This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2925830, IEEE Sensors Journal

Crosstalk Removal in Forward Scan Sonar Image Using Deep Learning for Object Detection

Minsung Sung, *Student Member, IEEE*, Hyeonwoo Cho, *Member, IEEE*, Taesik Kim,
Hangil Joe, *Student Member, IEEE*, and Son-Cheol Yu, *Member, IEEE*

***Abstract*—This study proposes the detection and removal of crosstalk noise using a convolutional neural network in images of forward scan sonar. Because crosstalk noise occurs near an underwater object and distorts the shape of the object, underwater object detection is limited. The proposed method can detect crosstalk noise using the neural network and remove crosstalk noise based on the detection result. Thus, the proposed method can be applied to other sonar-image-based algorithms and enhance the reliability of those algorithms. We applied the proposed method to a three-dimensional point cloud generation and generated a more accurate point cloud. We verified the performance of the proposed method by performing multiple indoor and field experiments.**

***Index Terms*—crosstalk detection, sonar image crosstalk, underwater sonar crosstalk, underwater object detection.**

I. INTRODUCTION

UNDERWATER object detection is necessary for autonomous underwater vehicles (AUVs) to accomplish various underwater missions [1]-[6]. Furthermore, forward scan sonar (FSS) is one of the widely used sensors in underwater operations [7]-[10]. FSS exhibits a long operating range and high resolution compared with other sonar sensors and visibility in a turbid and dark environment. Therefore, many object detection algorithms using FSS have been developed.

FSS provides sonar images for its forward scene. Therefore, image processing algorithms that detect objects using FSS has been used in conventional approaches. Cho *et al.* [11] detected a target object in various angles of view using beam-based template matching between a simulated image and an actual sonar image. Kim *et al.* [12] detected an object by combining multiple Haar-like features with adaptive boosting. Bennett *et al.* [13] proposed a method for an AUV to track a target using adaptive feature mapping. However, in the sonar image, the shape of the object changes significantly depending on the sonar's view point. Thus, these algorithms exhibits limited accuracy and relatively high false positive rate, especially in field applications.

Thus, algorithms for detecting an object by reconstructing three-dimensional (3D) data from sonar images have been proposed. Yu *et al.* [14] recognized objects by emulating sonar images from a 3D model based on ray tracing. Lorenson *et al.* [15] formed the 3D model of an object based on voxels and recognized the object by two-dimensional (2D) coding from the constructed model. Cho *et al.* [16] developed a method to detect objects by generating the 3D point cloud of underwater objects from successive images acquired using an AUV's mobility. The 3D data such as the shape, range, and direction of object enable a more reliable object detection.

However, a characteristic noise, *crosstalk,* degrades the accuracy of sonar-based object detection algorithms. Crosstalk occurs inevitably owing to the imaging mechanism of the sonar sensor. It occurs near the underwater object and exhibits a similar intensity to the highlight of object. Thus, crosstalk noise distorts the shape of an object in the FSS. If crosstalk noise is removed, AUVs can recognize the objects and environments with higher reliability.

We herein propose a method to detect and remove crosstalk noise using a deep neural network (DNN). Recently, the DNN has been employed for object detection in sonar images [17]-[21]. Because crosstalk noise has its own highlight, the DNN can detect crosstalk noise in low-resolution and noisy sonar images through its deep architecture.

Collecting training images is a challenge in using the DNN for sonar images. The DNN requires an enormous amount of images capturing the target object. In the case of optical images in a terrain field, obtaining images involving various shapes of the target object is relatively easy because optical cameras have become popular and the development of the Internet has allowed for copyright-free images to be obtained with small cost. Meanwhile, underwater sonar images are typically not available to the public. Thus, obtaining sonar images requires manual experiments that require significant cost and time. Moreover, because the shape of the target object changes in the sonar images depending on the environments and the AUV's view point, acquiring sufficient numbers of training images is difficult.

Training the DNN to detect crosstalk noise is easier than training the DNN to detect a specific target object. Crosstalk noise has a feature that do not depend on the object type, the sonar's viewpoint, and the environment. Thus, the feature used to detect crosstalk noise in one sonar image is reusable in other sonar images captured in different environments. Furthermore, because crosstalk noise occurs frequently, collecting images containing crosstalk noise is easier than capturing the sonar

The authors are with the Department of Creative IT Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea (e-mail: ms.sung@postech.ac.kr; lighto@postech.ac.kr; weeds3450@posetch.ac.kr; roboticist@postech.ac.kr; sncyu@postech.ac.kr).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2925830, IEEE Sensors Journal

images of a specific object. In this study, we obtained 1,173 images containing crosstalk noise. Using the DNN trained with these images, we can detect and remove crosstalk noise in various environments successfully.

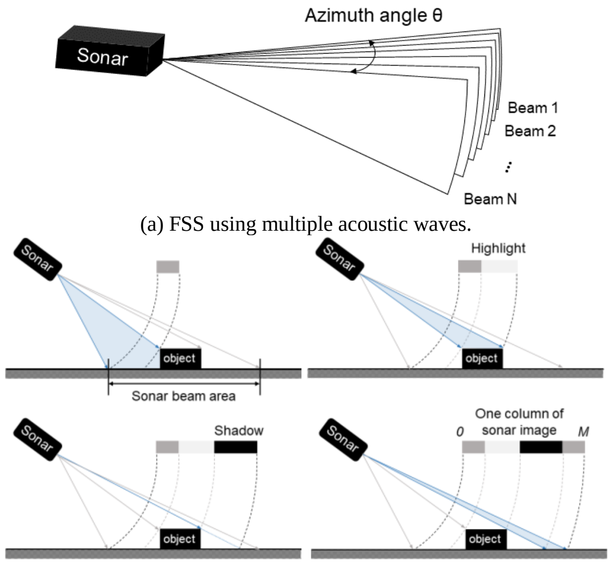
We apply the proposed method to a 3D-data-generation-based object detection algorithm [16]. This algorithm suffers from crosstalk noise as well. Moreover, the highlight of the seabed can be misperceived as the object. The proposed method removes crosstalk noise and allows for the true highlight of the underwater object to be extracted. Thus, the proposed method can prevent errors and generate a more accurate 3D point cloud. Likewise, the crosstalk-free images generated by the proposed method can be utilized in other sonar-image-based detection, localization, and navigation algorithms [22]-[26], and enhance the reliability of those algorithms.

This paper is organized as follows: In section II, we explain the geometry of the FSS and summarize the 3D object detection method proposed by Cho *et al.* [16]. In section III, we describe the causes and characteristics of crosstalk noise and difficulty in crosstalk removal using the conventional methods. Section IV describes the proposed method to detect and remove crosstalk noise. Section V presents the experimental results for verifying the reliability of the proposed method. The paper ends with the conclusion in section VI.

II. 3D-Data-Calculation-Based Object Detection

A sonar image loses the elevation angle of a captured scene because of the imaging mechanism used by sonar sensors [27]. The sonar sensor generates a sonar image by mapping the intensity of acoustic waves according to the distance and azimuth angle between the sonar sensor and reflected point. Thus, all the points with the same radius and azimuth angle around the sonar sensor are mapped as the same point in the sonar images. Therefore, restoring the elevation angle is an ill-posed problem.

Cho *et al.* [16] developed a method to reconstruct the elevation information of an object by analyzing sonar geometry with the AUV mobility. Furthermore, their method can generate a 3D point cloud of an underwater object by sequentially capturing the sonar images of the object, calculating the elevation information in every frame, and mapping the 3D information according to the AUV position. In other words, this method can extract 3D data from two-dimensional (2D) sonar images. Subsequently, they can recognize the underwater target object by comparing the calculated 3D data and the ground truth. We applied the proposed crosstalk detection and removal method to this object detection algorithm to verify the performance of the proposed method. In this section, we explain the geometry of the FSS and introduce this object detection method in more detail.



