Revolutionize the Oceanic Drone RGB Imagery with Pioneering Sun Glint Detection and Removal Techniques

Jiangying Qin[†1], Ming Li[*†1,2], Jie Zhao[†3], Jiageng Zhong[1], Hanqi Zhang[1]
[1]Wuhan University, [2]ETH Zürich, [3]Technische Universität München
Mingli39@ethz.ch, {jy_qin, zhongjiageng, hqzhang}@whu.edu.cn, jie.zhao@tum.de

Abstract

The issue of sun glint poses a significant challenge for ocean remote sensing with high-resolution ocean drone imagery, as it contaminates images and obstructs crucial features in shallow-waters, leading to inaccurate benthic substrates identification. While various physics-based statistical solutions have been proposed to address this optical issue in remote sensing, there is a lack of sun glint detection and removal methods specifically designed for high-resolution consumer-grade drone RGB imagery. In this paper, we present a pioneering pipeline for sun glint detection and removal in high-resolution drone RGB images, aiming to restore the real features that are hindered by sun glint. Our approach involves the development of a Foreground Attention-based Semantic Segmentation Network (FANet) for accurate and precise sun glint detection, while effective sun glint removal is achieved through pixel propagation using an optical flow field. Experimental results demonstrate the effectiveness of our FANet in identifying sun glint, achieving IoU accuracy of 81.34% for sun glint pixels and 99.52% for non-sun glint background pixels. Furthermore, the quantitative evaluation of sun glint removal using two well-known metrics show that our method outperforms the GAN-based image restoration method (DeepFillv2) and the conventional image interpolation method (Fast Marching Method, hereafter referred to as FMM). Thus, our pipeline lays the foundation for accurate and precise marine coastal ecological monitoring and seafloor topographic mapping using consumer-grade drone at a low cost.

1. Introduction

Sun glint occurs when sunlight reflects directly towards the sensor from the water surface, causing issues in ocean observations [23]. This phenomenon is problematic in coastal ecological monitoring and benthic habitat mapping based on optical images. It can lead to the loss of visible features of benthic communities, resulting in incomplete and inaccurate data [1]. One approach to reduce sun glint is acquiring images on overcast days or using specific observation angles and sensor fields, but this may not always be possible [8, 26]. Filters like UV and ND have been proposed to mitigate sun glint, but their effectiveness depends on variables like sunlight angle and design, and they can compromise image quality [17]. Additionally, the large number of archived images affected by sun glint requires post-processing for effective utilization.

Sun glint studies mainly focus on developing detection and removal methods for sun glint in optical satellite imagery [3,37]. Two main categories of methods are used: sea surface-based and band information-based [7]. Sea surface-based methods predict water leaving reflectivity by integrating radiative transfer models and statistical models of surface water [30, 33]. Cox and Munk's statistical method established a relationship between sun glint statistics and surface slopes [2]. These methods require extrinsic information and are suitable for low-to-medium resolution imagery [27, 40]. Band information-based methods use the near-infrared band as an indicator of sun glint by assuming negligible water penetration [12, 15]. The covariance between visible bands and the NIR band is used to establish the relationship [21]. However, these methods are not universally effective for high-resolution images due to residual radiance and biases in complex shallow water environments [9]. Traditional physic-based methods accurately describe sun glint generation and propagation but are susceptible to environmental variations and noise. They have limited adaptability and high computational costs.

Nowadays, marine scientists and managers increasingly prefer drone-based remote sensing for precise coastal ecological monitoring. Drones with RGB cameras are mature, efficient, and easily deployable remote sensing platforms. They provide rich information, low cost, centimeter-level spatial resolution, and a high signal-to-noise ratio [18, 38]. However, traditional sun glint detection and removal methods for optical images are not applicable to drone RGB im-

[†] Equal technical contribution
[*] Corresponding author

ages due to resolution and sensor differences [5, 6]. Sun glint detection algorithms for drone images are in early stages and require further research. Only one published study proposes a sun glint detection method using a deep learning semantic segmentation model [11]. This method departs from traditional approaches by using deep learning to detect sun glint. These models learn features related to sun glint by training on annotated image datasets, allowing for detection in diverse optical scenarios. They also provide fast predictions for real-time processing of marine optical imagery. However, no study has investigated a complete sun glint detection and removal pipeline specifically designed for high-resolution oceanic drone RGB images. This gap in research results in low efficiency or even the uselessness of sun-glint-affected high-resolution RGB images obtained through drone remote sensing.

Thus, in order to fill the gaps in sun glint detection and removal using high-resolution drone RGB images, we propose a pioneering pipeline for sun glint detection and removal in high-resolution drone RGB images using the Foreground Attention Module and an optical-flow-based approach. It includes the development of the Foreground Attention-based Semantic Segmentation Network (FANet) to distinguish sun glint from the oceanic drone RGB imagery background. We also design an effective strategy using optical flow to accurately clear sun glint and restore real features. To evaluate our pipeline, we compare FANet with state-of-the-art (SOTA) image segmentation networks for sun glint detection. Additionally, since ground truth data for evaluating the restoration ability of our sun glint removal strategy is lacking in real-world scenarios, we apply our method to contamination-free drone images with artificial sun glint added. We conduct a quantitative evaluation by comparing our approach with other image restoration algorithms.

2. Related Work

As mentioned in section 1, our proposed pipeline contains two steps: sun glint detection and removal. So, in this section, the SOTA approaches pertaining to the two steps are summarized respectively.

