

Marine Snow Removal Using Internally Generated Pseudo Ground Truth

Alexandra Malyugina¹, Guoxi Huang¹, Eduardo Ruiz², Benjamin Leslie² and Nantheera Anantrasirichai¹

¹*Visual Information Laboratory, University of Bristol, Bristol, UK*

²*Beam, Bristol, UK*

alex.malyugina@bristol.ac.uk, guoxi.huang@bristol.ac.uk, er19@beam.global,

bl18@beam.global, n.anantrasirichai@bristol.ac.uk

Abstract—Underwater videos often suffer from degraded quality due to light absorption, scattering, and various noise sources. Among these, marine snow—suspended organic particles appearing as bright spots or noise—significantly impacts machine vision tasks, particularly those involving feature matching. Existing methods for removing marine snow are ineffective due to the lack of paired training data. To address this challenge, this paper proposes a novel enhancement framework that introduces a new approach to generating paired datasets from raw underwater videos. The resulting dataset consists of paired images of generated snowy and snow-free underwater videos, enabling supervised training for video enhancement. We describe the dataset creation process, key characteristics, and demonstrate its effectiveness in enhancing underwater image restoration with the absence of groundtruth.

Index Terms—Underwater video restoration, video enhancement, marine snow, datasets.

I. INTRODUCTION

Video enhancement plays an important role in many computer vision applications, improving the visibility of objects and features under challenging conditions. It is also an important preprocessing step for tasks such as object detection, tracking, and simultaneous localization and mapping (SLAM), where degraded visual quality can significantly affect algorithm performance. In underwater environments, video degradation may be caused by light absorption and scattering, as well as various types of noise and marine snow. This makes video enhancement a critical preprocessing step for marine robotics, and autonomous vehicles.

Among underwater visibility issues, marine snow is a common and severe problem that can affect the performance of machine vision tasks such as SLAM. Marine snow consists of particles of near-neutral buoyancy suspended in the water column. When collecting video data with artificial illumination marine snow appears as bright spots or noise in the video, leading to quality degradation and obstruction of key features.. Marine snow is highly variable and difficult to model, making its removal challenging [1]. Traditional methods for mitigating marine snow include histogram manipulation and filtering techniques, model-based methods and machine learning-based approaches. However, these approaches have limitations: filtering methods may remove valuable details, histogram methods

are hard to tune for optimal performance and deep learning based approaches, despite promising results, often rely on synthetic data that fails to capture the complexity of real-world distortions.

Comprehensive training data is a key component of deep learning based methods. Obtaining high-quality paired data is particularly difficult in underwater settings, where complex conditions and lack of ground truth complicate dataset creation. To address these limitations, we propose a simple dataset creation method that transforms raw, unpaired data captured at various depths into a paired dataset without the need for ground truth. Our approach is based on straightforward but efficient frame blending to simulate paired data directly from raw underwater footage with minimal manual annotation effort. Moreover, we propose an underwater video enhancement method designed to improve SLAM performance in low-light environments affected by marine snow. Unlike previous approaches that discard or filter the keypoints, our method enhances entire video frames, preserving feature-rich regions while suppressing noise.

The key contributions of the paper are:

- A video enhancement method for marine snow mitigation, improving SLAM performance without explicit keypoint filtering.
- A data extraction approach that eliminates the need for paired or manually cleaned data, using raw underwater video captured at different depths for training.
- A demonstration of the method's effectiveness, showing that enhancing entire video frames improves feature tracking and 3D reconstruction.

II. RELATED WORK

Numerous methods have been developed to address distortions caused by atmospheric and underwater particles in videos. Early approaches could predominantly be applied to a single image and relied on handcrafted features such as guidance images, histogram of gradients [2]–[5] and filtering techniques [6], [7], to separate particles from the background. However these methods come at a cost of oversmoothing and losing details or do not guarantee the removal of the particles and struggle with temporal consistency when applied to a video sequence. More recent atmospheric snow removal methods are based on Gaussian Mixture Models [8], Maximum

This work is supported by EPSRC IAA, University of Bristol, and EPSRC ECR (EP/Y002490/1).