(a) FSS using multiple acoustic waves.

(b) Cross-sectional view of imaging mechanism in a single scan of the FSS.

Fig. 1. Imaging mechanism used by the FSS

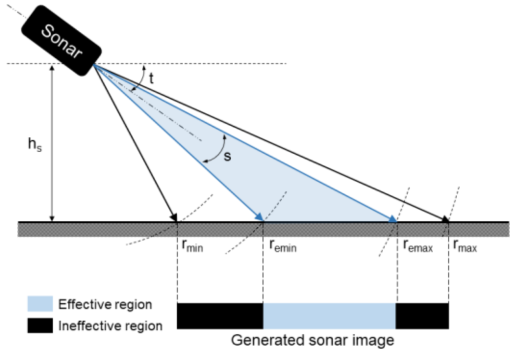


Fig. 2. Terms for the FSS image.

Fig. 1 describes the imaging principles used by the FSS to sense an underwater object. First, the FSS transmits multiple fan-shaped acoustic waves at various azimuth angles [28] as shown in Fig. 1a. Fig. 1b illustrates that one acoustic wave forms one column of the sonar image in a single scan. The acoustic wave is reflected from the seabed or surface of the underwater object and returns to the FSS. Subsequently, the FSS measures the time-of-flights (TOF) and the intensity of the reflected waves. Finally, the FSS generates a 2D sonar image by mapping the intensity of the reflected waves according to the range and angle [29]. The TOF is converted to the range from the FSS to reflected point by multiplying the speed of the acoustic waves, and the intensity is converted to a grayscale of the image through normalization. The multibeam FSS generates a sonar image of size *M* by *N* by transmitting *N* beams and mapping the signal measured by each beam into *M* pixels.

The method proposed by Cho *et al.* calculates 3D data from 2D sonar images by analyzing the geometrical relationship between the FSS and underwater object. They first defined some terminologies for the FSS image, as shown in Fig. 2. The FSS scans the range between r_{min} and $r_{max,}$ which are the user-defined window sizes. Furthermore, the acoustic waves have a finite vertical beam spreading angle *s*. Consequently, when there is no object, the highlights of the seabed is mapped to a specific range between the r_{min} and $r_{max.}$ This range is determined by the sonar tilt and vertical beam spreading angle, and defined as r_{emin} and r_{emax} denoted by

$$r_{emin} = \frac{h_s}{\sin(t+\frac{1}{2}s)}, \qquad r_{emax} = \frac{h_s}{\sin(t-\frac{1}{2}s)}, \qquad (1)$$

where h_s is the altitude of the FSS, *t* is the tilt angle of the FSS, and *s* is the vertical beam spreading angle. Cho *et al.* defined the region between r_{emin} and r_{emax} in the sonar image as the effe-

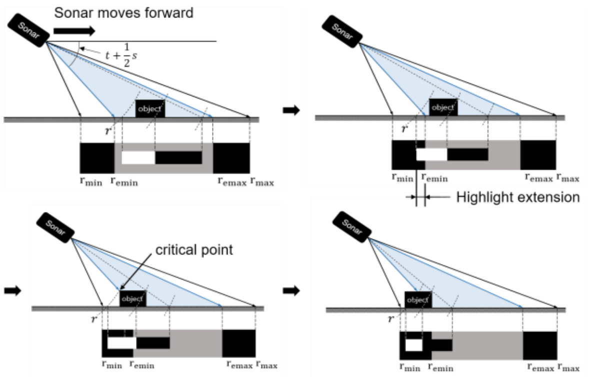


Fig. 3. Highlight extension according to movement of the FSS.

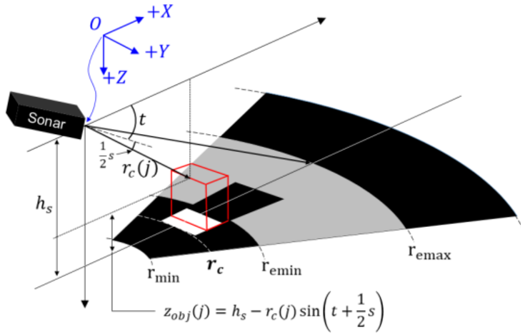


Fig. 4. Calculation of object height using highlight extension in the local coordinate system of the FSS.

ctive region, and the region outside this range as the ineffective region.

The geometry of the FSS in a particular situation is then analyzed to restore elevation of the underwater object. The FSS has a sweet spot, where the strength of the acoustic beams is concentrated and the signal-to-noise ratio is high, in the effective region. Therefore, when observing the object using the FSS, locating the object and its complete shadow in the sweet spot is preferable for an ideal identification [30], [31]. However, the authors focused on the situation where the underwater object exited the effective region. As shown in Fig. 3, the FSS scans the underwater object sequentially as it approaches the object. When the FSS is closer to the object than a certain distance, the reflection on the object occurred earlier in the slant range than the seabed, and the highlight is mapped in the ineffective area, named "*highlight extension*." As the FSS approaches the underwater object closer, the highlight extension increases until it reaches the critical point where the reflection occurs on the frontmost and uppermost of the object.

The critical point is determined by the sonar tilt, vertical beam spreading angle, and height of the object. Thus, the elevation information can be restored by measuring the length of the highlight extension at the critical point. Fig. 4 shows the condition at the critical point in three dimensions. We can calculate the position of the point on the object scanned by the j^{th} acoustic beam ($x_{obj}(j)$, $y_{obj}(j)$, $z_{obj}(j)$) as

$$x_{obj}(j) = x_s + r_c(j)\sqrt{1 - sin^2(\theta(j)) - sin^2\left(t+\frac{1}{2}s\right)}, \quad (2)$$

$$y_{obj}(j) = y_s + r_c(j)\sin(\theta(j)), \quad (3)$$

$$z_{obj}(j) = h_s - r_c(j)\sin\left(t+\frac{1}{2}s\right), \quad (4)$$

for $1 \leq j \leq N$, where j is the index of acoustic waves in the FSS, (x_s, y_s, h_s) is the position of the FSS, $r_c(j)$ is the distance between the FSS and the reflection point at the critical point, and $\theta(j)$ is the azimuth angle of the j^{th} acoustic beam.

Through (2)–(4), we can calculate the 3D data of the object only if $r_c(j)$ is obtained. The FSS maps [r_{min}, r_{max}] into [1, M] pixels. Thus, we can calculate $r_c(j)$ by extracting the pixel coordinates of the extended highlights in the sonar image using the following equation:

$$r_c(j) = r_{min} + (r_{max} - r_{min})\frac{p_c(j)}{M}, \quad (5)$$

for $1 \leq j \leq N$, where $p_c(j)$ is the row pixel index of the extended highlight in the j^{th} column of the sonar image.

