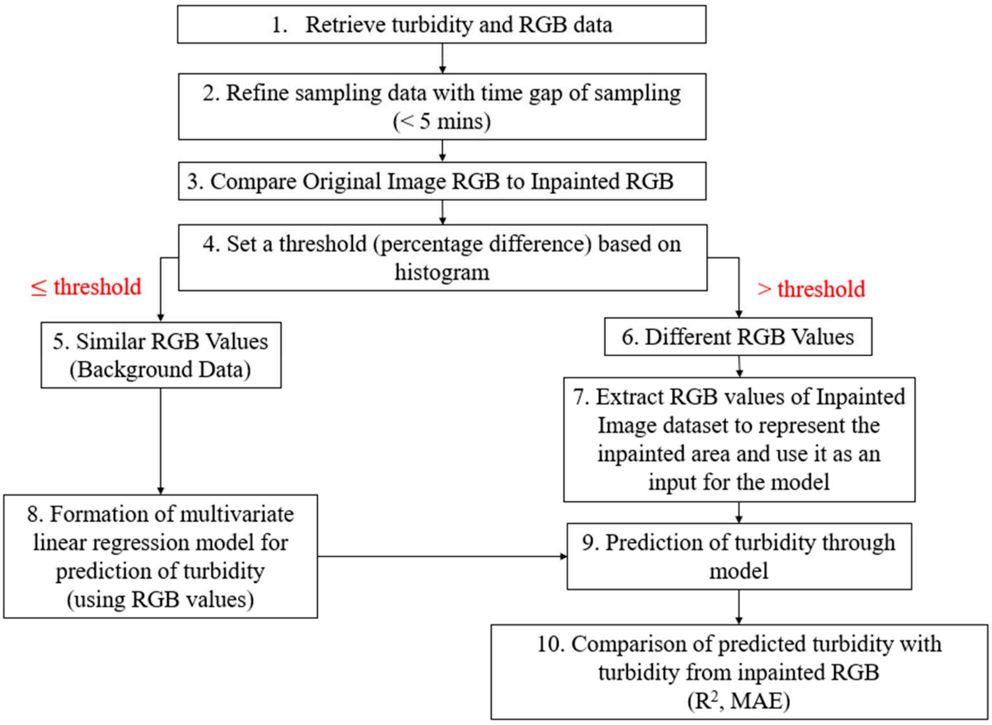
1. Retrieve turbidity and RGB data
2. Refine sampling data with time gap of sampling (< 5 mins)
3. Compare Original Image RGB to Inpainted RGB
4. Set a threshold (percentage difference) based on histogram
≤ threshold
> threshold
5. Similar RGB Values (Background Data)
6. Different RGB Values
7. Extract RGB values of Inpainted Image dataset to represent the inpainted area and use it as an input for the model
8. Formation of multivariate linear regression model for prediction of turbidity (using RGB values)
9. Prediction of turbidity through model
10. Comparison of predicted turbidity with turbidity from inpainted RGB (R², MAE)

Figure 6. Procedures for evaluating model performance for turbidity retrieval.

We adopted the optimal band ratio analysis (OBRA) approach [41,42] to determine the predictors of the regression function. The OBRA approach evaluates reflectance ratios, $X_{ij} = R(\lambda_i)/R(\lambda_j)$, where $R(\lambda_i)$ and $R(\lambda_j)$ present reflectance values from different band combinations and determines the most suitable function for predicting the target parameter (i.e., turbidity). From the dataset, the possible predictors are $R(\lambda_r)$, $R(\lambda_g)$, and $R(\lambda_b)$, representing the reflectance value of the red, green, and blue channels, respectively. In addition, OBRA also generates three reflectance ratio combinations of $X_{rg} = R(\lambda_r)/R(\lambda_g)$, $X_{rb} = R(\lambda_r)/R(\lambda_b)$, and $X_{gb} = R(\lambda_g)/R(\lambda_b)$. Multivariate regression functions are established for these three combinations of predictors. Comparison of the predictor combinations in Table 5 indicates that using the three bands, namely, $R(\lambda_r)$, $R(\lambda_g)$, and $R(\lambda_b)$, yields a high adjusted R^2 score and the lowest root mean square error (RMSE) among three combinations. Hence, this combination was used for the assessment of turbidity retrieval in the following analysis.

Table 5. Comparison of regression models using different combinations of predictors.