2.1. Sun glint detection

Traditional methods lack the capability to detect sun glint consistently and reliably due to the variations in sun glint patterns across different times and locations. The rapid development of machine learning, particularly deep learning, has provided promising solutions for sun glint detection. Specifically, various semantic segmentation algorithms exhibit significant potential for sun glint detection. For example, UNet originally proposed for biomedical image segmentation adopts the encoder-decoder structure, using skip connections and a symmetric U-shaped network including compression paths and expansion paths [24]. By combining the low-level and high-level features of the input data through the skip connections, UNet is able to produce more accurate segmentation results, particularly in situations where the amount of training data is limited. Therefore, UNet is quite popular in semantic segmentation in many remote sensing applications [13, 41] suffering from limited annotated data, which is also the case in sun glint detection [11]. Another class of encoder-decoder structured image segmentation networks is the DeepLab series [4], proposing to use Atrous Spatial Pyramid Pooling (ASPP) in the spatial dimension to capture multi-scale information in order to generate accurate results. Although those encoder-decoder structured networks show good performance in image segmentation, information loss is always unavoidable during the downsampling of high-resolution feature maps and upsampling of low-resolution feature maps. Therefore, the high-resolution convolution network, i.e., HRNet [29], is proposed in order to keep high-resolution representation during the multi-scale feature extraction in parallel. Additionally, ConvNeXt [20] is proposed using only convolutional structures, yet it achieves high ImageNet Top 1 accuracy without relying on a particularly complex or innovative architecture. In numerous studies, the aforementioned deep learning (DL) models have demonstrated commendable performance in image segmentation.

When employing semantic segmentation for sun glint detection, [11] proposed an enhanced version of the classic UNet network (hereafter named as UNetglint), which is specially designed to address the challenges associated with sun glint detection in oceanic drone RGB imagery and is the SOTA algorithm in this domain. UNetglint incorporates dropout layers and batch normalization to address overfitting. However, detecting sun glint poses challenges due to its small size and inter-class imbalance. UNetglint's segmentation accuracy is subpar in this task. To address these issues, we propose a novel sun glint detection network that combines UNet with our Foreground Attention Module (FAM) [10, 14, 35]. FAM highlights valuable features of sun glint, improving semantic segmentation performance. Additionally, we use a hybrid loss function to mitigate inter-class imbalance and facilitate effective network training.

2.2. Sun glint removal

The primary objective of sun glint removal is to clear sun glint and restore the occluded real underwater texture features. Currently, some conventional image restoration methods rely on known appearance information to fill missing areas typically through image interpolation, typically Fast Marching Method (FMM) [32]. Alternatively, the image restoration can also be regarded as a conditional image generation task, carried out via Generative Adversarial Network (GAN), such as DeepFillv2 [39]. It should

be noted that the GAN-generated results are inconsistent with the ground truth regardless of their visual rationality and structural consistency with the real images. Simultaneously, some studies have attempted to treat sun glint as noise and utilize the Total Variation (TV) method for its removal [6]. However, this approach often leads to excessively smoothed outcomes, resulting in the loss of image details. In our study, the aim is to remove the sun glint contaminated pixels in RGB images and restore the accurate underwater features. One possibility for obtaining authentic benthic substrates can be realized by dense feature matching on image sequences, which is computationally expensive. Moreover, it is difficult to maintain temporal consistency in image sequences by matching and complementing individual pixels or image patches. Fortunately, the optical flow method [31] which characterizes the correspondence between adjacent frames by describing the instantaneous motion state of pixels of moving objects (*e.g.*, cameras or observed objects), thereby calculating the motion information of objects between adjacent frames can be introduced in sun glint removal due to its ability in maintaining pixels' temporal consistency and less computational complexity [36]. Moreover, drone imagery that is obtained in marine coastal monitoring characterizing high overlap, continuous shooting, and consistent brightness create good prerequisites for optical-flow-based sun glint removal. Furthermore, the available high-accuracy optical flow estimation methods such as GMA [16] also lay the foundation for the application of the optical flow field in sun glint removal. GMA effectively addresses the issue of accurate optical flow estimation in the presence of occlusion through the incorporation of image self-similarity modeling. Thus, we propose a sun glint removal strategy embedding an advanced optical flow method based on deep learning (*i.e.*, GMA) in this study to clear sun glint and restore the sun glint-contaminated pixels.

3. Materials and Methods

The workflow of our study is shown in Figure 1. First, the data collection and our study sites are introduced in section 3.1, followed by the data pre-processing description including the annotated dataset preparation for sun glint detection and artificial sun-glint-affected datasets preparation in section 3.2. Finally, the proposed FANet is detailed described in section 3.3 while the optical-flow-based sun glint removal using GMA [16] is presented in section 3.4.

3.1. Drone data collection and study sites

The drone coastal images were collected in January 2016 from the east of Cook's Bay in Moorea Island, French Polynesia (17°30′S, 149°50′W) (Figure 2), with more than 60% overlapping in adjacent images. In order to maintain a Ground Sampling Distance (GSD) of centimeter-level as required by this study, the camera was positioned at a fixed height above the objects or terrain as much as possible. It should be noted that all images were acquired on a sunny day with good lighting, resulting in inevitable sun glint contamination. This is because marine ecological monitoring requires images with clear texture, high visibility, and high-quality remote sensing images.

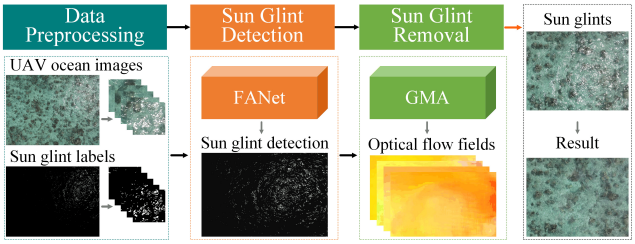

Figure 1. The pipeline of proposed sun glint detection and removal.