Finally, the authors extracted extended highlight in the ineffective region by applying a difference filter D such as (6) and selecting pixels that exceed a threshold value.

$$D = \begin{bmatrix} -1 & -1 & \cdots & -1 \\ 0 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{3\times N} \quad (6)$$

III. Difficulty in Object Detection In The FSS

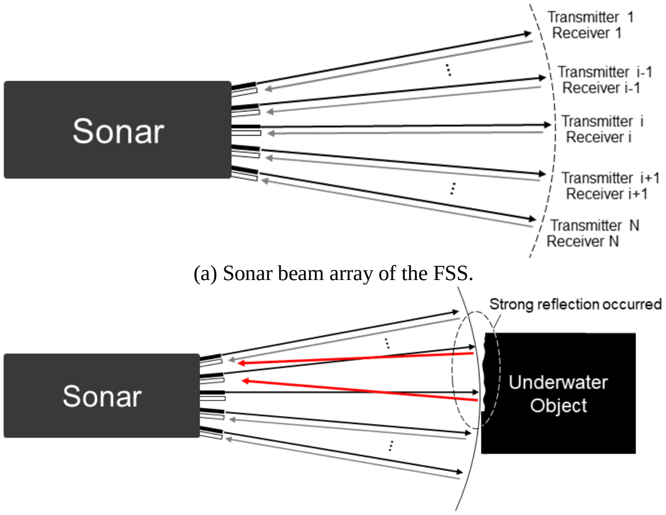
A. Problem Statement

Many object detection algorithms using FSS images can be degraded by two primary factors: crosstalk noise and ambiguous highlight. Crosstalk noise occurs near the object and has its own highlight. Thus, crosstalk noise distorts the highlight of one object and hides the shadow of neighboring objects, both of which are important sources for identifying the objects. The seabed has its own highlight as well; therefore, distinguishing the true highlight of the object is difficult. This is named the ambiguous highlight problem.

The method proposed by Cho *et al*. [16] was degraded by these two factors as well. They used a simple difference filter when extracting the highlight of the object to handle FSS images containing scanty information. The difference filter extracts the highlight from an image by detecting the sudden change in pixel value from dark to bright or from bright to dark. However, this approach may not well distinguish the highlight of crosstalk noise and the seabed from the highlight of the object. In this section, we discuss the crosstalk noise and ambiguous highlight in more detail and how these two factors cause errors in the algorithm.

B. Crosstalk Noise

The FSS is a multibeam sonar, and crosstalk is noise that typically occurs near an underwater object in a multibeam sonar image. Multibeam sonar uses a sonar array and transmits multiple acoustic waves to scan a region, not a line. To prevent interference among the waves, the multibeam sonar transmits each waves at a time interval. Thus, ideally, each receiver in the sonar array receives the reflected beam from the corresponding transmitter, as in Fig. 5a. However, the time interval required to obtain images at a high frame rates is a few milliseconds. Therefore, as shown in Fig. 5b, adjacent receivers may incorrectly receive the strong reflection that occurred on the surface of the underwater object and returned to the FSS in a sh-



(a) Sonar beam array of the FSS.

(b) Reflection occurring on the surface of the underwater object

Fig. 5. Cause of crosstalk noise in the FSS.

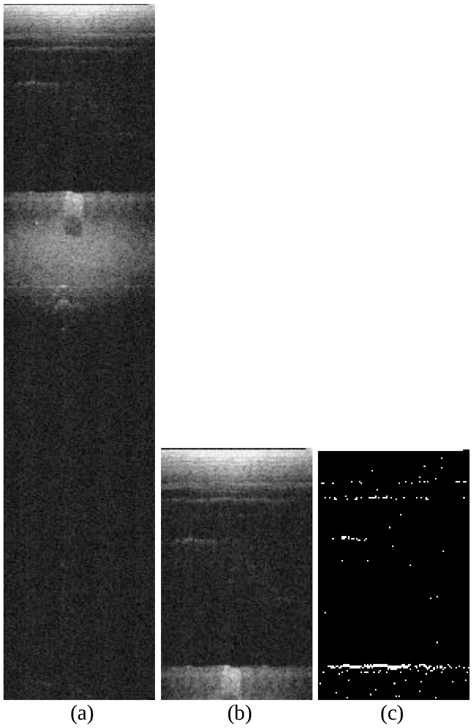


Fig. 6. Crosstalk noise that causes errors in 3D data calculation. (a) Crosstalk noise in the FSS image, (b) Crosstalk noise in the ineffective region, (c) Incorrectly perceived crosstalk noise as the object by the the difference filter.

ort TOF. This condition causes the highlight to spread around the object and is called crosstalk noise.

Distinguishing the true highlight of an object from crosstalk noise is difficult as crosstalk noise contains highlight as well. Therefore, the difference filter extracts a highlight wider than that of the object, as shown in Fig. 6. Crosstalk noise in the sonar image should be filtered to detect underwater objects more accurately.

Because the acoustic signal causes crosstalk noise, the conventional approach for crosstalk noise reduction in the sonar image is to use signal processing techniques [32], [33] through the following steps: First, the sonar image is divided into acoustic signal components using transformations such as Fourier transform or wavelet transform. Next, among the acoustic signal components, the acoustic signal that causes crosstalk noise is identified and filtered. Finally, a crosstalk-free image is generated by transforming the signals back to the spatial domain. Additionally, [34], [35] provide examples of crosstalk noise elimination using these approaches.

However, the methods, based on signal processing, exhibit limitations. Developing a general and automatic algorithm that identifies the crosstalk signal can be difficult owing to two reasons: First, the characteristic of the signals that construct the sonar image can vary for each image according to the environments where the image is captured. Next, because the conventional methods process the entire image, every pixel in the image is affected when applying signal processing. Consequently, these methods can cause undesired effects, such as information loss.

To remove crosstalk noise more accurately, we address several of its characteristics according to its cause of occurrence. First, a strong reflection of acoustic waves causes crosstalk noise; thus, crosstalk noise occurs primarily near the object. Next, the misperception of adjacent receivers causes crosstalk noise. Therefore, crosstalk noise exhibits a slightly lower intensity compared with the true highlight of the object. Finally, several adjacent receivers may incorrectly receive the reflected waves. As the adjacent receiver is farther from the corresponding receiver, the acoustic wave travels longer; thus, the intensity of the wave becomes weaker according to a parabolic curvature. Consequently, crosstalk noise exhibits a characteristic gradation pattern.

From these characteristics, several image processing-based methods to remove crosstalk noise have been used. Crosstalk noise occurs near an object when the periphery region is darker. Thus, in some studies [22], [36], crosstalk noise was removed by tracking the highlight and shadow of an underwater object from a frame where no crosstalk noise appears. However, this method presents two limitations. First, in sonar images, objects appear differently depending on the distance and viewpoint; therefore, identifying the exact highlight of the object is difficult in the current frame, even if the object is being tracked. Next, this method requires sequential frame information and causes additional computation.

Eliminating crosstalk noise using thresholding has been performed, because crosstalk noise exhibits a slightly weaker intensity compared with the highlight of the objects. However, the intensity value of crosstalk noise varies depending on the various factors such as the material of the underwater object, tilt angle of sonar sensors, and the captured scene. Therefore, the intensity of crosstalk noise in sonar images may be different in every experiment. Furthermore, the intensity of crosstalk noise can be different even in sequentially captured images. Consequently, setting a general threshold that can filter crosstalk noise is difficult, as shown in Fig. 7. Figs. 7a and 7b show the FSS images of the same object. Moreover, these images were captured at intervals of a few frames in the same scanning trial of the object. However, because the intensities of crosstalk noise were different in the two images, eliminating crosstalk noise using the same threshold value was not success-