Predictor Combination	Adjusted R^2	RMSE (FNU)
$R(\lambda_r)$, $R(\lambda_g)$, $R(\lambda_b)$	0.786	0.192
X_{rg}, X_{rb}, X_{gb}	0.345	0.361
$R(\lambda_r)$, $R(\lambda_g)$, $R(\lambda_b)$, X_{rg}, X_{rb}, X_{gb}	0.793	0.410

The selection of model output to formulate the dataset for the turbidity retrieval function was carefully considered based on the specific features in the inpainting mechanism of the two models. While the DSTT model selectively fills in the masked areas only, leaving the other regions unchanged, the DIP model restores the entire image and thus introduces uncertainties in the non-masked regions at the same time. Therefore, for the enhanced accuracy of ground-truth data, the filtered data (both original and inpainted) from the DSTT model are used to establish the turbidity retrieval function.

3. Results

3.1. Qualitative Evaluation

A notable discrepancy occurs in the output resolution during the inpainting process. The DSTT model, constrained by the torch size limitation, reduces the output resolution to 432 × 240 pixels. The resolution reduction leads to compromised image fidelity, as shown in Figure 7, where the surface waves in the inpainted image can be seen to be smoothened compared to the original image.

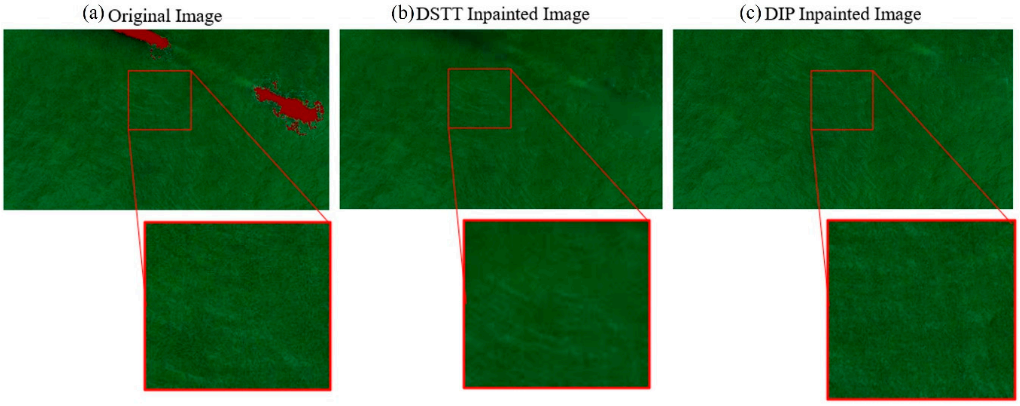

Figure 7. (**a**) Original images and inpainted images from the (**b**) DSTT and (**c**) DIP models. The water regions near the vessel are enlarged for comparison.

The DIP model outperforms the DSTT model in retaining the high-resolution output, again as evidenced by the fidelity of the surface waves in the inpainted image. However, the performance of the DIP model in reconstructing the detailed features is highly contingent on its hyperparameters, particularly the number of iterations. Figure 8 indicates the model performance across different numbers of iterations. It is evident that, at lower numbers (e.g., 3000, 5000, and 6000), the image shows large black areas which are indicative of incomplete inpainting performance. Satisfactory results can only be achieved at higher numbers of iterations (e.g., 20,000, 25,000, and 30,000).

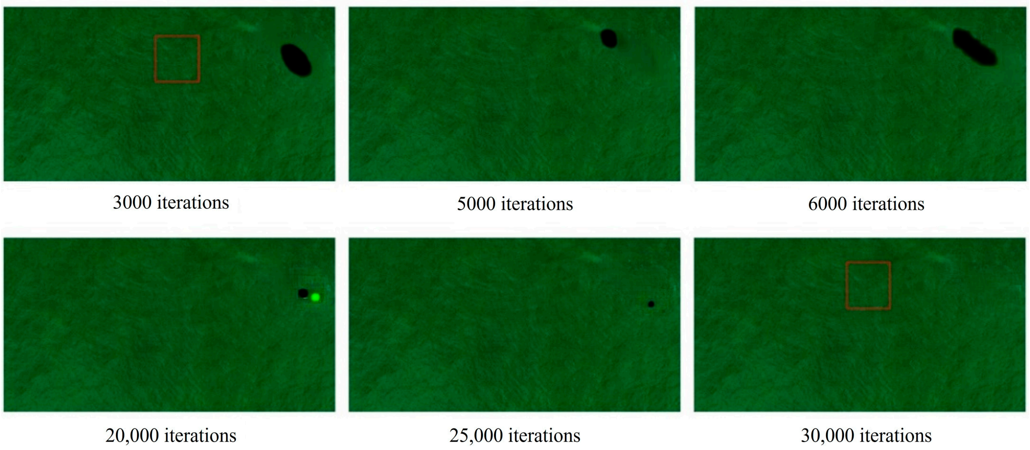

Figure 8. Effect of number of iterations on DIP model performance.