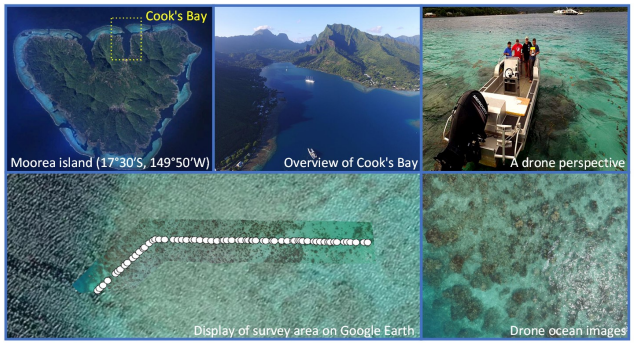

Figure 2. The study site and data collection during our fieldwork.

3.2. Data preprocessing

The annotated data was generated using a semi-automatic method. Specifically, the initial sun glint labels with a sufficient number of Regions of Interest (ROIs) were manually generated through visual inspection. Then, Support Vector Machine (SVM) algorithm was employed within those manually-derived annotated data in order to generate the labels for all images. It should be noted that there is a visual similarity in the sun glint and bleached corals, resulting in misclassification and omissions in the SVM results. Thus, experts checked all the SVM-derived labels and corrected all the labels manually to get the final annotated dataset. In this study, 82 coastal coral images with 4000×3000 pixels were obtained by drone. All images were randomly cropped into 256×256 pixels image patches as initial data. In order to reduce the imbalance problem between sun glint and its background, image patches covering too little or even no sun glint were discarded. At last, we

got 4869 image patches for training, 1217 image patches for validation and 742 image patches for testing. Data augmentation techniques, including hue transformation, random contrast transformation, random translation, random rotation (*i.e.*, 90°,180°, 270°), and random flip, have been applied in order to enhance the generalizability of our model and avoid overfitting. Furthermore, images patches were randomly cropped at multiple scales (*i.e.*, 256×256 pixels, 224×224 pixels, 168×168 pixels) and adjusted to a fixed size of 224×224 pixels in order to help the model learn multi-scale information.

Besides the annotated real dataset for sun glint detection, another artificial sun glint dataset was prepared for the evaluation of sun glint removal. Since it is difficult to evaluate the performance of sun glint removal methods due to a lack of accurate ground truth in practice images, we added sun glint by setting specific pixel saturation on the contamination-free image patch sequences as sun glint-affected images while the original contamination-free images can be used as ground truth.

3.3. Our sun glint detection neural network

The Foreground Attention-based Semantic Segmentation Network (*i.e.*, FANet) employs the classical encoder-decoder structure of UNet to simultaneously capture high-level global contextual information and low-level image details with skip connections, as shown in Figure 3. It has been proved that the supervision of the last layers of the decoder using ground truth in side output is able to reduce the overfitting, as is reported in the BASNet [25]. Besides, we also noticed that the attention mechanism should be considered in this study as sun glint typically manifests based on the fact that the sun glint often has intense contrast and luminosity, and individuals can easily and precisely spot sun glint in the image upon first inspection. Additionally, the sun glint's distribution tends to be regionally concentrated, frequently occurring in a small portion of the image as small objects, causing a significant data imbalance issue. Thus, a Foreground Attention Module (FAM) was employed to the imbalanced sun glint dataset in order to emphasize crucial features of sun glint while minimizing irrelevant ones [19]. The FAM was inserted into the last layers of different decoding stages, followed by the supervision of ground truth. The detailed architecture of FAM is shown in Figure 3, following a general attention design idea.

Given a feature tensor $F \in R^{W\times H\times C}$ where W, H, C present the number of width, height and channel of the feature map, respectively, a basic block is applied to F to get a feature F_{up} with a unified number of channels as 64:

$$F_{up} = ReLU\left\{BN\left[Conv_{3\times 3}\left(F\right)\right]\right\} \tag{1}$$

Then the Squeeze-and-Excitation (SE) attention module [14] is employed to get rescaled the features map F_{re} while

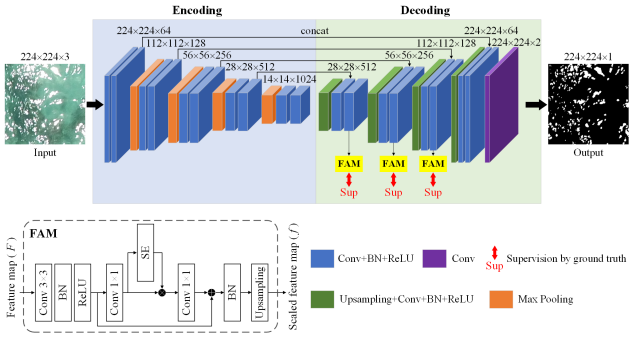

Figure 3. The architecture of propsed FANet which embeds FAM to UNet.

the two 1×1 convolutions are used to maintain the channel number, as follows:

$$F_{re} = Conv_{1\times 1}\left\{Conv_{1\times 1}\left(F_{up}\right) \times SE\left[Conv_{1\times 1}\left(F_{up}\right)\right]\right\} \tag{2}$$

In order to highlight the salient attributes of the foreground, we compute the scaled feature map f with a fixed size of 224×224×2 by combining F_{up} and F_{re}, followed by BN and upsampling (Equation 3). This f is then used for supervision with ground truth to guide the subsequent training iterations.

$$f = upsampling\left[BN\left(F_{up} + F_{re}\right)\right] \tag{3}$$

It is expected that FAM enhances the foreground features of interest while preserving the background information by adding and multiplying the foreground feature maps extracted by SE module with the original feature maps.