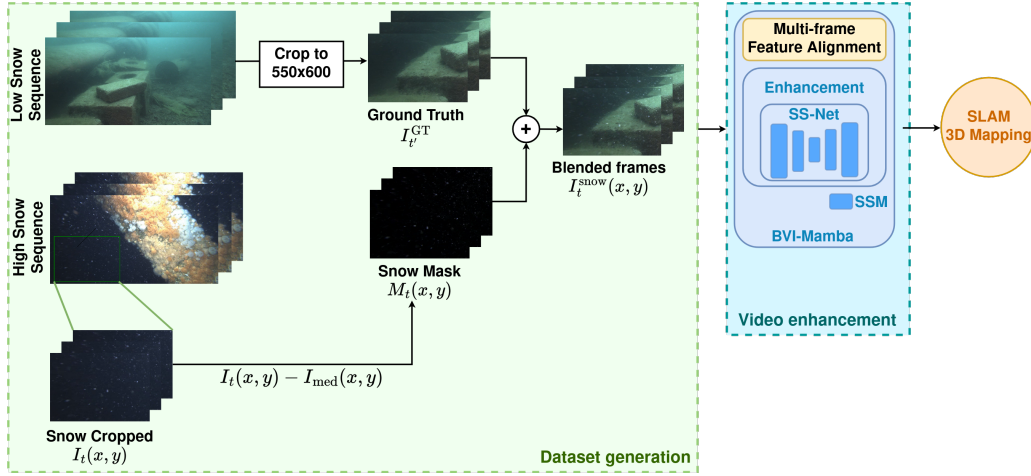


Fig. 1. Diagram of the proposed framework

a Posteriori framework [9]. There are also some recent deep learning approaches, such as contrastive learning models with temporal aggregation [10], which, like many others, rely on synthetic data [10], [11].

Despite atmospheric snow removal having been extensively studied for both images and videos, these methods do not fully address the challenges of marine snow removal. Video atmospheric snow removal methods assume that snowflakes, being ruled by gravitation, move in relatively uniform downward trajectories. Underwater snow is composed of suspended organic and inorganic particles and exhibits complex and non-uniform motion influenced by water currents and turbulence. These particles are often large and dense and do not follow simple gravitational paths but move in multiple directions, making their removal considerably more challenging.

Numerous methods have been proposed for image-based underwater enhancement [1], [12]–[15], with some methods focusing on supervised learning trained on synthetic data generated by manual pairwise voting on enhanced images [13] or a generative network (GAN) [1]. However, research on marine snow removal in underwater videos remains limited.

Cyganek and Gongola [16] developed a real-time filtering method based on spatiotemporal patch analysis and 3D median filtering, which is effective for remotely operated vehicles (ROVs) but relies on handcrafted rules, limiting adaptability to diverse marine snow conditions. More recently, Coffelt et al. [11] proposed a fully synthetic dataset for marine snow simulation and trained a UNet-based deep learning model for snow removal. While this synthetic data provides controlled training conditions, the lack of real-world marine snow samples may result in poor generalization. Hodne et. al in [17] focus on filtering keypoints for SLAM by classifying marine snow in keypoint detection, which requires integrating keypoint rejection into SLAM pipelines and training classifiers. The data generation process involves manual collection of marine snow samples from specific conditions and then superimposing them onto background.

Unlike these methods, our approach uses real underwater video sequences to extract marine snow masks via median frame subtraction, enabling supervised training on realistic data while maintaining temporal consistency.

III. METHODOLOGY

The overview of our framework is shown in Fig. 1, consisting of dataset generation and video enhancement modules.