Although the DIP model can generate high-resolution images with commendable inpainting quality, there are instances where certain areas are inaccurately inpainted due to model randomization (Figure 9). Multiple reruns of the DIP model with the same settings and inputs can result in varying outputs, each potentially exhibiting improved or diminished performance. On the other hand, the DSTT model consistently maintains high-quality visual inpainting irrespective of the model reruns despite yielding lower resolution outputs, which can be attributed to its nature of referencing the images both before and after the frame to retrieve the data.

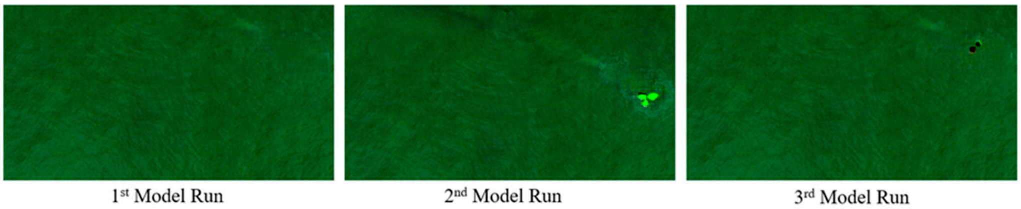

Figure 9. Inconsistency of DIP model results.

The output images from both the DSTT and DIP models were stitched together to assess the model performance across the entire flight line. Figure 10 shows the stitched images from the original, DSTT, and DIP inpainted images for both Line 1 and Line 10 data. For Line 1 data with relatively low obstruction percentages, both models perform relatively well in restoring the areas obstructed by the objects. There are, however, discrepancies in the stitched output of the DSTT model, which can be attributed to its mechanism that exclusively processes only the annotated pixels. In contrast, the DIP model operates on the whole image and thus can attempt to refine the discrepancies during the inpainting process.

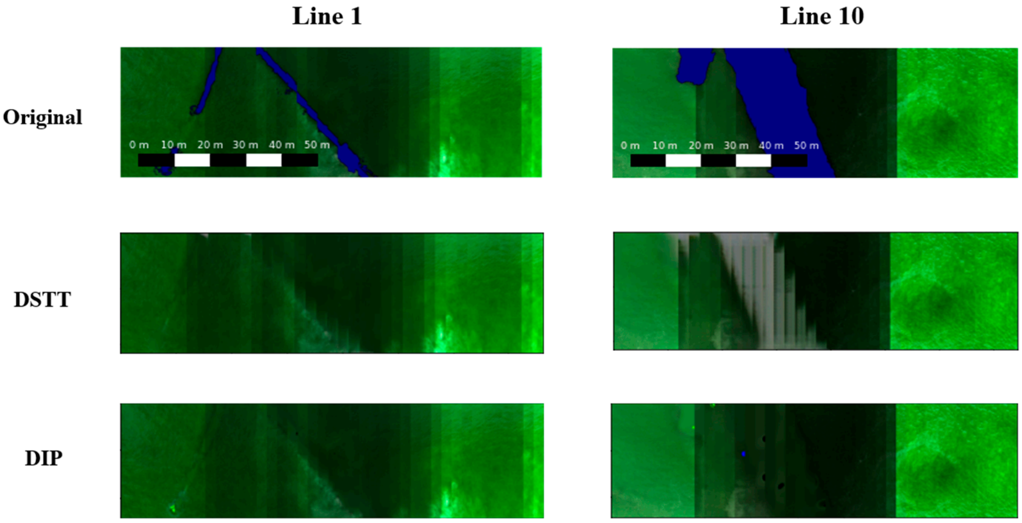

Figure 10. Mosaic images of Line 1 and Line 10 stitched from original, DSTT, and DIP inpainted images. The vessel and marine object are masked with blue color due to confidentiality.