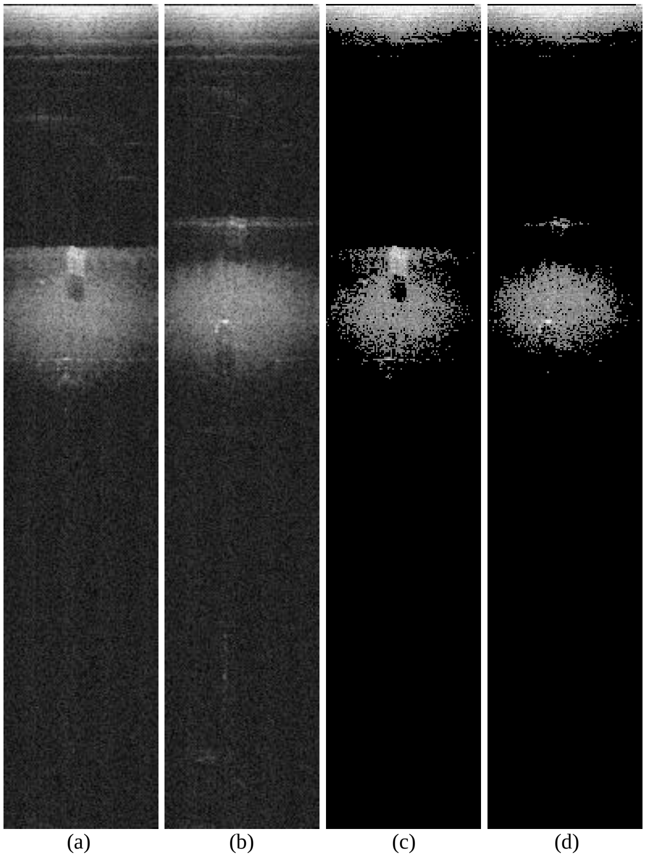


Fig. 7. Difficulty in crosstalk elimination using the threshold. (a) and (b) Original image of the same object, (c) and (d) Images obtained for the threshold value of 110.

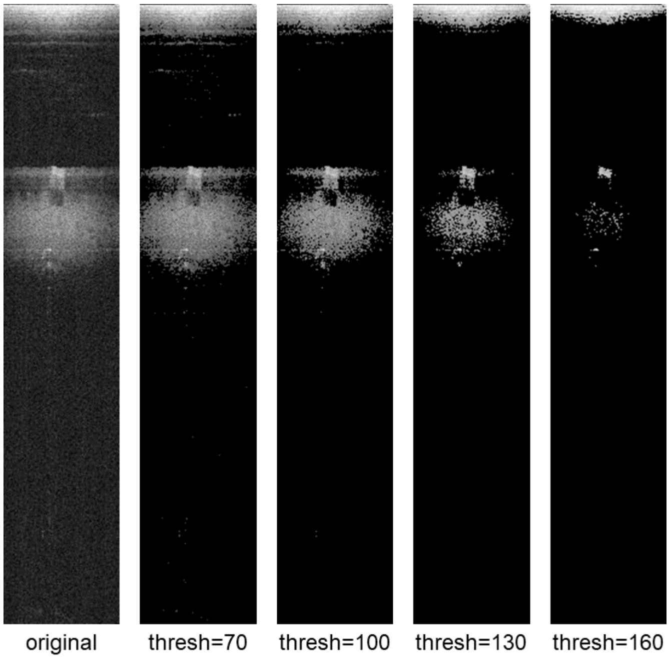


Fig. 8. Difficulty in crosstalk elimination using the threshold.

ful. When using 110 as the threshold value, we could remove crosstalk noise in one image, as shown in Fig. 7d. However, crosstalk noise still appeared and degraded the image in Fig. 7c.

Moreover, the intensity level of crosstalk was similar to that of the highlight from the seabed. The intensity of crosstalk noise immediately next to the underwater object is similar to that of the object. Thus, setting an appropriate intensity value is difficult. As shown in Fig. 8, a low threshold value cannot eliminate crosstalk noise properly. On the contrary, setting a high threshold value may remove valuable information such as the highlight of the seabed and that of the underwater object.

Therefore, the detection of crosstalk noise should precede for the accurate and efficient elimination of crosstalk noise. Detecting which part of the image is crosstalk noise is necessary to analyze the characteristics of crosstalk noise such as intensity values and patterns. Furthermore, we should detect the region where crosstalk occurs and process only the detected region to maintain other valuable information such as the highlight of the underwater object or seabed.

We present object detection techniques to detect crosstalk noise in the FSS images. Although crosstalk is a type of noise, it has its own highlight and shape. Furthermore, crosstalk noise exhibits the characteristic gradation pattern. Therefore, extracting features which allows to detect crosstalk noise is possible.

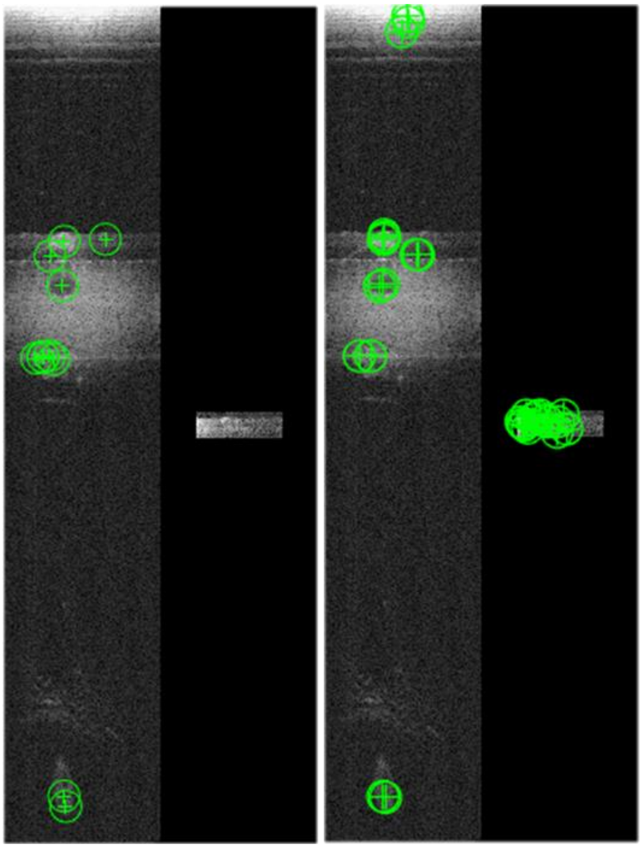
However, the conventional feature-based object detection algorithms did not perform well for detecting crosstalk noise in the given underwater sonar images. The sonar image is of low-resolution and has low signal-to-noise ratio. In other words, the sonar image contains scanty information. Therefore, it is difficult to extract low-level features and recognize highlights in the sonar images. Moreover, although the gradation pattern of crosstalk noise appeared similar, other characteristics such as intensity and size varied depending on the environment such as object type and setting of the FSS.

Fig. 9 shows the limitation of crosstalk noise detection using the conventional object detection algorithms. Fig. 9a shows the result of applying the speeded-up robust features (SURF) algorithm [37]. We thought that the SURF feature is suitable for the sonar image as it is robust to image blurring. However, we failed to extract the SURF feature from crosstalk noise, as the FSS image contains a faint highlight and the contrast is not large. Furthermore, we used the KAZE feature [38] in Fig. 9b. The KAZE feature extracts features in a nonlinear scale space; therefore, it can handle low-resolution and noisy sonar images. However, owing to scanty information, the KAZE features extracted in the whole image did not match with the features extracted in crosstalk.