A. Dataset Generation Module

We introduce a new design to generate a paired dataset for marine snow removal in underwater videos. This dataset enables supervised learning by providing aligned image pairs, where one frame is affected by marine snow, and the other is used as a clean reference. This process can be applied for footage made in different conditions, generating customised datasets that suit the needs of particular applications. Our dataset generation process is as follows:

First, we manually select underwater video sequences with minimal marine snow from the data available to us, which serve as our clean reference frames (ground truth). These sequences contain minimal visual noise, making sure they represent the true scene structure.

Then, we identify sequences that contain visible marine snow but do not have moving objects in the background. However, such sequences were not readily available in our dataset. To overcome this, we extract static background regions from snow-affected frames by selecting fixed-size patches of 550×600 pixels. These patches are chosen to ensure they contain snow particles while avoiding any foreground objects. To isolate the snow, we compute a median frame across each snow sequence and subtract it from each frame, producing a set of dynamic snow masks that capture the movement of marine snow particles. Given a sequence of N frames, denoted as $I_t(x, y)$ where $t \in \{1, 2, \dots, N\}$, the median frame $I_{med}(x, y)$ is computed as:

$$I_{med}(x, y) = \text{median}(I_1(x, y), I_2(x, y), \dots, I_N(x, y)). \quad (1)$$

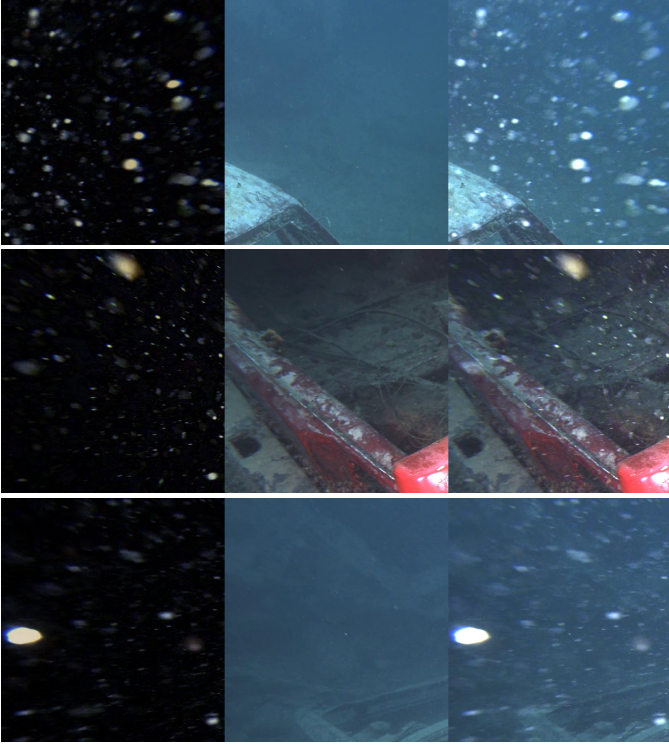


Fig. 2. First column: extracted snow mask; second column: ground truth image; last column: resulting snow-affected image with overlaid snow mask.

The snow mask sequence $M_t(x, y)$ is then obtained by subtracting the median frame from each individual frame:

$$M_t(x, y) = I_t(x, y) - I_{\text{med}}(x, y). \quad (2)$$

This operation removes the static background, leaving only the dynamic sequence with snow particles. Then, $M_t(x, y)$ is overlaid onto the clean reference frames $I_{\text{GT}}^{t'}$ to generate snow-affected frames $I_{\text{snow}}^t(x, y)$. Since the extracted snow mask sequences may be shorter than the reference sequences, we apply a sliding window approach. The temporal alignment of the mask with the reference sequence is randomized, introducing natural variations in motion patterns. Additionally, the spatial position of the snow mask is randomized across the reference frame, ensuring variability in how marine snow appears in different locations:

$$I_t^{\text{snow}}(x, y) = I_{t'}^{\text{GT}}(x + \Delta_x, y + \Delta_y) + M_t(x, y), \quad (3)$$

where Δ_x, Δ_y are spatial offsets used to randomize the overlay position. The temporal index t' is randomly selected within a sliding window, ensuring variation.

