Marine Snow Removal in High-Resolution Underwater Videos Using U-Net Segmentation and Inpainting

Jung-kyu Park[1], Giseop Noh[2]

[1]BE Course, Department of Software and Communications Engineering, Hongik University, South Korea

[2]Professor, Department of Software and Communications Engineering, Hongik University, South Korea

Corresponding author: Giseop Noh (kafa46@hongik.ac.kr)

Abstract

In this study, we conducted experiments using U-Net segmentation and LaMa-based inpainting to remove marine snow from high-resolution underwater videos. Underwater videos often suffer from visual degradation caused by light scattering, absorption, and suspended particles. In particular, marine snow appears as small bright dots or blurred spots, reducing the accuracy of subsequent tasks such as object detection and structural analysis. To preserve the detailed information of the original 4K videos, the frames were divided into 512×512 tiles, and a U-Net-based segmentation model was trained to detect marine snow regions as binary masks. The generated masks were then used as input to the LaMa inpainting model to restore the particle regions with surrounding background information. The experimental results showed that tile-based U-Net training is applicable for detecting small marine snow regions in high-resolution underwater videos. However, background noise with visual characteristics similar to marine snow, such as barnacles on underwater structures and bright reflection regions, was also detected. In addition, the LaMa-based inpainting method showed limitations in naturally restoring pixel-level small particle regions. When undetected residual particles remained around the masked region, the restored area was sometimes regenerated in a form similar to marine snow. Therefore, this study confirms the applicability and limitations of U-Net-based marine snow detection and LaMa-based inpainting removal in high-resolution underwater videos. The results suggest that further research is needed on temporal-information-based detection methods and restoration models specialized for marine snow removal.

Keywords: Marine Snow, U-Net, segmentation, inpainting, Underwater Video

1. Introduction

Recently, the investigation and analysis of underwater environments have become increasingly important in various fields, including marine structure inspection, underwater ecosystem observation, marine resource exploration, and the operation of underwater robots and drones. In particular, advances in underwater cameras and equipment such as ROVs (Remotely Operated Vehicles) and AUVs (Autonomous Underwater Vehicles) have made it possible to acquire high-resolution underwater images. Such image data are used as essential resources for the precise analysis of marine environments.

However, unlike images captured in terrestrial environments, underwater images suffer from various types of visual degradation due to the characteristics of water as a medium. Light scattering and absorption reduce image contrast and sharpness, while fine particles or suspended matter drifting underwater appear in images as bright spots, stains, blurring, or snow-like patterns. This phenomenon is generally referred to as marine snow, and it is a major factor that degrades the visibility of underwater images and reduces the accuracy of subsequent processing tasks such as object detection, structure analysis, and image restoration [1].

Marine snow does not have consistent size, brightness, or shape within images, and it appears in various forms depending on camera motion, lighting conditions, and underwater turbidity [2]. In particular, in videos captured by a moving camera such as an underwater drone, background structures and suspended particles may move together or overlap visually. Therefore, it is difficult to

 ISSN (Online): 3140-5312

stably remove only suspended particles using simple filtering or threshold-based methods. In addition, excessive image restoration may damage the texture of background structures, making it important to balance suspended particle removal and preservation of the original image.

Existing studies on underwater image enhancement have mainly focused on methods such as color correction, contrast enhancement, and denoising. Although these methods can be effective in improving overall image quality, they have limitations in precisely detecting and removing marine snow located at specific positions in an image. Recently, deep learning-based removal studies have been conducted; however, when suspended particles in underwater images have irregular shapes and are difficult to distinguish from the background, it remains difficult to obtain stable removal results [2].

In this study, a pipeline-based approach consisting of segmentation and inpainting is experimentally applied to remove marine snow present in underwater images. First, a segmentation model is used to extract marine snow regions as masks. Then, based on the extracted masks, an inpainting method is applied to naturally restore the regions where suspended particles were present using the surrounding background. Through this process, this study aims to construct an underwater image restoration method that explicitly detects and removes suspended particle regions, rather than simply enhancing the image.

The purpose of this study is to specifically identify the major problems that occur when attempting to remove marine snow from real underwater images. In the experimental process, variations in the size and brightness of suspended particles, similarity to background structures, temporal instability caused by camera motion, and background damage caused by excessive restoration were analyzed. Based on this analysis, this study summarizes the limitations of existing image processing and deep learning-based restoration methods in real underwater environments and suggests directions for improvement toward the development of more stable marine snow removal techniques.

2. Related Work and Limitations

Marine snow appearing in underwater images takes the form of small dots, bright particles, or blurred stains, degrading both the visual quality of images and the performance of subsequent analysis tasks [2]. To mitigate this problem, filtering-based methods and segmentation-based detection methods can be used in existing studies and image processing procedures. Filtering-based methods aim to reduce noise across the entire image, whereas segmentation-based methods aim to detect suspended particle regions to be removed at the pixel or object level. However, in real underwater images, both approaches have limitations because the size and shape of marine snow are irregular, and visually similar elements often coexist with background structures.