Another innovation of this method is in loss function implementation. Firstly, a hybrid loss function comprising pixel-level Weighted Cross Entropy (WCE) loss function [22] and patch-level Structural Similarity ($SSIM_{loss}$) loss function [34] is employed to guide the training of the network. The WCE loss function is a variant of the Cross Entropy (CE) loss function that weights positive and negative samples, aiming to solve the imbalance issue in sun glint and its background. The WCE loss function l^{WCE} is defined as shown in Equation 4.

$$l^{WCE} = -\sum_{i=1}^{N} w_i y_i log(p_i) \tag{4}$$

where N represents the total number of classes, with $N = 2$ denoting the two classes in this study (*i.e.* sun glint and background). The weights w_i corresponds to a specific class i and is determined by the proportion of pixels belonging to that class relative to all pixels. y_i and p_i denote the ground truth label and the predicted probability for class i, respectively.

The $SSIM_{loss}$, initially developed for the evaluation of image quality, demonstrates the capability to capture the inherent structural information within an image and assess the structural similarity between the predicted image and the original image. It assigns higher weights to boundaries by considering the local neighborhood of pixels, making the network more focused on the structural features of foreground classes. The definition of $SSIM_{loss}$ loss function $l^{SSIM_{loss}}$ for two given image patches x and y is presented in Equation 5.

$$l^{SSIM_{loss}} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5)$$

where μ_x and μ_y refer to the means of x and y, respectively, while σ_x and σ_y denote the standard deviations associated with x and y. σ_{xy} is their covariance. The constants $C_1 = 0.01^2$ and $C_2 = 0.03^2$ are small values utilized for ensuring numerical stability during the calculations.

Specially, the side output of the last layers of the first three decoder stages is supervised by the WCE loss function only. The network output is supervised by the WCE and $SSIM_{loss}$ hybrid loss function, which is defined in Equation 6.

$$l_{network}^{hybrid} = l_{network}^{WCE} + \alpha l_{network}^{SSIM_{loss}} \quad (6)$$

where $l_{network}^{hybrid}$, $l_{network}^{WCE}$, $l_{network}^{SSIM_{loss}}$ represents the hybrid loss function, WCE loss function and $SSIM_{loss}$ loss function for the network output, respectively. α is a hyperparameter, which is set as 0.4 according to trail-and-error.

The final training loss, denoted as l, is defined as the sum of all side outputs supervised by the WCE loss function, along with the network output supervised by the hybrid loss function, as shown in Equation 7.

$$l = \sum_{m=1}^{M} \beta l_m^{WCE} + l_{network}^{hybrid} \quad (7)$$

l_m^{WCE} represents the mth side output result after FAM and supervised by ground truth, and the total number of side outputs is $M = 3$. β is the weight of each loss, which is set to 0.2 according to trial-and-error.

3.4. Our sun glint removal strategy

Our optical-flow-based sun glint removal strategy (Figure 4) consists of three steps: (1) optical flow field estimation, (2) pixel propagation, and (3) unseen regions inpainting.

To remove sun glint from images, accurate optical flow estimation is crucial as it guides correct pixel propagation and inpainting of unseen regions. However, sun glint occlusion can cause correspondence ambiguity, leading to outliers in the cost volume, which affects optical flow decoding. To address this issue, we use the Global Motion Aggregation (GMA) method [16] for optical flow estimation,

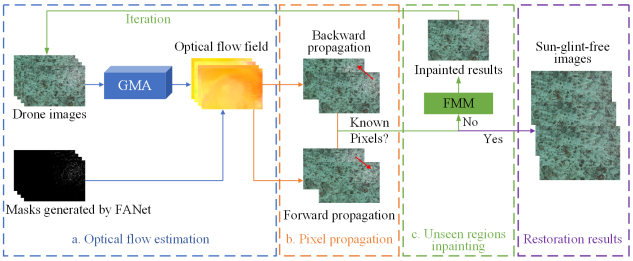

Figure 4. A workflow for the proposed sun glint removal strategy.

which propagates motion information from non-occluded background pixels to occluded ones (*i.e.*, sun glint pixels) through object self-similarity modeling. GMA takes visual context features and local motion features as input and outputs an aggregated global motion feature that is concatenated with local motion features and visual context features, decoded by GRU into a residual flow, and then the final optical flow is obtained. We would like to point out that, in our study, the utilization of a pre-trained GMA for optical flow estimation is justified by its robustness and efficacy across various image textures. Despite the dissimilarity between the GMA training dataset and our own, the optical flow field focuses on the motion data of pixels within successive image frames, rather than the image's texture.

Then, optical-flow-based pixel propagation is utilized to complete the sun-glint-contaminated pixels. This involves propagating through forward and backward optical flow until two known pixels are reached. A forward-backward consistency check is then performed to verify the validity of the pixel propagation. This check ensures that the forward optical flow and backward optical flow of two frames are equal in value and opposite in direction [42]. The consistency is measured using the forward-backward consistency error, which is defined as the two-norm of the sum of forward optical flow and backward optical flow. For a given pixel x, where x represents its 2D location within the image, the definition of the error is shown in Equation 8:

$$err_{ij}(x) = \|f_{ij}(x) + f_{ji}(x + f_{ij}(x))\|_2^2 \quad (8)$$

where err_{ij} represents forward-backward consistency error between frame i and frame j, and f_{ij} represents the optical flow value from i to j. Moreover, weights have been assigned to the two known pixels obtained by optical flow propagation through the forward and backward consistency error and the final pixel is completed via weighted fusion. In this study, the weights w is defined as an exponential function (Equation 9):

$$w_x = e^{-\frac{err_x}{m}} \quad (9)$$

where m is set to 0.1 to reduce possible excessive errors.