Both models indicate a notable decline in performance when processing data with high obstruction percentages, as shown in the output results for Line 10. While the DIP model introduces minor errors in its outputs, the DSTT model fails to predict the background water surface beneath the large vessel, resulting in inaccurate reconstructions across the entire annotated region. This poor performance can be attributed directly to the insufficient information from the adjacent frames for reconstruction. In addition to its capability to utilize information from surrounding pixels, another notable feature of the DSTT model is its ability to track object movement and exploit information from adjacent images to recover the obstructed region. We note that the dredging vessel remained stationary throughout the UAV flight in this study. Hence, the information about the obstruction region is not available in subsequent images, leading to the failure of inpainting.

3.2. Quantitative Evaluation

3.2.1. Ground-Truth Assessment

The assessment of inpainting accuracy with respect to ground truths is not possible for remote sensing data as the underlying background information is naturally obstructed and cannot be retrieved. Therefore, in the literature, the inpainting algorithms are typically validated by using the synthetic approach whereby the object is artificially created on clear images of which the background information is already known. Subsequently, one can then assess the model performance in reconstructing the obstructed regions of the image.

Table 6 and Figure 11 show the R^2 scores and MAE values of the DSTT and DIP models in reconstructing the reflectance values at synthetic vessel areas in Line 1 and Line 10. Both models exhibit high R^2 scores, indicating strong reconstruction performance. Overall, the DIP model outperforms the DSTT model, achieving a higher R^2 value but a lower MAE score. However, for input images with a low obstruction percentage (i.e., Line 1), the DSTT model achieves a better R^2 score than the DIP model.

Table 6. Comparison of overall R^2 and MAE of the DSTT and DIP models.

	Line 1		**Line 10**		**Overall**	
Model	**DSTT**	**DIP**	**DSTT**	**DIP**	**DSTT**	**DIP**
R^2 score	0.872	0.866	0.761	0.795	0.82	0.831
MAE	122.4	106	131.5	70.04	126.5	88.4

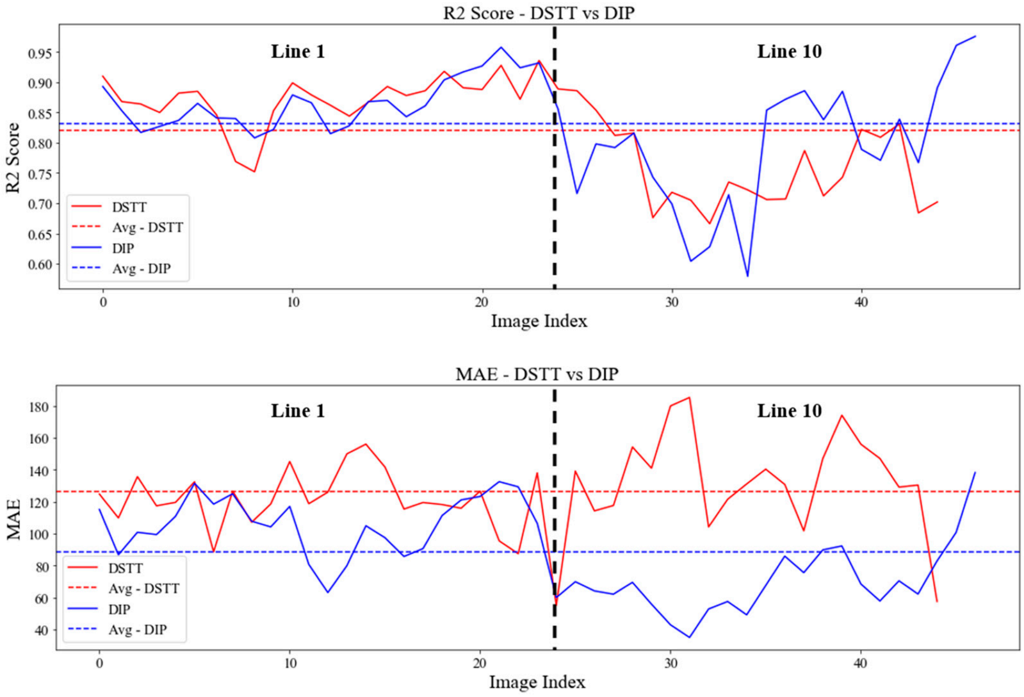

Figure 11. R^2 and MAE of the DSTT and DIP models with ground truth information. The images are indexed sequentially flowing the flight path of Lines 1 and 10.