The median filter is a representative nonlinear filter widely used for noise removal in OpenCV-based image processing. However, since the median filter is applied uniformly to the entire image, it cannot selectively remove only marine snow regions. When the kernel size of the filter is set to a large value, small suspended particles or dot-like noise can be partially reduced, but the boundaries and fine textures of background structures may also be damaged [3]. In particular, the contours of underwater structures, surface textures, and fine patterns in the seafloor background can serve as important information for subsequent analysis. Therefore, excessive filtering may cause a decrease in image sharpness and blurring of boundaries. Conversely, when the kernel size is set to a small value, the detailed information of the original image is relatively well preserved, but marine snow that appears small or blurred may not be sufficiently removed. In addition, in underwater videos captured by a moving camera, the positional relationship between suspended particles and the background changes from frame to frame. Thus, it is difficult to obtain temporally stable removal results using simple spatial filtering alone. Therefore, although the median filter can be used to reduce noise in underwater images, it has limitations as a method for precisely detecting marine snow or selectively removing it without damaging the background.

Segmentation-based methods play an important role in image restoration tasks that require explicit specification of the target regions to be removed, because they can separate specific objects or regions in an image at the pixel level. In particular, deep learning-based segmentation models can learn features such as color, brightness, boundaries, and shape from input images and predict object regions in the form of masks. The generated masks can then be used as inputs for restoration techniques such as inpainting.

However, existing segmentation-based methods often focus on problems in which object shapes and boundaries are relatively clear or the target regions are larger than a certain size. In contrast, the marine snow addressed in this study appears as very small dots, blurred particles, or bright stains within images, and occupies a very low pixel ratio in the entire frame. Such small suspended particles do not have clear boundaries with the background, and their size and brightness vary from frame to frame.

 ISSN (Online): 3140-5312

Therefore, it is difficult to detect them stably using general segmentation methods alone.

Furthermore, underwater images contain background elements with visual characteristics similar to those of marine snow. Barnacles on the surfaces of underwater structures, lighting reflections, and bright textures in the seafloor background may have colors and shapes similar to suspended particles, causing false detections. Conversely, blurred or very small marine snow may be classified as background and remain undetected. Therefore, for marine snow detection, it is necessary to construct separate training data and design a segmentation model that considers the characteristics of small pixel-level suspended particles, rather than directly applying a general object segmentation model. Accordingly, in this study, high-resolution underwater images were used to train a U-Net-based segmentation model to detect marine snow regions as binary masks [4]. The generated masks were then used as inputs to the LaMa inpainting model, and experiments were conducted to restore the detected suspended particle regions using surrounding background information [5].

3. Research Method

3.1 U-Net Based Segmentation

3.1.1 Dataset Construction

In this study, a U-Net-based binary segmentation model was trained to detect marine snow regions in underwater images. To this end, videos captured by an underwater drone were divided into individual frames, and the marine snow regions present in each frame were manually annotated at the pixel level to construct the dataset. The input images consisted of actual underwater video frames, while the ground truth masks were generated as binary images in which marine snow regions were represented in white and background regions in black.

In the ground truth masks, the white regions indicate marine snow regions that the model should detect, whereas the black regions indicate background regions that should be preserved. By using these binary masks, the U-Net model was trained as a segmentation problem that separates marine snow regions from the input image at the pixel level.

When high-resolution underwater frames are used for training at their original size, GPU memory usage can increase rapidly, making training difficult. On the other hand, when the entire image is simply resized to reduce the input size, it was experimentally observed that marine snow pixels, which occupy a very small proportion of the total pixels, can be lost or weakened during the resizing process. Therefore, in this study, to preserve the detailed information of the original 4K images as much as possible while reducing the GPU memory burden, the entire images were not directly resized but were instead divided into 512 × 512 tiles to construct the training data.

Finally, the U-Net training dataset consisted of 21,711 image-mask pairs, and the validation dataset consisted of 5,258 image-mask pairs.

3.1.2 U-Net Training

The U-Net model was trained in a supervised learning manner using pairs of original images and their corresponding binary masks. The input images were loaded in RGB format and normalized to values between 0 and 1. For the masks, regions with pixel values greater than 127 were converted into marine snow regions, while the remaining regions were converted into background regions, thereby producing binary masks. During training, both the input images and masks were used at a size of 512 × 512.

The input image consists of a 512 × 512 RGB image and is converted into 64 feature channels through the first convolution block. Subsequently, in the encoder path, the spatial resolution of the feature maps is gradually reduced using max pooling, while the number of channels is increased to 64, 128, 256, 512, and 1024 to extract high-level features from the image.