As a result of this process, we generate a dataset of total 83,000+ paired frames, where each image pair consists of a snow-affected frame and a corresponding clean frame. The generated dataset contains 300 training sequences having a resolution of 550×600 pixels and covering a wide range of marine snow conditions and 10 testing sequences for evaluation. Fig. 2 shows the example frames from the dataset.

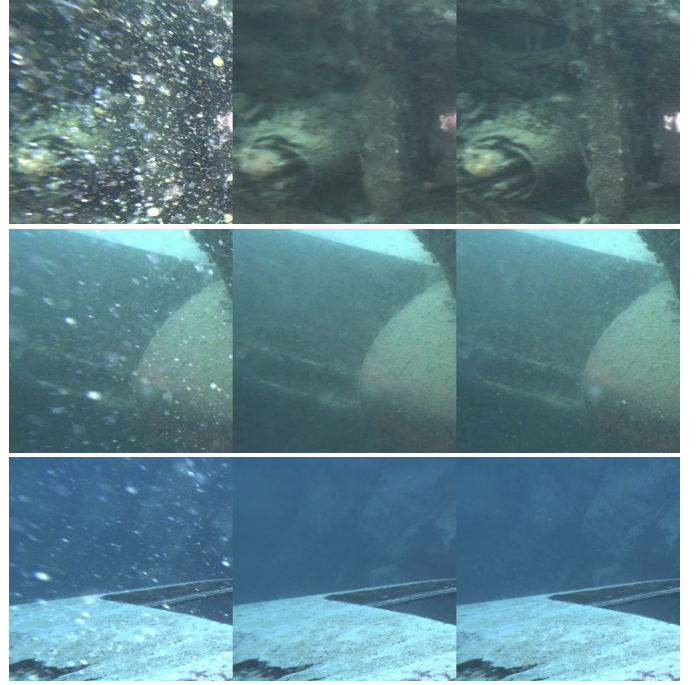


Fig. 3. Subjective results on test set. First column: Raw input image with real marine snow; second column: result; last column: ground truth image.

B. Video Enhancement Module

We employ BVI-Mamba [18], an architecture that incorporates feature-level frame alignment and integrates State Space Models (SSMs) with the 2D Selective Scan (SS2D) module [19]. This design efficiently models long-range dependencies while maintaining a computationally efficient framework. The model captures temporal correlations across frames with linear complexity, making it effective for handling long sequences while preserving fine spatial details.

IV. EXPERIMENTS AND RESULTS

A. Datasets and Training Setup

Underwater sequences were provided by Beam¹, a company that uses underwater robots to service offshore wind farms. These sequences are in RGB format with resolution 2048×1080 , captured using Beam's mapping system in both commercial and controlled testing environments. Each sequence includes inspections of submerged structures, such as jacket members, anodes, and a fuselage, with durations ranging from 8.5 minutes to 2 hours 39 mins. For our experiments, shorter 1-2-minute clips are extracted from the full sequences.

Eight sequences for task-specific performance evaluation, each containing between 1,241 and 2,841 frames, were kept separate from the main dataset for further task-based evaluation.

For training, we used 512×512 patch size, input sequence length of 5 frames, learning rate of 0.0001 and l_1 loss. The training results on the test set are shown in Fig. 3.

¹<https://beam.global/>

B. Performance Metrics

Evaluating enhancement performance with a generated dataset like ours (even if that is derived from real-world data) is challenging. Since our paired data is derived from frame blending, the “ground truth” is only the cleanest available data, not a truly clean underwater scene. This makes it difficult to assess whether enhancement restores true lost details or introduces artifacts. Also, standard objective image quality metrics like PSNR and SSIM that rely on pixel-wise differences may not provide meaningful insights into real-world performance. Pixel-level performance does not necessarily correlate with the performance of downstream tasks, like SLAM, which depend on preserving key visual features rather than exact pixel-level reconstruction.