3.2.2. Temporal Consistency

To quantitatively assess the temporal consistency of the inpainting performance, the variance score between the corresponding regions of adjacent inpainted images is calculated as shown in Figure 12. From the figure, the DSTT model exhibits better temporal consistency in Line 1, characterized by consistently lower variance scores across frames compared to the DIP model. However, in Line 10, the variance scores of the DSTT model are significantly higher, indicating a notable inconsistency in the inpainted images. This observation is consistent with the qualitative assessment of Line 10, highlighting the poor performance of the DSTT model on images with substantial obstruction percentages.

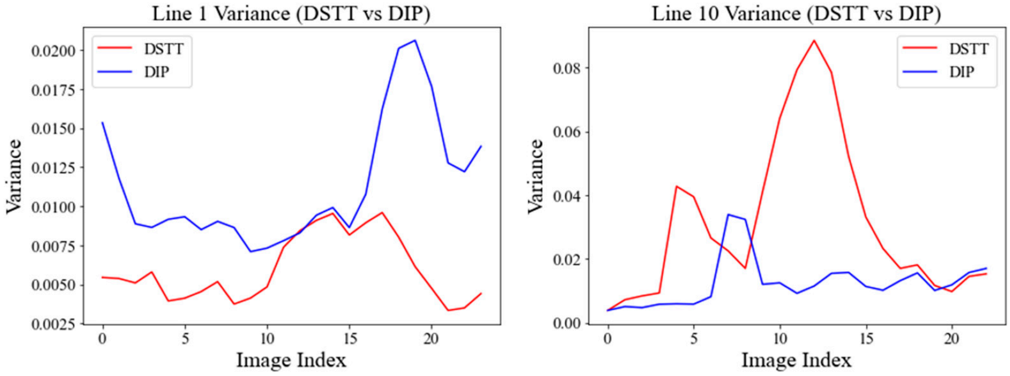

Figure 12. Variance in the inpainted image with previous and next frames for DSTT and DIP. The images are indexed sequentially flowing the flight paths of Lines 1 and 10.

The obstruction percentage of the images in both Lines 1 and 10 are calculated and plotted in accordance with the variance scores of the DSTT and DIP models in Figure 13. The average obstruction percentages in Line 1 and Line 10 are approximately 11.5% and 33.4%, respectively. Notably, the variance score of the DSTT model demonstrates a high sensitivity to the obstruction percentage, as evidenced by a strong linear correlation with a coefficient of 0.913. Conversely, the DIP model shows a remarkably lower correlation coefficient of 0.219, indicating that its inpainting quality remains similar regardless of obstruction percentages due to its randomized learning approach. These results highlight the suitability of the DSTT model for images with smaller objects while reaffirming the superiority of the DIP model for images with high obstruction percentages.

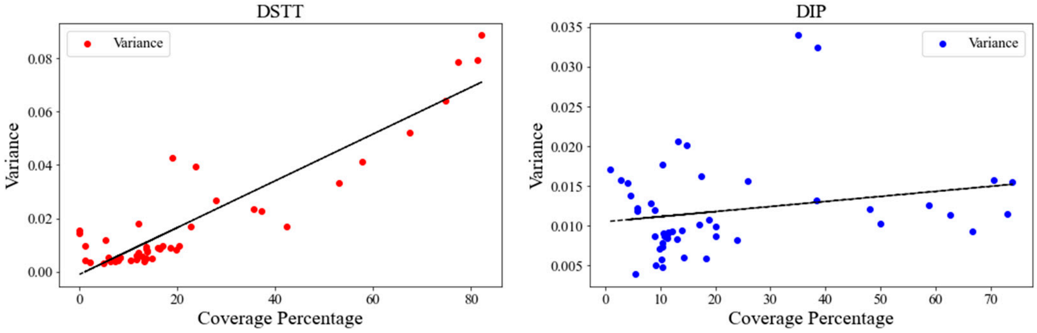

Figure 13. Correlation plot of variance and obstacle coverage percentage.