If some pixels cannot be filled through the pixel propagation due to the presence of sun glint in all image sequences or problems in optical flow estimation, single image restoration using FMM [32] is used as a solution. Then FMM-based restored image is used as input for the next iteration until all missing pixels are completed, ultimately producing sun-glint-free images.

4. Experiments and results analysis

The overall program is operationalized on a desk computer configured with two NVIDIA GeForce RTX 3060s. For details, sun glint semantic segmentation networks are implemented based on PyTorch. Segmentation networks are trained with a learning rate of 0.0001 and the epoch number is 100. Momentum is set to 0.9 and weight decay is set to 0.0005. The $mIoU$ (Mean Intersection over Union) and IoU ($IoU_{background}$ represents the IoU (Intersection over Union) of the background class and $IoU_{sunglint}$ represents the IoU of the sun glint class) are used to evaluate the performance of sun glint detection. Structural Similarity Measure ($SSIM_{metric}$) is a metric employed for quantifying the similarity between two images, considering factors including luminance, contrast, and structural elements. A higher $SSIM_{metric}$ value indicates a higher similarity between the images, with an $SSIM_{metric}$ value of 1 denoting complete similarity between the two images [34]. Peak Signal-to-Noise Ratio ($PSNR_{metric}$) is a quantitative metric that measures the extent of distortion present between the reconstructed signal and the original signal, with higher $PSNR_{metric}$ values indicating lower levels of distortion [28]. In sun glint removal, $SSIM_{metric}$ and $PSNR_{metric}$ are used to quantitatively measure the similarity between the restored images and the ground truth thereby evaluating the sun glint removal performance.

4.1. Comparison of sun glint detection networks

The result of our FANet is evaluated and compared with several representative semantic segmentation networks (*i.e.*, UNet [24], DeepLabv3+ [4], HRNet [29], ConvNeXt [20] and one network developed specifically for sun glint detection (*i.e.*, UNetglint [11])) regarding the performance in sun glint detection. Table 1 shows that all methods exhibit high $mIoU$ values, primarily because the $IoU_{background}$ is consistently above 98%. The results differ when considering the $IoU_{sunglint}$: our FANet achieves the highest accuracy in both sun glint and background segmentation; UNet-related methods, namely UNet and UNetglint, outperform other classic semantic segmentation networks, with an $IoU_{sunglint}$ of over 75%. In comparison, other classic semantic segmentation methods, such as DeepLabv3+, HRNet, and ConvNeXt, exhibit $IoU_{sunglint}$ ranging between 56.53% and 62.73%, suggesting that UNet is an effective backbone for sun glint detection. Moreover, a visual inspection was carried out as shown in Figure S2 of the supplementary material. The results indicate that DeepLabv3+, HRNet, and ConvNeXt tend to identify sun glint edges that are larger than the true edges. Moreover, these methods may group different sun glint with close distances into a single sun glint, resulting in incorrect segmentation of many background pixels as sun glint. In contrast, UNet, UNetglint, and FANet exhibit more accurate sun glint boundaries, with a higher number of small sun glint being identified. One possible explanation is that UNet's design is optimized for situations with limited training data, which is not always the case for other classic semantic segmentation methods developed for image segmentation in the computer vision domain, where larger annotated datasets are more common. Additionally, UNet's use of skip connections, which combines multi-scale low-level and high-level features, preserves spatial information for small objects like sun glint, while other classic semantic segmentation methods may struggle to detect small objects. Thus, our results indicate the superiority of FANet in accurate sun glint detection, preparing the solid foundation for subsequent sun glint removal precisely.

Methods	$mIoU$	$IoU_{background}$	$IoU_{sunglint}$
UNet	87.57%	99.41%	75.74%
UNetglint	89.30%	99.48%	79.13%
DeepLabv3+	77.54%	98.55%	56.53%
HRNet	80.28%	98.70%	61.85%
ConvNeXt	80.78%	98.32%	62.73%
FANet	**90.43%**	**99.52%**	**81.34%**

Table 1. Comparison results for different deep learning models in sun glint detection. The highest values in $mIoU$, $IoU_{background}$ and $IoU_{sunglint}$ are shown in bold.

4.2. Evaluation of sun glint removal strategies

To evaluate our sun glint removal strategy, we created a specialized artificial dataset by artificially introducing sun glint into previously sun-glint-free regions of UAV coral images. We conducted experiment on this specialized artificial dataset and compared our result with the GAN-based method (*i.e.*, DeepFillv2) derived result and the conventional image interpolation method (*i.e.*, FMM) derived result. The visual comparison is depicted in Figure 5 while the quantitative evaluation results are summarized in Table 2. In Figure 5, the first row consists of our chosen UAV coral images without sun glint, the same images with artificial sun glint, and sun glint removal results generated by DeepFillv2, FMM and ours. The second row presents enlarged views of the red rectangular dashed boxes in the first row, enabling a comprehensive visual inspection of their differences. The visualized results of the three methods for image

	DeepFillv2	FMM	**Ours**
$SSIM_{metric}$	0.962	0.975	**0.998**
$PSNR_{metric}$	39.48	41.47	**54.95**

Table 2. Quantitative comparison of different removal methods. The highest value is indicated in bold.