Therefore, a better approach is to assess the impact of video enhancement on SLAM performance. Since SLAM relies on feature tracking, matching, spatial and temporal consistency, improvements in these aspects provide a more relevant measure of the quality of the enhancement. If an enhancement method improves SLAM robustness in challenging conditions under water, it is likely more useful than one that optimizes PSNR, SSIM or other objective pixel-level metrics on generated test data.

For performance evaluation we chose three task-relevant metrics: the number of keypoints per frame in SLAM, the number of frame-to-frame feature matches in SLAM and the number of points in the reconstructed 3D-models (dense mapping).

C. Results

We compared the performance of Beam’s proprietary mapping system, which supports real-time, high-resolution, and globally consistent mapping; the system includes a feature extraction and matching module used as part of the mapping pipeline. The SLAM and dense mapping systems are run on the original sequences without enhancement, the sequences enhanced with the Bayesian enhancement model [15] and our method. We chose 8 real video sequences to evaluate the performance. The sequences starting with 2022_ are affected by heavy marine snow, whereas the sequences starting with 2023_ have low levels of snow.

Fig. 4 shows the number of keypoints per frame, where high numbers indicate high levels of snow. Overall, our results show a lower number of keypoints compared to sequences without enhancement and those enhanced with BEM. That suggests that our method may suppress noise caused by snow particles. The example of the feature point detection is shown in Fig. 5. Note how for our method, features are focused mainly on the seabed and cable regions rather than snow particles.

Our method also consistently achieves a higher number of frame-to-frame matches, as shown in Fig. 6, that means the preserved or enhanced features are more consistent across frames, improving feature tracking stability in SLAM. This suggests that although fewer keypoints are detected, they are of higher quality and better suited for matching.

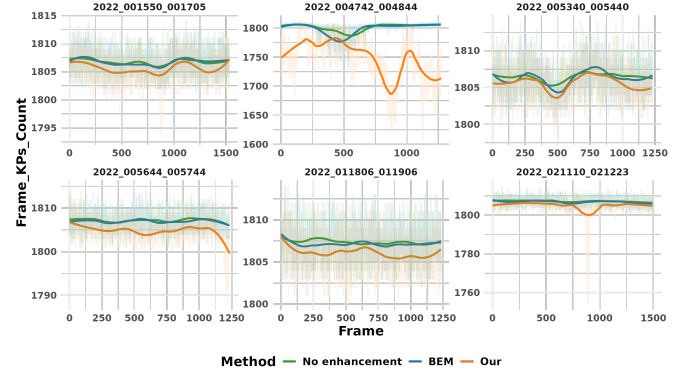


Fig. 4. Number of keypoints per frame (lower is better, indicating fewer snow particles). Smoothing was applied to the data for clarity.

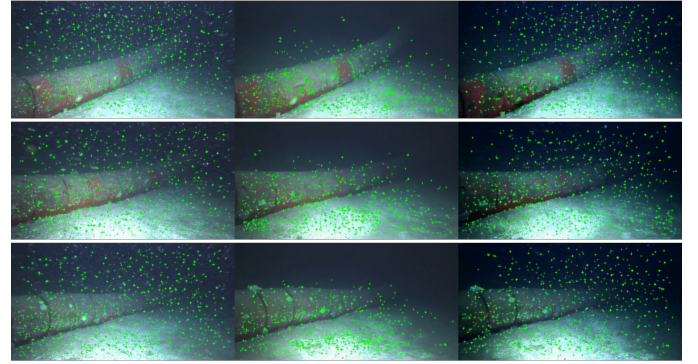


Fig. 5. Example frames and feature detection (feature points shown in green) for sequence 2022_004742_004844. Feature detection is performed on left column: no enhancement; central column: our method; right column: BEM.