C. Ambiguous Highlight

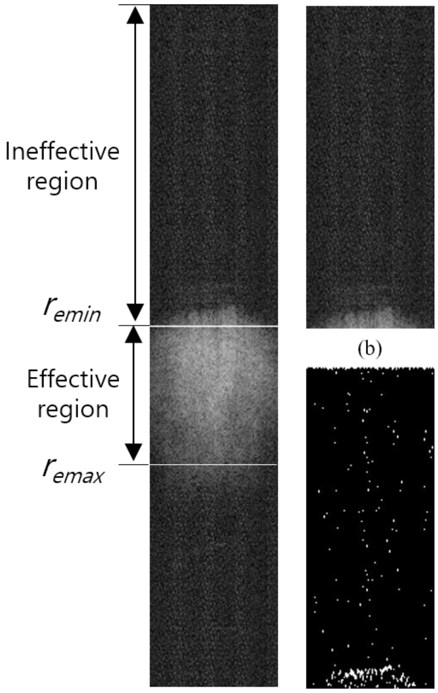
The method to restore the 3D information of the underwater object is based on the concept that underwater objects protrudes from the seabed. Therefore, the highlights in the ineffective region are regarded as the object. To find the ineffective region, r_{emin} is calculated by (1) using the altitude and tilt angle of the FSS.

However, the field conditions are not always ideal. The acoustic waves are scattered and the seabed is not flat. Therefore, the highlight of the seabed may also appear in the ineffective region. Moreover, the highlight of the seabed is similar to that of the underwater object and is difficult to be distinguished from the true highlight of the object. We call this condition the *ambiguous highlight*. Fig. 10 describes the ambi-



(a) (b)

Fig. 9. Difficulty in feature-based object detection. (a) Feature extraction and matching using SURF feature, (b) Feature extraction and matching using KAZE feature.



(a) (c)

Fig. 10. Ambiguous highlight problem. (a) Highlight of the seabed, (b) Highlight of the seabed that extends in the ineffective region, (c) Seabed mis-detected as the object by the difference filter.

guous highlight. The upper white horizontal line in Fig. 10a represents r_{emin}. Although no object exists on the seabed, the highlight of the seabed is extended in the ineffective region, such as in Fig. 10b, and the difference filter detects the seabed as the object, as shown in Fig. 10c. This incorrect information causes an error in the shape of the generated 3D point cloud. Thus, ambiguous highlights should be classified and the highlight of the seabed should be filtered to generate accurate 3D data.

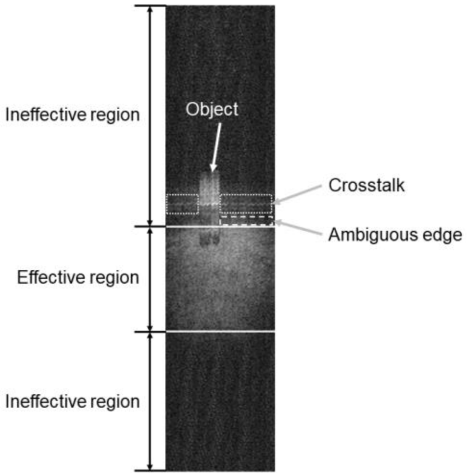


Fig. 11. Crosstalk noise and ambiguous highlight in the FSS image.

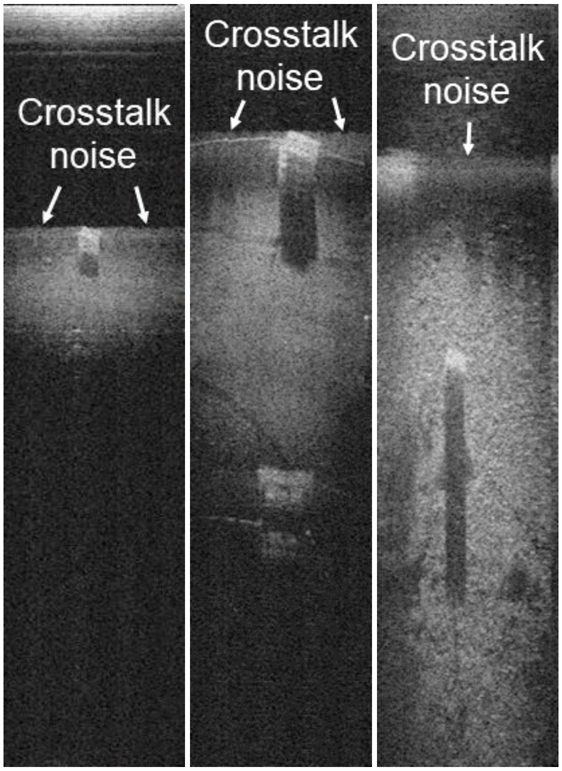
IV. Crosstalk Detection and Removal Using DNN

In this study, we propose a method to detect and remove crosstalk noise and ambiguous highlight that cause errors in detecting the highlight of an object in an ineffective region, as shown in Fig. 11. After removing the highlight of the crosstalk noise and seabed, we can identify the true highlight of the object. Subsequently, by applying the proposed method to the algorithm [16], we can calculate the accurate 3D data and recognize the underwater object. Moreover, the crosstalk-free images generated by the proposed method can be utilized in many sonar-image-based algorithms and enhance the reliability of those algorithms.

The detection of crosstalk noise is necessary for the accurate and efficient elimination of crosstalk noise. We introduced the DNN for the detection according to the characteristics of crosstalk noise. Crosstalk noise exhibits a parabolic gradation pattern. Further, this gradation pattern appears almost the same regardless of the environment such as in Fig. 12. Figs. 12a and 12b show the FSS images of the same brick. Although two images are captured in different sonar tilt angles and distances, a similar pattern is observed near the object. Crosstalk noise occurred similarly in the sea next to natural terrains such as rocks, as shown in Fig. 12c. Therefore, the DNN can detect crosstalk noise from a single given sonar image using this gradation pattern as a feature.

Next, the DNN is used to classify the ambiguous highlight into the highlight of the object, seabed, and crosstalk noise. The DNN exhibits an outstanding performance in object classification. Using its deep structure, the DNN can distinguish slight differences among ambiguous highlights and classify them into a seabed and object.

If the DNN detects the region where crosstalk noise occurred, we then apply crosstalk noise removal on the detected region.



(a) (b) (c)

Fig. 12. Crosstalk noise occurring in various environments. (a) and (b) Crosstalk noise around a brick, (c) Crosstalk noise around natural terrain of sea.

By applying the removal on the detected region, important information of the other areas are preserved. Furthermore, an accurate removal is possible by analyzing the characteristics of the detected crosstalk noise.

Finally, we applied the proposed method to the 3D reconstruction-based object detection algorithm. Because peripheral highlights such as crosstalk noise and seabed are removed, we can obtain more accurate 3D data to recognize the underwater object. In this section, we explain three parts: the DNN to detect crosstalk noise and classify ambiguous highlight; the method to remove crosstalk noise using the detection result; and the method to calculate accurate 3D data using the proposed method.

A. Convolutional Neural Network

Among various DNNs, we used the convolutional neural network (CNN) to detect crosstalk noise and classify ambiguous highlights in given sonar images. The CNN has demonstrated outstanding performance in object detection and classification. Unlike the conventional low-level feature-based object detection algorithms, the CNN produces high-level features by pooling the extracted feature through its deep architecture. Thus, the CNN can detect crosstalk noise even from low-resolution and noisy sonar images.

The algorithms used in the AUVs require fast processing speed and low computational complexity for three reasons. First, the AUVs have limited battery capacity; thus, they must limit their power consumption. Next, conducting underwater experiments with AUVs is time consuming and expensive. Finally, recording the absolute location of the AUVs is difficult because localization equipment such as GPS does not function underwater. Therefore, underwater experiments exhibit low reproducibility. Thus, we attempted to develop a real-time method to detect and remove crosstalk noise.