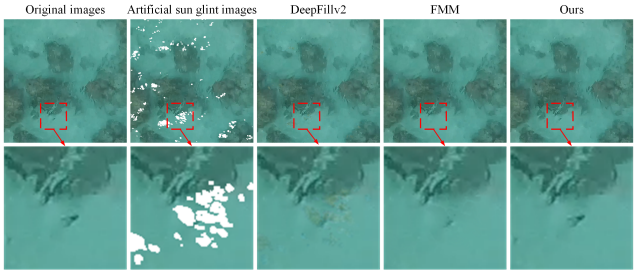

Figure 5. A visual comparison of our sun glint removal strategy with DeepFillv2 and FMM.

restoration at the sun-glint-dense area in Figure 5 reveal that our method achieves the richest texture and visually similar restoration to the original texture. DeepFillv2 produces new artificial textures that are inconsistent with real texture features, while FMM yields excessively smoothed outcomes, leading to the loss of fine texture details. We also provide more instances of sun glint image restoration results of our methods in Figure S3 of the supplementary material for further visual inspection. Moreover, we quantitatively compare their results with the original sun-glint-free images using $SSIM_{metric}$ and $PSNR_{metric}$ in Table 2, which consistently shows that our method produces the most realistic and best-detailed restoration results with the highest $SSIM_{metric}$ 0.998 and $PSNR_{metric}$ 54.95. Hence, our method effectively solves those problems as it clears almost all sun glint and restores images with real features, improving image quality.

4.3. Ablation study of FANet

To evaluate the effectiveness of the FAM module and hybrid loss function in detecting sun glint, we performed ablation experiments in this section. The baseline is the original UNet using the CE loss function, which aligns with the loss function used in the original UNet network. Then the UNet with FAM using CE loss function, as well as our proposed FANet combining the UNet with FAM and the hybrid loss function is evaluated. The results of these experiments presented in Table 3 confirm the effectiveness of our proposed improvement. Our introduction of the FAM module has resulted in a 3.36% improvement in $IoU_{sunglint}$ compared to the baseline, indicating that the FAM module guides the network's attention towards foreground information, thereby enabling the acquisition of more efficient and discriminative sun glint features. Furthermore, the use of the hybrid loss function has addressed class imbalance and further improved semantic segmentation accuracy. Ultimately, our FANet achieved a notable 5.6% improvement in $IoU_{sunglint}$ over the baseline, demonstrating its outstanding performance.

Methods	$mIoU$	$IoU_{background}$	$IoU_{sunglint}$
UNet	87.57%	99.41%	75.74%
UNet+FAM	89.28%	99.47%	79.10%
UNet+FAM+hybrid loss	**90.43%**	**99.52%**	**81.34%**

Table 3. The accuracy comparison of ablation study. The highest values in $mIoU$, $IoU_{background}$ and $IoU_{sunglint}$ are shown in bold.

Images	$mIoU$	$IoU_{background}$	IoU_{coral}
Sun glint images	65.24%	63.64%	66.85%
Sun-glint-free images	**78.78%**	**79.03%**	**78.91%**

Table 4. Comparative results of the impact of sun glint on the benthic coral semantic segmentation task. The highest values in $mIoU$, $IoU_{background}$ and IoU_{coral} are shown in bold.

4.4. Applications of sun-glint-free drone images

The utilization of oceanic drone RGB imagery for marine coastal ecological monitoring is the final goal of this paper. Within this domain, accurately identifying and analyzing benthic organisms, encompassing their species, distribution, and ecological responses based on these data is the focus. We conducted coral identification experiments specifically targeting the dominant benthic organism coral to quantitatively evaluate the negative implications of sun glint on marine coastal ecological monitoring tasks. The experimental setup involved applying pre-trained weights of the UNet model to classify coral and background on both original images with sun glint and images that underwent sun glint removal. The experimental results are presented in Table 4. Quantitatively, the removal of sun glint led to a notable improvement in the $mIoU$ by approximately 13.5% and IoU_{coral} by 12% for coral segmentation.

To facilitate visual inspection, we further provide visual comparison results. The visual comparison reveals that the presence of sun glint has a detrimental effect on the accuracy of the underwater coral segmentation results. It leads to considerable misclassifications, omissions and introduces numerous small artifacts and noise. However, through sun glint removal, these issues notably diminish, resulting in improved segmentation accuracy and sharper coral boundaries. This improvement in accuracy contributes to a more reliable data foundation and support for marine ecological monitoring. Therefore, our experiments demonstrate

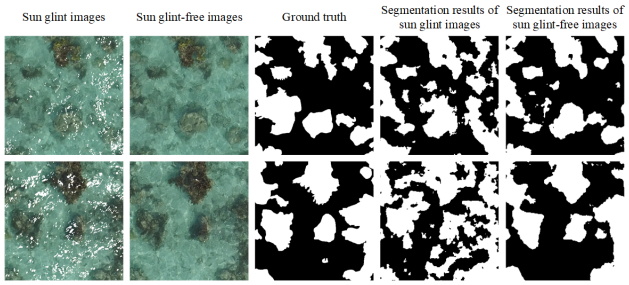

Figure 6. A visual comparison of the sun glint's impact on downstream benthic coral semantic segmentation task.

the detrimental impact of sun glint on the understanding of coral distribution, while highlighting the substantial quality improvement achieved through its removal.