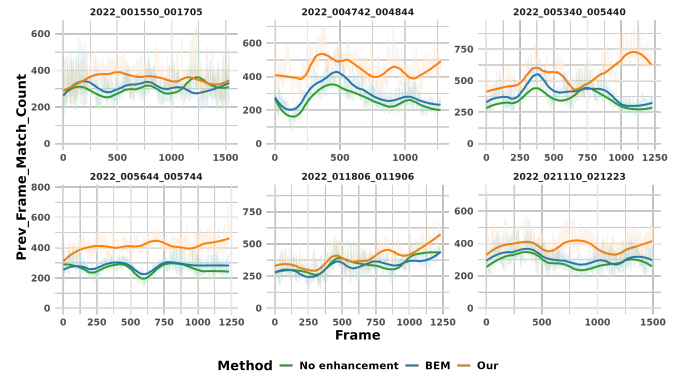


Fig. 6. Number of frame-to-frame feature matches (higher is better, indicating more useful features for SLAM). Smoothing was applied to the data for clarity.

Table I demonstrates the result for average number of points in dense mapping. Our approach results in a higher average number which demonstrates improved scene reconstruction quality. This implies that the enhancement reduces noise and artifacts that would otherwise disrupt dense geometry estimation, leading to more reliable 3D models. Figure 7 shows the example of the reconstructed 3D model (point cloud) for the sequence 2022_004742_004844. Our method shows details not appearing on the reconstructed model obtained from a non-

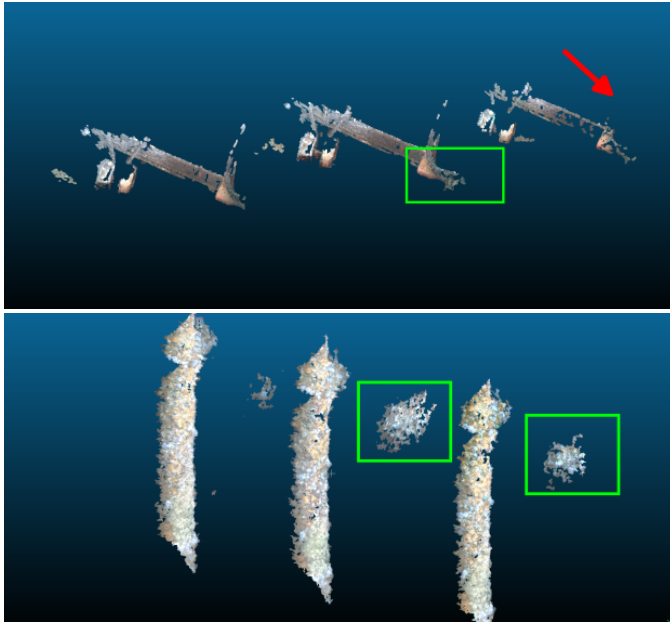


Fig. 7. 3D models (point cloud) reconstructed for sequence 2023_000015_000135 (top) and sequence 2022_005644_005744. Methods order (left to right): no enhancement, our method, BEM. The red arrow highlights key differences (missing parts or sparsity) of the enhanced video with respect to the non-enhanced baseline. The green box highlights improved reconstruction with respect to the baseline

TABLE I
AVERAGE NUMBER OF POINTS PER METHOD IN 3D-RECONSTRUCTION

Method	Average Num of Points
No Enhancement	893,076
BEM	802,084.6
Our	1,142,521.5

enhanced sequence and sequence processed with BEM.

V. CONCLUSION

This paper introduced a novel dataset generation approach and a video enhancement method for marine snow removal in underwater videos. The framework generates a paired dataset from real snow artefacts, providing correct characteristics of distortion for the enhancement network. Our results show that the proposed method improves SLAM performance by enhancing key visual features while preserving spatial and temporal consistency. This work provides a practical solution for underwater vision tasks affected by marine snow.

REFERENCES

- [1] F. Galetto and G. Deng, "A deep learning approach for marine snow synthesis and removal," *Signal, Image and Video Processing*, vol. 19, no. 1, p. 1, 2025.
- [2] J. Xu, W. Zhao, P. Liu, and X. Tang, "An improved guidance image based method to remove rain and snow in a single image," *Computer and Information Science*, vol. 5, no. 3, p. 49, 2012.