3.2.3. Applicability in Turbidity Retrieval Function

Table 7 illustrates the R^2 score and MAE of the turbidity retrieval function using the inpainted image from the DSTT and DIP models as inputs. From the table, the use of inpainted images from the DSTT model in Line 1 results in relatively more accurate predictions of water turbidity, surpassing those from the DIP model. However, the turbidity predictions based on the inpainted image from the DSTT model in Line 10 exhibit large errors in both R^2 score and MAE, which can be attributed again to the inaccurate inpainting performance. In contrast, the inpainted image from the DIP model consistently yields high-quality predictions for turbidity retrieval in both Lines 1 and 10. Once again, these findings highlight the superiority of the DIP model over the DSTT model for inpainting removal in coastal remote sensing applications.

Table 7. Comparison of inpainted images from the DSTT and DIP models for turbidity retrieval.

	Line 1		Line 10	
Model	**DSTT**	**DIP**	**DSTT**	**DIP**
R^2 score	0.872	0.866	0.761	0.795
MAE	122.4	106	131.5	70.04

4. Discussion and Conclusions

The qualitative and quantitative evaluations presented above reveal the strengths and weaknesses of the DSTT and DIP models for the inpainting removal of obstructions to improve UAV multispectral imagery for water turbidity retrieval as follows:

(a) The DSTT model excels at generating high-quality visual inpainting with low obstruction percentages, but encounters resolution constraints, reducing the output resolution by approximately threefold. This limitation can significantly affect the utility of the retrieved data, particularly for detailed environmental monitoring. Further improvements in the model architecture or specific transfer learning approaches will be necessary to address this limitation in the future. Moreover, the DSTT model struggles to effectively reconstruct areas with higher obstruction percentages due to its inability to leverage surrounding information for restoring extensively obstructed regions. As a result, its efficacy rapidly diminishes as obstruction percentages increase due to its reliance on adjacent frame data. This may result in incomplete or imprecise data restoration, potentially mispresenting environmental changes in areas with substantial obstructions. Hence, it is essential to identify a threshold obstruction percentage beyond which the DSTT model should not be considered at all.

(b) The DIP model demonstrates remarkable consistency in inpainting quality across different obstruction percentages. This attribute, coupled with its superior overall R^2 and lower MAE scores, underscores its robustness and versatility. The DIP model also offers high adaptability with its flexible resolutions, but introduces variability in image quality due to its reliance on hyperparameters and network architecture. This inconsistency poses a challenge for tasks requiring high precision, such as sensitive environmental assessments that require high texture and consistency.

Comparing the two models, the DIP model outperforms in inpainting removal over a wider range of obstruction percentages in terms of temporal consistency. At the same time, the data processing time from the DIP model is approximately 92.95 h in this study using a single Intel Xeon Gold 6258R processor (Intel Corp., Santa Clara, CA, USA), which is much longer than the DSTT model with approximately 4.69 min only. The prolonged processing time emphasizes a crucial trade-off between processing speed and output quality, which may impede the DIP model's adoption for remote sensing applications where timely data retrieval is essential. Hence, future research should focus on optimization strategies, such as algorithmic developments and hardware upgrades, to further reduce processing time and enhance the model's applicability for specific remote sensing tasks.

Author Contributions: Conceptualization, H.T.K. and A.W.-K.L.; methodology, H.T.K., Y.S.Y. and A.W.-K.L.; formal analysis, H.T.K. and Y.S.Y.; data curation, H.T.K. and Y.S.Y.; writing—original draft preparation, H.T.K.; writing—review and editing, H.T.K., Y.S.Y., H.L.T. and A.W.-K.L.; visualization, H.T.K. and Y.S.Y.; supervision, A.W.-K.L.; project administration, H.L.T.; funding acquisition, A.W.-K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Nanyang Environment and Water Research Institute (Core Fund), Nanyang Technological University, Singapore.

Data Availability Statement: Restrictions apply to the datasets.