Among the CNNs developed for object detection, we adopted the “You Only Look Once (YOLO)” proposed by Redmon *et al.* [39]. Unlike the conventional two-stage CNN for object detection such as the fast R-CNN [40] or faster R-CNN [41], YOLO is a unified CNN. The single CNN selects the candidate region and classifies the selected region simultaneously. Thus, YOLO recorded a real-time processing speed of over 45 frames per second for the terrestrial images. Despite this fast processing speed, it recorded a high detection accuracy that is not less than that of the existing object-detection CNN.

The YOLO network comprises three versions. The latest version is YOLOv3 [42]. The YOLOv3 recorded the highest detection accuracy with state-of-the-art techniques such as batch normalization, anchor box, and multiscale prediction. Thus, we adopted the architecture of YOLOv3. However, because the underwater sonar images have lower resolution and contain less information compared with the terrestrial images, we modified some layers. Fig. 13 shows the architecture of the CNN we used. The CNN consists of 59 convolutional layers. For every convolutional layer, batch normalization [43] is applied and the activation function is leaky ReLU. Compared with the original architecture [42], we reduced the size of the input layers. Furthermore, we reduced the filter size of the last convolutional layer that makes final prediction of the class probabilities and bounding box into one-fourth.

Subsequently, we trained the CNN using a custom underwat-

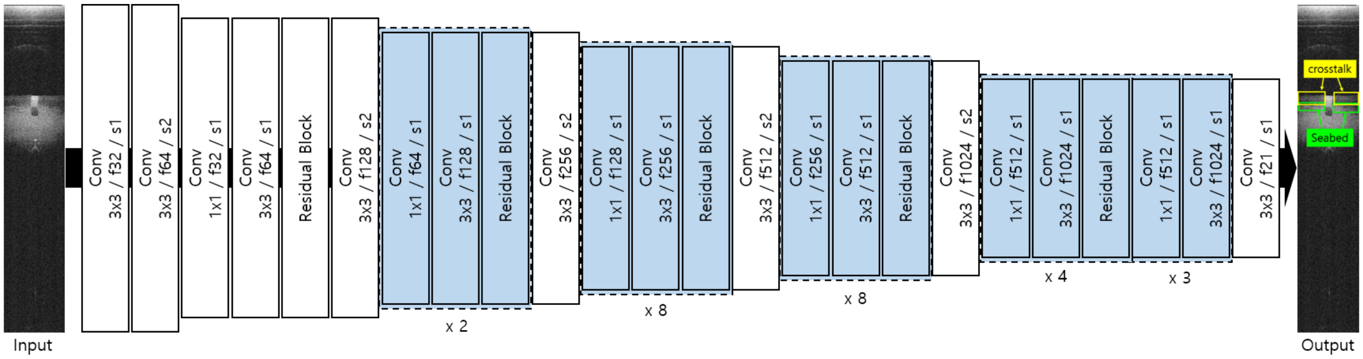


Fig. 13. Architecture of the CNN in the proposed method. Conv is convolutional layer. 1x1 or 3x3 is the size of the convolutional layer, f means the filter depth of the convolutional layers, and s means stride of the convolutional layers.

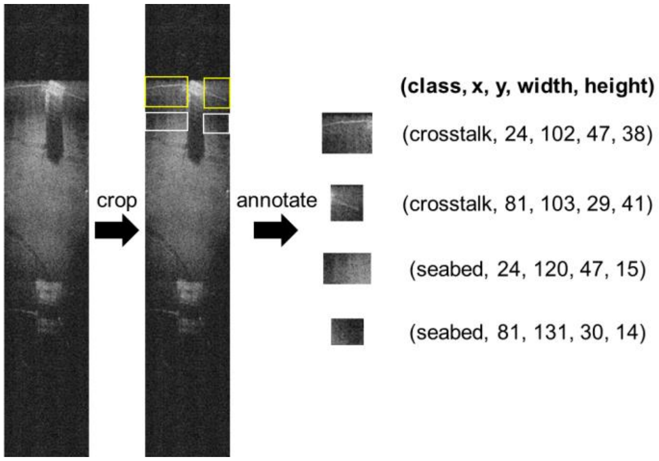

crop
annotate
(class, x, y, width, height)
(crosstalk, 24, 102, 47, 38)
(crosstalk, 81, 103, 29, 41)
(seabed, 24, 120, 47, 15)
(seabed, 81, 131, 30, 14)

Fig. 14. Training data generation process.

er sonar image dataset. We used supervised learning to train the network to detect the crosstalk noise and highlight of the seabed in the given sonar images. For supervised learning, we should provide the bounding box data of the target object as a label for the CNN. We manually cropped the crosstalk noise and seabed in the FSS image and recorded the x and y coordinates, width, and height to generate the training data.

Labeling the image such that the characteristics of the target objects are clearly visible is important to improve the detection accuracy of the CNN. Fig. 14 illustrates the labeling process. We addressed that crosstalk occurred on both sides of the underwater object. Therefore, when we cropped the bounding box of crosstalk noise, we created the bounding box to include the left or right boundary of the highlight of the object. Subsequently, the CNN would detect the crosstalk region including the underwater object after the training. We could solve this problem by verifying the intensity of the left and right boundaries and reducing the size of the detected bounding box. Next, because the acoustic waves spread out from the FSS in a fan shape, the edge of the seabed appeared as an arc in the FSS image. We cropped these arc-shaped edges and labeled them as the seabed.

Consequently, the trained CNN receives a single FSS frame as input. Subsequently, the CNN detects crosstalk noise and seabed in the image, and outputs the position and size of the region of the crosstalk noise and seabed.

B. Crosstalk Noise Removal

After the CNN detects the region where crosstalk noise occurs, we remove crosstalk noise by applying image-processing algorithms on the detected region. We can remove crosstalk noise by simply converting the detected crosstalk noise into a shadow. The CNN we built detects crosstalk noise in the ineffective area. If no object exists, highlights do not occur in the ineffective region. Moreover, crosstalk noise may not be visible if the seabed or another object exists behind the crosstalk noise. Because these objects cause strongly reflected waves, the receiver of the sonar sensor can receive the reflected wave from its corresponding transmitter, instead of misreceiving it from the adjacent transmitter. In other words, the region detected by the CNN where crosstalk noise occurs is originally a shadow. Thus, we can remove crosstalk noise by simply converting the pixels in the detected region into a shadow.

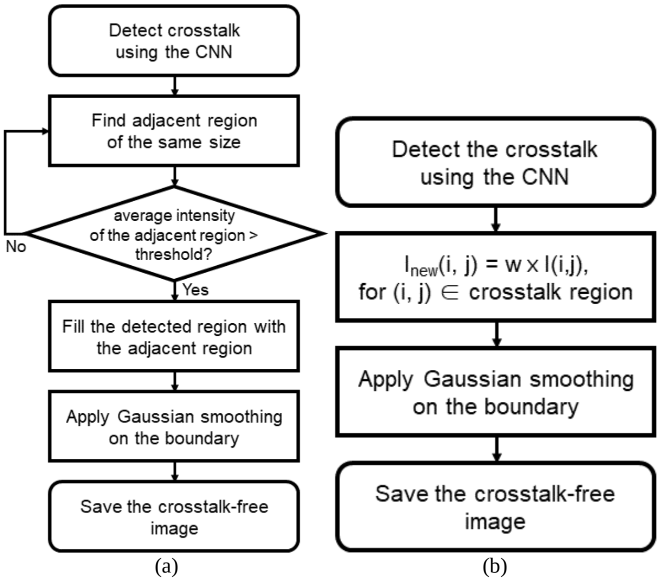

Detect crosstalk using the CNN
Find adjacent region of the same size
average intensity of the adjacent region > threshold?
No
Yes
Fill the detected region with the adjacent region
Apply Gaussian smoothing on the boundary
Save the crosstalk-free image
Detect the crosstalk using the CNN
$I_{new}(i, j) = w \times I(i,j)$, for $(i, j) \in$ crosstalk region
Apply Gaussian smoothing on the boundary
Save the crosstalk-free image
(a)
(b)

Fig. 15. Flowchart of crosstalk removal. (a) Crosstalk removal based on inpainting, (b) Crosstalk removal based on intensity adjustment.