- [3] S. Yu, Y. Zhao, Y. Mou, J. Wu, L. Han, X. Yang, and B. Zhao, "Content-adaptive rain and snow removal algorithms for single image," in *Advances in Neural Networks-ISNN 2014: 11th International Symposium on Neural Networks, ISNN 2014, Hong Kong and Macao, China, November 28-December 1, 2014. Proceedings 11*. Springer, 2014, pp. 439–448.
- [4] S.-C. Pei, Y.-T. Tsai, and C.-Y. Lee, "Removing rain and snow in a single image using saturation and visibility features," in *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 2014, pp. 1–6.
- [5] J. Bossu, N. Hautiere, and J.-P. Tarel, "Rain or snow detection in image sequences through use of a histogram of orientation of streaks," *International journal of computer vision*, vol. 93, pp. 348–367, 2011.
- [6] X. Ding, L. Chen, X. Zheng, Y. Huang, and D. Zeng, "Single image rain and snow removal via guided l0 smoothing filter," *Multimedia Tools and Applications*, vol. 75, pp. 2697–2712, 2016.
- [7] J. Xu, W. Zhao, P. Liu, and X. Tang, "Removing rain and snow in a single image using guided filter," in *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, vol. 2. IEEE, 2012, pp. 304–307.
- [8] B. Yang, Z. Jia, J. Yang, and N. K. Kasabov, "Video snow removal based on self-adaptation snow detection and patch-based gaussian mixture model," *IEEE Access*, vol. 8, pp. 160 188–160 201, 2020.
- [9] M. Li, X. Cao, Q. Zhao, L. Zhang, and D. Meng, "Online rain/snow removal from surveillance videos," *IEEE Transactions on Image Processing*, vol. 30, pp. 2029–2044, 2021.
- [10] H. Chen, J. Ren, J. Gu, H. Wu, X. Lu, H. Cai, and L. Zhu, "Snow removal in video: A new dataset and a novel method," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2023, pp. 13 165–13 176.
- [11] J. P. Coffelt, N. Nowald, and P. Kampmann, "Marine snow simulation and elimination in video," in *2023 IEEE Underwater Technology (UT)*. IEEE, 2023, pp. 1–10.
- [12] Y. Wang, X. Yu, D. An, and Y. Wei, "Underwater image enhancement and marine snow removal for fishery based on integrated dual-channel neural network," *Computers and Electronics in Agriculture*, vol. 186, p. 106182, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S016816992100199X>
- [13] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2020.
- [14] Y. Li, B. Niu, T. Zhou, and H. Gao, "Underwater image restoration based on marine snow removal techniques," in *2024 2nd International Conference on Signal Processing and Intelligent Computing (SPIC)*, 2024, pp. 651–655.
- [15] G. Huang, N. Anantrasirichai, F. Ye, Z. Qi, R. Lin, Q. Yang, and D. Bull, "Bayesian neural networks for one-to-many mapping in image enhancement," *arXiv preprint arXiv:2501.14265*, 2025.
- [16] B. Cyganek and K. Gongola, "Real-time marine snow noise removal from underwater video sequences," *Journal of Electronic Imaging*, vol. 27, no. 4, pp. 043 002–043 002, 2018.
- [17] L. M. Hodne, E. Leikvoll, M. Yip, A. L. Teigen, A. Stahl, and R. Mester, "Detecting and suppressing marine snow for underwater visual slam," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5101–5109.
- [18] R. Lin, N. Anantrasirichai, G. Huang, J. Lin, Q. Sun, A. Malyugina, and D. R. Bull, "Bvi-rlv: A fully registered dataset and benchmarks for low-light video enhancement," *arXiv preprint arXiv:2407.03535*, 2024.
- [19] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.