Acknowledgments: The authors would like to thank Li Min and Dawn Pang for the procurement of equipment and Pak Hui Ying for assistance in the UAV operation and image data acquisition. We would like to acknowledge the support from the Environmental Process Modelling Centre (EPMC) for providing the equipment and facilities.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Klemas, V. Remote Sensing Techniques for Studying Coastal Ecosystems: An Overview. *J. Coast. Res.* **2011**, *27*, 2–17.
2. Kieu, H.T.; Law, A.W.-K. Remote sensing of coastal hydro-environment with portable unmanned aerial vehicles (pUAVs) a state-of-the-art review. *J. Hydroenviron. Res.* **2021**, *37*, 32–45. [CrossRef]
3. Guo, Y.; He, J.; Huang, J.; Jing, Y.; Xu, S.; Wang, L.; Li, S.; Zheng, G. Effects of the Spatial Resolution of UAV Images on the Prediction and Transferability of Nitrogen Content Model for Winter Wheat. *Drones* **2022**, *6*, 299. [CrossRef]
4. Domingo, D.; Ørka, H.O.; Næsset, E.; Kachamba, D.; Gobakken, T. Effects of UAV Image Resolution, Camera Type, and Image Overlap on Accuracy of Biomass Predictions in a Tropical Woodland. *Remote Sens.* **2019**, *11*, 948. [CrossRef]
5. Shastry, A.; Carter, E.; Coltin, B.; Sleeter, R.; McMichael, S.; Eggleston, J. Mapping floods from remote sensing data and quantifying the effects of surface obstruction by clouds and vegetation. *Remote Sens. Environ.* **2023**, *291*, 113556. [CrossRef]
6. Chen, X.; Xu, X.; Yang, Y.; Wu, H.; Tang, J.; Zhao, J. Augmented Ship Tracking Under Occlusion Conditions from Maritime Surveillance Videos. *IEEE Access* **2020**, *8*, 42884–42897. [CrossRef]
7. Elharrouss, O.; Almaadeed, N.; Al-Maadeed, S.; Akbari, Y. Image Inpainting: A Review. *Neural Process. Lett.* **2020**, *51*, 2007–2028. [CrossRef]
8. Zhang, X.; Zhai, D.; Li, T.; Zhou, Y.; Lin, Y. Image inpainting based on deep learning: A review. *Inf. Fusion* **2023**, *90*, 74–94. [CrossRef]
9. Moskalenko, A.; Erofeev, M.; Vatolin, D. Method for Enhancing High-Resolution Image Inpainting with Two-Stage Approach. *Program. Comput. Soft.* **2021**, *47*, 201–206. [CrossRef]
10. Niknejad, M.; Bioucas-Dias, J.M.; Figueiredo, M.A.T. Image Restoration Using Conditional Random Fields and Scale Mixtures of Gaussians. *arXiv* **2018**, arXiv:1807.03027.
11. Salem, A.; Mansour, Y.; Eldaly, H. Generative vs. Non-Generative AI: Analyzing the Effects of AI on the Architectural Design Process. *Eng. Res. J.* **2024**, *53*, 119–128. [CrossRef]
12. Jiang, Y.; Xu, J.; Yang, B.; Xu, J.; Zhu, J. Image Inpainting Based on Generative Adversarial Networks. *IEEE Access* **2018**, *8*, 22884–22892. [CrossRef]
13. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
14. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.C.; Tao, A.; Catanzaro, B. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
15. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-form image inpainting with gated convolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
16. Kim, D.; Woo, S.; Lee, J.Y.; Kweon, I.S. Recurrent temporal aggregation framework for deep video inpainting. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 1038–1052. [CrossRef]
17. Liu, R.; Deng, H.; Huang, Y.; Shi, X.; Lu, L.; Sun, W.; Wang, X.; Dai, J.; Li, H. Decoupled spatial-temporal transformer for video inpainting. *arXiv* **2021**, arXiv:2104.06637.
18. Liu, R.; Deng, H.; Huang, Y.; Shi, X.; Lu, L.; Sun, W.; Wang, X.; Dai, J.; Li, H. Fuseformer: Fusing fine-grained information in transformers for video inpainting. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
19. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep image prior. *Int. J. Comput. Vis.* **2020**, *128*, 1867–1888. [CrossRef]
20. Givkashi, M.H.; Hadipour, M.; Zanganeh, A.P.; Nabizadeh, Z.; Karimi, N.; Samavi, S. Image Inpainting Using AutoEncoder and Guided Selection of Predicted Pixels. *arXiv* **2021**, arXiv:2112.09262.
21. Tu, C.T.; Chen, Y.F. Facial image inpainting with variational autoencoder. In Proceedings of the IEEE 2nd International Conference of Intelligent Robotic and Control Engineering (IRCE), Singapore, 25–28 August 2019.
22. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
23. Rad, M.S.; Bozorgtabar, B.; Marti, U.V.; Basler, M.; Ekenel, H.K.; Thiran, J.P. Srobb: Targeted perceptual loss for single image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
24. Yang, Q.; Yan, P.; Zhang, Y.; Yu, H.; Shi, Y.; Mou, X.; Kalra, M.K.; Zhang, Y.; Sun, L.; Wang, G. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans. Med. Imaging* **2018**, *37*, 1348–1357. [CrossRef]
25. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016.