The production of shallow water habitat maps utilizing coastal drone remote sensing imagery is another fundamental task in marine ecological monitoring. Given the restricted coverage area of a single high-resolution image captured by drones, the generation of large-scale mosaic images through photogrammetric processing of single drone images obtained from the survey area is imperative to facilitate a more comprehensive understanding and analysis. In this study, in order to demonstrate the effect of sun glint removal in producing benthic habitat maps, we have chosen a set of coastal drone images for stitching through image feature matching and pixel fusion. The image stitching results with and without sun glint removal are shown in Figure 7. In Figure 7(a), a considerable number of sun glint instances contaminate a substantial portion of the upper and right area, leading to those areas being unusable. Furthermore, sun glint poses challenges in image matching, evident by pronounced ghosting and blurring in the image. Conversely, the sun glint removal outcome in Figure 7(b) exhibits a distinct restoration of the shape and distribution of benthic coral substrates, accompanied by a notable improvement in mitigating image contamination. The findings demonstrate the significant impact of sun glint removal on producing benthic habitat maps, underscoring its crucial role in enabling accurate habitat mapping.

5. Conclusions

Sun glint is crucial to consider in remote sensing of habitats using high-resolution drone images of inland water or coastal ocean areas. It can contaminate images, making the object of interest indiscernible and compromising image processing and interpretation. Our study addresses this problem by proposing a pipeline with a Foreground Attention-based sun glint detection module and an optical-flow-based removal strategy. The Foreground Attention Module enhances the detection of sun glint characteristics,

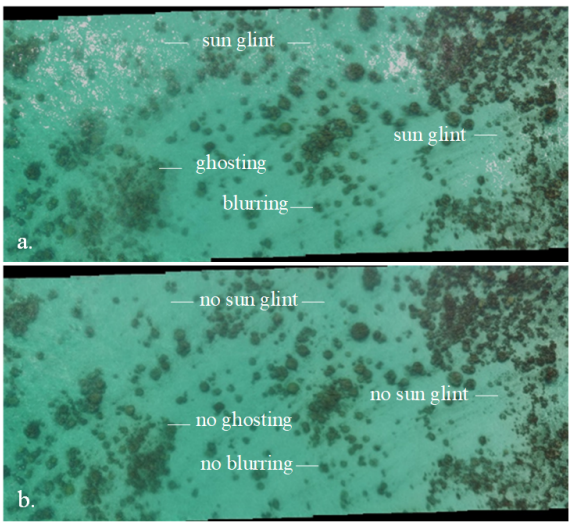

Figure 7. Comparison of stitched images. (a) is the original image stitching result without sun glint removal. (b) is the image stitching result with sun-glint-free images.

improving identification and classification accuracy. Experimental results show exceptional sun glint segmentation performance. However, segmentation alone doesn't solve image contamination. Sun glint's distribution area is not extensive but can still impact a wide range of images. This poses a challenge for refined habitat identification and classification based on remote sensing imagery. For instance, when identifying and classifying individual corals, certain pillar corals may have a limited spatial distribution on the plane but possess a significant three-dimensional presence. These corals play a crucial role in the ecosystem and their significance could rival that of dwarf corals with a larger distribution area. Consequently, even a small number of random sun glint pixels contaminating the image can render the identification and classification task nearly impossible. For this, we propose an optical-flow-based strategy to restore the missing sun glint area, effectively clearing sun glint and restoring accurate image features. Our pioneering techniques for sun glint correction significantly enhance the accuracy of information extraction and mapping tasks in marine ecological monitoring.

6. Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number: NSFC-41901407), ETH Postdoctoral Research Project: ETH MOOREA, and German Federal Ministry for Economic Affairs and Climate Action Project (grant number:50EE2201C).

References

[1] Guillaume Brunier, Emma Michaud, Jules Fleury, Edward J Anthony, Sylvain Morvan, and Antoine Gardel. Assessing the relationship between macro-faunal burrowing activity and mudflat geomorphology from uav-based structure-from-motion photogrammetry. *Remote Sensing of Environment*, 241(2020):111717–111734, 2020. 1

[2] Cox Charles and Walter Munk. Measurement of the roughness of the sea surface from photographs of the sun's glitter. *Josa*, 44(11):838–850, 1954. 1

[3] Jun Chen, Xianqiang He, Zhongli Liu, Nan Lin, Qianguo Xing, and Delu Pan. Sun glint correction with an inherent optical properties data processing system. *International Journal of Remote Sensing*, 42(2):617–638, 2021. 1

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, pages 801–818, 2018. 2, 6

[5] Wei Sheng Chong, Nurul Hidayah Mat Zaki, Mohammad Shawkat Hossain, Aidy M Muslim, and Amin Beiranvand Pour. Introducing theil-sen estimator for sun glint correction of drone data for coral mapping. *Geocarto International*, 37(15):4527–4556, 2022. 2

[6] Aijun Cui, Jingyu Zhang, Yi Ma, and Xi Zhang. A noise de-correlation based sun glint correction method and its effect on shallow bathymetry inversion. *Remote Sensing*, 14(23):5981–6003, 2022. 2, 3

[7] Puhong Duan, Jibao Lai, Jian Kang, Xudong Kang, Pedram Ghamisi, and Shutao Li. Texture-aware total variation-based removal of sun glint in hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166(2020):359–372, 2020. 1

[8] James P Duffy, Andrew M Cunliffe, Leon DeBell, Chris Sandbrook, Serge A Wich, Jamie D Shutler, Isla H Myers-Smith, Miguel R Varela, and Karen Anderson. Location, location, location: considerations when using lightweight drones in challenging environments. *Remote Sensing in Ecology and Conservation*, 4(1):7–19, 2018. 1

[9] Vahtmäe Ele, Kutser Tiit, Martin Georg, and Kotta Jonne. Feasibility of hyperspectral remote sensing for mapping benthic macroalgal cover in turbid coastal waters—a baltic sea case study. *Remote Sensing of Environment*, 101(3):342–351, 2006. 1