We propose two methods that transform the detected crosstalk noise region into a shadow. The first method is inpainting. This method eliminates crosstalk noise by filling the detected region with adjacent pixel values that are also the shadow. Fig. 15a illustrates this method. When the CNN detects the crosstalk, the algorithm searches for the adjacent region of the same size with the detected bounding box. Subsequently, the algorithm verifies if the searched area is a shadow by comparing the average pixel value in the area and a threshold value. Threshold value is determined dynamically according to the pixel intensity of the detected crosstalk noise. The algorithm subsequently copies the pixel values of the selected area and paints the pixel values on the crosstalk noise region. Because repainting may cause the images to appear unnatural, we applied Gaussian smoothing on the boundary of the repainted region as the final step. The second method is intensity adjustment. This method mitigates crosstalk noise by multiplying a small weight value w in each pixels in the detected crosstalk noise region. The value of w is also determined by analyzing the pixel intensity of the detected crosstalk noise. Fig. 15b shows the flowchart of this method. After the CNN detects the crosstalk region, the algorithm multiplies w for each pixel value and constructs new images. The result of this method may appear unnatural as well; thus, Gaussian smoothing was applied on the boundary.

We can generate crosstalk-free images using these two algorithms. The generated crosstalk-free images can be utilized in many sonar-image-based algorithms, such as localization and navigation, because they can provide more accurate information for underwater landmarks or objects.

C. 3D Reconstruction for Underwater Object

As one application of the proposed method, we propose a precise 3D-data-based object detection using crosstalk removal method and the algorithm [16]. We apply the proposed method

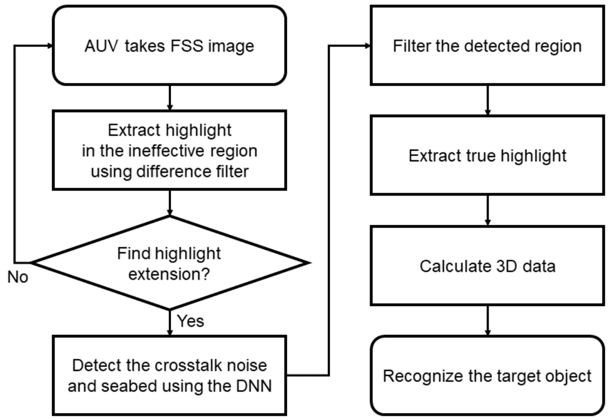

AUV takes FSS image
Extract highlight in the ineffective region using difference filter
Find highlight extension?
No
Yes
Detect the crosstalk noise and seabed using the DNN
Filter the detected region
Extract true highlight
Calculate 3D data
Recognize the target object

Fig. 16. Flowchart of the proposed method for detecting underwater objects.

to the 3D point cloud generation algorithm, such as in Fig. 16, for object detection. First, the proposed method divides the sonar image into an effective and ineffective region based on the tilt angle and altitude of the FSS. Subsequently, the proposed method extracts the extended highlights in the ineffective area using the difference filter. Next, the CNN detects the bounding box data of the crosstalk noise and seabed. If the extracted highlights by the difference filter are included in the region detected by the DNN, those pixels are filtered; subsequently, the true highlight of the underwater object can be identified. Next, the proposed method measures the length of the highlight extension from a distance between the extracted true highlights and r_{emin}. Subsequently, the 3D coordinate values of the object are calculated from the length of the highlight extension through (2)–(4). Finally, we reconstruct the object in three dimensions by calculating the coordinate values from sequential scanning images of the object and mapping the calculated value in the global coordinate system. We can recognize the underwater object by comparing the reconstructed 3D data with the ground truths.

This method is different from conventional approaches to detect an object using the CNN. The conventional object detection approaches use a CNN that is trained to detect the specific target object. However, because of the difficulty in predicting the shape of an object in a sonar image and obtaining training images of the target object, we propose training the CNN to detect crosstalk noise. Because crosstalk noise exhibits similar characteristics regardless of the underwater object, obtaining the training data is relatively easy, and the CNN can detect crosstalk noise with high accuracy. Furthermore, crosstalk noise occurs near an object; therefore, we can detect the highlight of the object next to the crosstalk noise. Finally, we can recognize the object by generating 3D data using the highlight of the object.

V. Experiment

A. Experimental Setup

We conducted indoor water tank experiments to obtain actual underwater FSS images to train the CNN and verify the proposed method. In the indoor water tank, the seabed indicates the floor of the water tank.

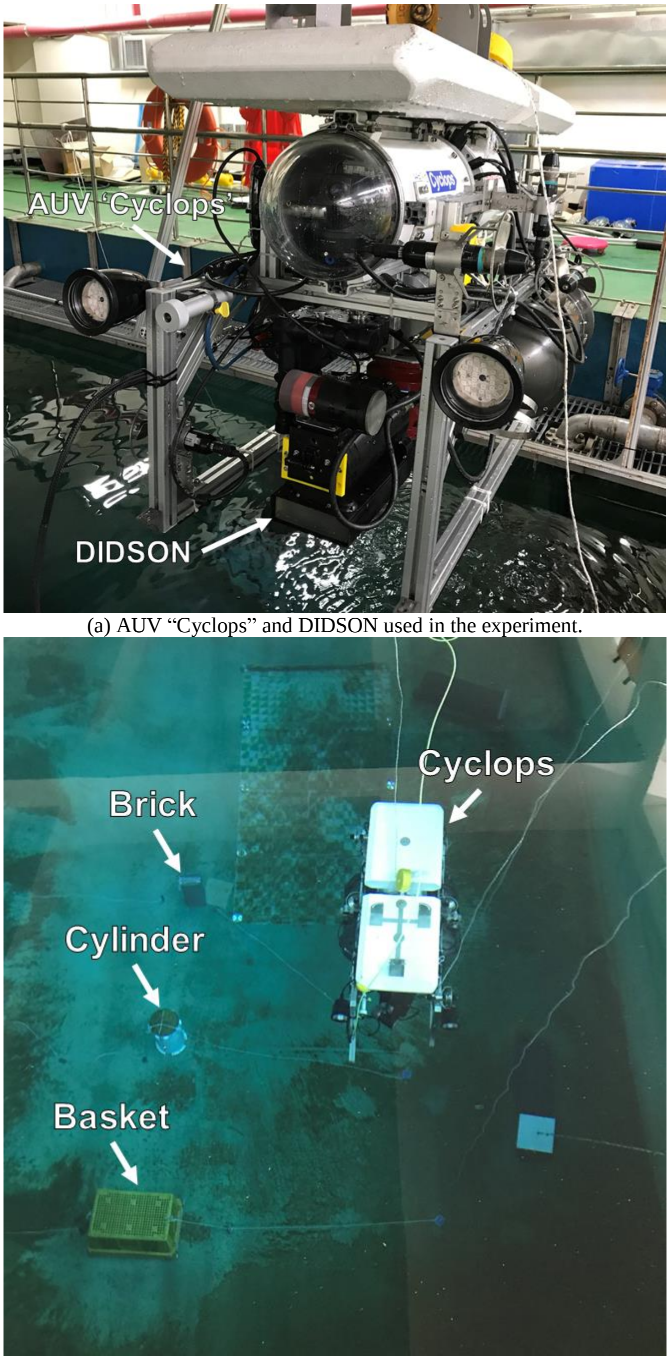
TABLE I.
Specifications of the DIDSON

Parameter	Value
Operating frequency	1.8 MHz
Vertical beam spreading angle	14 °
Azimuth field of view	29 °
Number of beams	96
Maximum resolution	0.3 °
Maximum imaging range	12 m
Image size	512 × 96
Frame rate	4–21 fps
Depth rating	300 m