26. Liu, Z.; Li, X.; Luo, P.; Loy, C.C.; Tang, X. Deep learning markov random field for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 1814–1828. [CrossRef]
27. Li, C.; Wand, M. Combining markov random fields and convolutional neural networks for image synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
28. Chen, Q.; Montesinos, P.; Sun, Q.S.; Heng, P.A. Adaptive total variation denoising based on difference curvature. *Image Vis. Comput.* **2010**, *28*, 298–306. [CrossRef]
29. Oliveira, J.P.; Bioucas-Dias, J.M.; Figueiredo, M.A. Adaptive total variation image deblurring: A majorization–minimization approach. *Signal Process.* **2009**, *89*, 1683–1693. [CrossRef]
30. Chan, T.; Esedoglu, S.; Park, F.; Yip, A. Recent developments in total variation image restoration. *Math. Models Comput. Vis.* **2005**, *17*, 17–31.
31. Cannas, E.D.; Mandelli, S.; Bestagini, P.; Tubaro, S.; Delp, E.J. Deep Image Prior Amplitude SAR Image Anonymization. *Remote Sens.* **2023**, *15*, 3750. [CrossRef]
32. Long, C.; Li, X.; Jing, Y.; Shen, H. Bishift Networks for Thick Cloud Removal with Multitemporal Remote Sensing Images. *Int. J. Intell. Syst.* **2023**, *2023*, 9953198. [CrossRef]
33. Park, J.; Chol, Y.K.; Kim, S. Deep learning-based UAV image segmentation and inpainting for generating vehicle-free orthomosaic. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *115*, 103111. [CrossRef]
34. Taha, A.; Rabah, M.; Mohie, R.; Elhadary, A.; Ghanem, E. Assessment of Using UAV Imagery over Featureless Surfaces for Topographic Applications. *MEJ. Mansoura Eng. J.* **2022**, *47*, 25–34. [CrossRef]
35. Pak, H.Y.; Kieu, H.T.; Lin, W.; Khoo, E.; Law, A.W.-K. CoastalWQL: An Open-Source Tool for Drone-Based Mapping of Coastal Turbidity Using Push Broom Hyperspectral Imagery. *Remote Sens.* **2024**, *16*, 708. [CrossRef]
36. Kieu, H.T.; Pak, H.Y.; Trinh, H.L.; Pang, D.S.C.; Khoo, E.; Law, A.W.-K. UAV-based remote sensing of turbidity in coastal environment for regulatory monitoring and assessment. *Mar. Pollut. Bull.* **2023**, *196*, 115482. [CrossRef]
37. Trinh, H.L.; Kieu, H.T.; Pak, H.Y.; Pang, D.S.C.; Tham, W.W.; Khoo, E.; Law, A.W.-K. A Comparative Study of Multi-Rotor Unmanned Aerial Vehicles (UAVs) with Spectral Sensors for Real-Time Turbidity Monitoring in the Coastal Environment. *Drones* **2024**, *8*, 52. [CrossRef]
38. Trinh, H.L.; Kieu, H.T.; Pak, H.Y.; Pang, D.S.C.; Cokro, A.A.; Law, A.W.-K. A Framework for Survey Planning Using Portable Unmanned Aerial Vehicles (pUAVs) in Coastal Hydro-Environment. *Remote Sens.* **2022**, *14*, 2283. [CrossRef]
39. Evangelidis, G.D.; Psarakis, E.Z. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1858–1865. [CrossRef] [PubMed]
40. Zhao, M.; Olsen, P.; Chandra, R. Seeing through clouds in satellite images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4704616. [CrossRef]
41. Bioresita, F.; Puissant, A.; Stumpf, A.; Malet, J.P. A method for automatic and rapid mapping of water surfaces from Sentinel-1 imagery. *Remote Sens.* **2018**, *10*, 217. [CrossRef]
42. Pak, H.Y.; Law, A.W.K.; Lin, W. Retrieval of total suspended solids concentration from hyperspectral sensing using hierarchical Bayesian model aggregation for optimal multiple band ratio analysis. *J. Hydroenviron. Res.* **2023**, *46*, 1–18. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.