[10] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *CVPR*, pages 3146–3154, 2019. 2

[11] Anna B Giles, James Edward Davies, Keven Ren, and Brendan Kelaher. A deep learning algorithm to detect and classify sun glint from high-resolution aerial imagery over shallow marine environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181(2021):20–26, 2021. 2, 6

[12] Eric J Hochberg, Andréfouët Serge, and Tyler R Misty. Sea surface correction of high spatial resolution ikonos images to improve bottom mapping in near-shore environments. *IEEE transactions on geoscience and remote sensing*, 41(7):1724–1729, 2003. 1

[13] Yuewu Hou, Zhaoying Liu, Ting Zhang, and Yujian Li. Cunet: Complement unet for remote sensing road extraction. *Sensors*, 21(6):2153–2174, 2021. 2

[14] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 2, 4

[15] Martin Javier, Eugenio Francisco, Marcello Javier, and Medina Anabella. Automatic sun glint removal of multispectral high-resolution worldview-2 imagery for retrieving coastal shallow water parameters. *Remote Sensing*, 8(1):37–53, 2016. 1

[16] Shihao Jiang, Dylan Campbell, Yao Lu, Hongdong Li, and Richard Hartley. Learning to estimate hidden motions with global motion aggregation. In *ICCV*, pages 9772–9781, 2021. 3, 5

[17] Joyce KE, Stephanie Duce, Susannah M Leahy, Javier Xavier Leon, and Maier SW. Principles and practice of acquiring drone-based image data in marine environments. *Marine and Freshwater Research*, 70(7):952–963, 2018. 1

[18] Brendan P Kelaher, Andrew P Colefax, Alejandro Tagliafico, Melanie J Bishop, Anna Giles, and Paul A Butcher. Assessing variation in assemblages of large marine fauna off ocean beaches using drones. *Marine and Freshwater Research*, 71(1):68–77, 2019. 1

[19] Zhengmin Kong, Zhuolin Fu, Feng Xiong, and Chenggang Zhang. Foreground feature attention module based on unsupervised saliency detector for few-shot learning. *IEEE Access*, 9(2021):51179–51188, 2021. 4

[20] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *CVPR*, pages 11976–11986, 2022. 2, 6

[21] David R Lyzenga, Norman P Malinas, and Fred J Tanis. Multispectral bathymetry using a simple physically based algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8):2251–2259, 2006. 1

[22] Jun Ma, Jianan Chen, Matthew Ng, Rui Huang, Yu Li, Chen Li, Xiaoping Yang, and Anne L Martel. Loss odyssey in medical image segmentation. *Medical Image Analysis*, 71(2021):102035–102048, 2021. 4

[23] Aidy M Muslim, Wei Sheng Chong, Che Din Mohd Safuan, Idham Khalil, and Mohammad Shawkat Hossain. Coral reef mapping of uav: A comparison of sun glint correction methods. *Remote Sensing*, 11(20):2422–2446, 2019. 1

[24] Ronneberger Olaf, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015. 2, 6

[25] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *CVPR*, pages 7479–7489, 2019. 4

[26] Mount Richard. Acquisition of through-water aerial survey images: Surface effects and the prediction of sun glitter and subsurface illumination. *Photogrammetric Engineering and Remote Sensing*, 71(12):1407–1415, 2005. 1

[27] Joseph A Shaw and James H Churnside. Scanning-laser glint measurements of sea-surface slope statistics. *Applied optics*, 36(18):4202–4213, 1997. 1

[28] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006. 6

[29] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *CVPR*, pages 5693–5703, 2019. 2, 6

[30] Kay Susan, John D Hedley, and Samantha Lavender. Sun glint correction of high and low spatial resolution images of aquatic scenes: a review of methods for visible and near-infrared wavelengths. *Remote sensing*, 1(4):697–730, 2009. 1

[31] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *ECCV*, pages 402–419, 2020. 3

[32] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1):23–34, 2004. 2, 6

[33] Menghua Wang and Sean W Bailey. Correction of sun glint contamination on the seawifs ocean and atmosphere products. *Applied Optics*, 40(27):4790–4798, 2001. 1

[34] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 4, 6

[35] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *ECCV*, pages 3–19, 2018. 2

[36] Rui Xu, Xiaoxiao Li, Bolei Zhou, and Chen Change Loy. Deep flow-guided video inpainting. In *CVPR*, pages 3723–3732, 2019. 3

[37] Qiumeng Xue and Li Guan. Identification of sun glint contamination in gmi measurements over the global ocean. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6473–6483, 2019. 1

[38] Zongyao Yang, Xueying Yu, Simon Dedman, Massimiliano Rosso, Jingmin Zhu, Jiaqi Yang, Yuxiang Xia, Yichao Tian, Guangping Zhang, and Jingzhen Wang. Drone remote sensing applications in marine monitoring: Knowledge visualization and review. *Science of The Total Environment*, 838(1):155939–155962, 2022. 1

[39] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *ICCV*, pages 4471–4480, 2019. 2

[40] Hao Zhang and Menghua Wang. Evaluation of sun glint models using modis measurements. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 111(3):492–506, 2010. 1

[41] Jie Zhao, Yu Li, Patrick Matgen, Ramona Pelich, Renaud Hostache, Wolfgang Wagner, and Marco Chini. Urban-aware u-net for large-scale urban flood mapping using multitemporal sentinel-1 intensity and interferometric coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 60(2022):1–21, 2022. 2

[42] Yuliang Zou, Zelun Luo, and Jia-Bin Huang. Df-net: Unsupervised joint learning of depth and flow using cross-task consistency. In *ECCV*, pages 36–53, 2018. 5