Each encoder stage used a DoubleConv block composed of two 3 × 3 convolutions, batch normalization, and ReLU activation functions. Through this structure, the model was trained to learn local features of marine snow, such as brightness, shape, and boundaries. In the bottleneck stage, 1024 feature channels were used at the smallest spatial resolution to represent the abstract features of the input image.

In the decoder path, the resolution of the feature maps was gradually restored using transposed convolution. At this stage, feature maps generated at the corresponding resolution levels in the encoder were connected through skip connections and combined with the decoder feature maps. Skip connections help compensate for positional information and detailed boundary information that may be lost during the downsampling process. This is important for detecting marine snow regions that appear as small dots or blurred stains at the pixel level [4].

In the final output layer, a 1 × 1 convolution was applied to generate a one-channel output mask. The output value represents the probability that each pixel

 ISSN (Online): 3140-5312

corresponds to marine snow, and a threshold was subsequently applied to convert it into a binary mask that separates marine snow regions from background regions.

Since marine snow occupies a very small pixel ratio in the entire image, using only general binary cross entropy may cause the model to be biased toward the background region during training [6]. To alleviate this problem, this study used a combination of Weighted Binary Cross Entropy Loss, which assigns a weight to the positive class, and Dice Loss [7]. Weighted BCE assigns higher importance to marine snow pixels, thereby compensating for the learning of small suspended particle regions, while Dice Loss directly reflects the area overlap between the predicted mask and the ground truth mask [8].

The final loss function was constructed by combining Weighted BCE Loss and Dice Loss at an equal ratio.

$$L_{U-Net} = L_{WBCE} + L_{Dice}$$

Table 1 shows the main hyperparameters used for training the U-Net model in this study.

Table 1. Training Hyperparameter Settings

Parameter	Value
Batch size	4
Epoch number	100
Learning rate	0.0001
Early Stopping Patience	15

3.2 LaMa-Based Inpainting

In this study, the LaMa inpainting model was applied based on the marine snow masks generated through U-Net segmentation. Inpainting is a technique that restores regions to be removed from an image or video using surrounding background information. In this study, the marine snow regions marked in white in the binary masks predicted by U-Net were set as the target regions for removal, and these regions were restored using the LaMa model.

LaMa is a deep learning-based image restoration model proposed for large-mask inpainting and is used to restore large missing regions and complex structures. In this experiment, each frame and its corresponding U-Net-predicted mask were used as inputs, and the suspended particle regions corresponding to the mask regions were filled in a form similar to the surrounding background.

The processing was performed on a frame-by-frame basis. The original frame was input into the U-Net model to generate a marine snow mask, and the generated mask and the original frame were then input into the LaMa model. Subsequently, the frames restored by LaMa were reassembled into a video according to the frame order and FPS of the original video.

However, the LaMa inpainting used in this study is a frame-level image restoration method and therefore does not directly consider temporal consistency between consecutive frames. As a result, in videos captured by an underwater drone, where camera motion and suspended particle motion occur simultaneously, the restoration results may differ slightly from frame to frame. Considering this characteristic, this study visually compared the LaMa-based inpainting results and analyzed the background restoration quality after marine snow removal as well as the occurrence of residual suspended particles.

4. Experimental Results

In this study, marine snow regions were detected using the trained U-Net model. The experimental results showed that small suspended particle regions distributed in the image could be detected by dividing the original 4K video into tiles for training, without directly downscaling the video. Figure 1 shows the original underwater video frame, and Figure 2 shows the marine snow segmentation result generated by inputting the corresponding frame into the U-Net model. Some small suspended particles that were difficult to identify visually in the entire frame were also detected, which is considered to be the result of training while preserving the detailed information of the high-resolution image.

However, the detection results of U-Net showed that regions other than suspended particles were also detected. In particular, barnacles on the surfaces of underwater structures and bright reflection regions tended to be falsely detected when they exhibited features similar to those of suspended particles. This can be regarded as a limitation caused by the fact that a single-frame-based segmentation model cannot consider the motion information of suspended particles or temporal differences from the background.

Figure 3 shows the marine snow removal result obtained by applying an existing inpainting method using the U-Net-based segmentation result as a mask. For suspended particles with relatively clear sizes and distinguishable boundaries from the surrounding background, the masked regions were restored in a form similar to the surrounding background. However, for

ISSN (Online): 3140-5312

suspended particles appearing at the pixel level, the mask regions were very small, and the restoration effect was not clearly observed. In some regions, traces similar to the removed suspended particles were generated again.

In addition, when undetected residual suspended particles remained in the surrounding area, the inpainting model recognized them as background information and generated patterns similar to the residual suspended particles again during the restoration process. Therefore, it was confirmed that existing inpainting methods can be applied to general object removal, but have limitations in completely removing fine marine snow that is irregularly distributed in underwater images.

Overall, this experiment confirmed that the U-Net trained using a 4K tile-based learning approach can be applied to the detection of small suspended particles. However, false detections of background noise similar to suspended particles and the limitations of existing inpainting methods in restoring fine suspended particle regions were observed together. These results indicate that, for marine snow removal in underwater images, additional approaches that consider temporal continuity or the motion characteristics of suspended particles are required, beyond simple segmentation and restoration methods.

Fig. 1. Natural Image

Fig. 2. U-Net Output Mask

Fig. 3. LaMa inpainting Output Image

5. Conclusion

In this study, to examine the feasibility of removing marine snow from 4K underwater videos, the original images were divided into 512 × 512 tiles, and a U-Net-based segmentation model was trained. Through this approach, it was confirmed that small suspended particle regions can be detected at the pixel level without directly downscaling high-resolution images. In particular, the tile-based learning method showed the advantage of preserving the detailed information of the original images while allowing the model to focus relatively more on marine snow regions, which occupy a very small proportion of the entire image.

However, the experimental results showed that although U-Net-based segmentation could detect small marine snow regions relatively sensitively, it also detected regions with visual characteristics similar to marine snow, such as barnacles on the surfaces of underwater structures, bright reflection regions, and fine background textures. This indicates that, because marine snow appears in images as small dots, blurred particles, or bright stains, it is difficult to stably distinguish suspended particles from background noise based only on the spatial features of a single frame.

In addition, when existing inpainting methods were applied based on the detected masks, limitations were observed in naturally restoring small suspended particle regions at the pixel level. In particular, when undetected residual suspended particles existed around the mask, the restoration process referred to the surrounding information and filled the corresponding regions again in a form similar to suspended particles. This implies that, for marine snow removal, it is important not only to restore the detected regions but also to generate precise masks that consider fine suspended particles that remain undetected.

Limitations were also observed in video-level application. The same marine snow was not consistently detected across consecutive frames, and cases occurred in which it was detected in some frames but not in others. As a result, a flickering phenomenon was observed in the removal results, where certain suspended particles disappeared and then reappeared. This phenomenon can be considered a problem caused by the frame-level segmentation model not sufficiently reflecting temporal continuity.

Therefore, future research should consider detection methods that go beyond conventional approaches relying on the spatial features of a single frame and jointly reflect motion information and temporal consistency between consecutive frames. In addition, further research is needed to analyze the effect of the weight ratio between Weighted BCE Loss and Dice Loss during U-Net training on detection sensitivity and false detections, and to derive an optimal loss function combination that can more stably distinguish background regions from marine snow regions.

Acknowledgement

This research was supported by the Ministry of Science and ICT(MSIT), Korea, under the Sejong SW Convergence Cluster 2.0 Project supervised by the National IT Industry Promotion Agency(NIPA) (Grant No. PJTS08012604100700010003002000).

References

[1] Fernando Galetto, Guang Deng "A deep learning approach for marine snow synthesis and removal.", 2023
[2] Jeremy Paul Coffelt, Nicolas Nowald, Peter Kampmann, "Marine Snow Simulation and Elimination in Video", 2023
[3] Reina Kaneko, Yuya Sato, Takumi Ueda et al., "Marine Snow Removal Benchmarking Dataset", 2024
[4] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", 2015
[5] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin et al., "Resolution-robust Large Mask Inpainting with Fourier Convolutions", 2022
[6] Saeid Asgari Taghanaki, Yefeng Zheng, S. Kevin Zhou et al., "Combo loss: Handling input and output imbalance in multi-organ segmentation", 2019
[7] Jun Ma, Jianan Chen, Matthew Ng et al., "Loss odyssey in medical image segmentation", 2021
[8] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb et al., "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation", 2022

Author Information

Jung-kyu Park
2022~ now: BE, Dept. of Software & Communications Engineering, Hongik University
Email: pjg367@naver.com
Research Interests: Deep Learning, Computer Vision, Image Processing

Giseop Noh
1998: BE, Engineering, Korea Air Force Academy
2009: MS, Computer Science, University of Colorado, USA
2014: PhD, Computer Engineering, Seoul National University
2016.12~2018.02: Assistant Professor, Dept. of Computer & Information Science, Korea Air Force Academy
2018.03~2025.02: Assistant Professor, Dept. of Artificial Intelligence Software, Cheongju University
2025.03~now: Assistant Professor, Dept. of Software & Communications Engineering, Hongik University
Email: kafa46@hongik.ac.kr
Research Interests: NLP, Neural Network Theory, Reinforcement Learning