TABLE II.
Specifications of the "Cyclops"

Parameter	Value
Dimension	0.9 m × 1.5 m × 0.9 m (width × length × height)
Weight	210 kg in air
Depth rating	100 m
Propulsion	8 thrusters (475 W)
Maximum speed	2 knots
Power source & batteries	24 VDC & 600 Wh Li-Po battery × 2
Computing system	PC-104 (Intel Atom @ 1.66 GHz) × 2
Sensors	1.1 MHz & 1.8 MHz Forward Scan Sonar Digital pressure transducer Doppler velocity logger Fiber-optic gyro

The data used to train the CNN affects the detection accuracy significantly. Thus, we designed the experiments with four points to obtain various FSS images. First, we installed various types of objects on the floor of the water tank: aluminum cylinder, brick, and plastic basket. Next, we attached the FSS on the AUV "Cyclops" [6] and obtained the sonar images by moving the AUV in lawnmower trajectory. Hence, we can vary the angle and distance between the FSS and the underwater object. Moreover, the ambiguous highlight is the most difficult to distinguish when the highlights of the crosstalk noise, seabed, and underwater object are overlapped. Therefore, we attempted to obtain many images in which those highlights are overlapped. If the AUV moved in the lawn mower trajectory, this condition occurred frequently because the sonar sensor moves back and forth with respect to the installed underwater object. For the FSS, we used a dual-frequency identification sonar (DIDSON) developed by Edward *et al.* [44]. Tables I and II list the specifications of the DIDSON and AUV "Cyclops," respectively. Next, we captured images by changing the tilt angle and altitude of the FSS. Consequently, we can cause crosstalk noise and the floor of the tank to appear in various sizes and at various locations. Finally, we conducted experiments in two indoor water tanks to increase the number of data and diversify the capturing environments. The dimensions of two indoor tanks is 8 m × 12 m × 6 m and 10 m × 85 m × 3.5 m (width × length × depth), and both tanks were filled with clear water. Thus, we can construct a dataset containing diverse highlights of the crosstalk noise and floor



(a) AUV "Cyclops" and DIDSON used in the experiment.

(b) "Cyclops" and the objects installed on the floor

Fig. 17. Experiment in the indoor water tank.

with different intensities, different sizes, and different positions in the images. Fig. 17 illustrates the experimental setup to obtain the dataset and Fig. 18 shows examples of images obtained in different condition.

We conducted 11 experiments scanning the underwater objects and took 16,792 frames of FSS images. Among these 16,792 frames, we acquired 4,254 frames that the crosstalk noise occurred near the underwater object. To filter out sequentially taken similar images, we randomly sampled 1,173 images and used them as training data. We then selected 384 images not included in the training data and used them as test data. Table III specifies the constructed training dataset and experimental settings for capturing those training images.

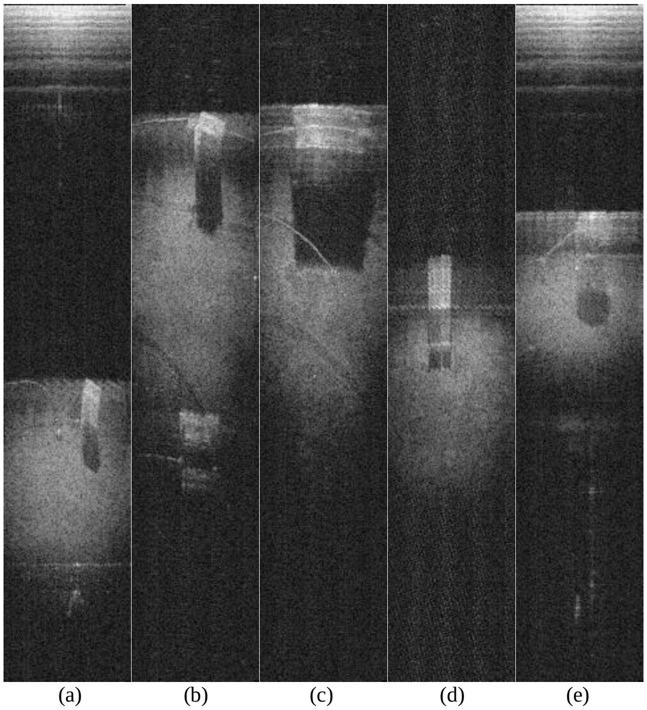


Fig. 18. Crosstalk noise in various environments. (a) Crosstalk noise around the brick at a distance, (b) Crosstalk noise around the brick nearby, (c) Crosstalk noise around the horizontally placed basket, (d) Crosstalk noise around the vertically placed basket, (e) Crosstalk noise around the cylinder.

TABLE III.
DESCRIPTION OF THE TRAINING DATASET

Environment (w × l × d)	Installed Objects	Sonar Setting (tilt / altitude / $r_{min} \sim r_{max}$)	# of Images
Water Tank 1 (8 m × 12 m × 6 m)	Brick, Cylinder, Basket	45.2 ° / 2.72 m 0.42–10.42 m	300
	Brick, Cylinder, Basket	45.2 ° / 2.69 m 0.42– 5.42 m	150
	Brick, Cylinder, Basket	45.2 ° / 1.68 m 0.42–5.42 m	360
	Brick, Basket	30.0 ° / 2.15 m 1.25–6.25 m	300
Water Tank 2 (10 m × 85 m × 3.5 m)	Basket	45.0 ° / 3.18 m / 1.67–6.67 m	63

B. *Experimental Result*

We trained the CNN for 22,700 epochs using 1,173 training images. To reduce the training time, we used the transfer learning and the CNN was trained from a pre-trained model on ImageNet [45]. Consequently, the training lasted 11 hours using the graphics processing unit (GPU) Titan X. Fig. 19 shows the CNN loss with the training epochs. The CNN loss was calculated as the summation of the sum-squared error between the predicted bounding box and ground-truth bounding boxes and the sum-squared error between the predicted class probability and ground-truth class. We measured the loss value for every 100 training epochs using a batch of training dataset. We stopped the training if the loss value did not decrease significantly for 1,000 training epochs. The final loss value was 0.189. Fig. 20 shows the outputs of the CNN according to the training epochs. The CNN misclassifies the ambiguous highlights in the early stage. As the training progressed, the CNN could classify the ambiguous highlights

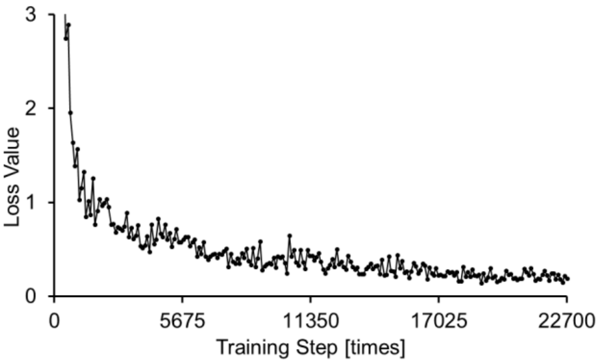


Fig. 19. Graph of loss value over training steps.

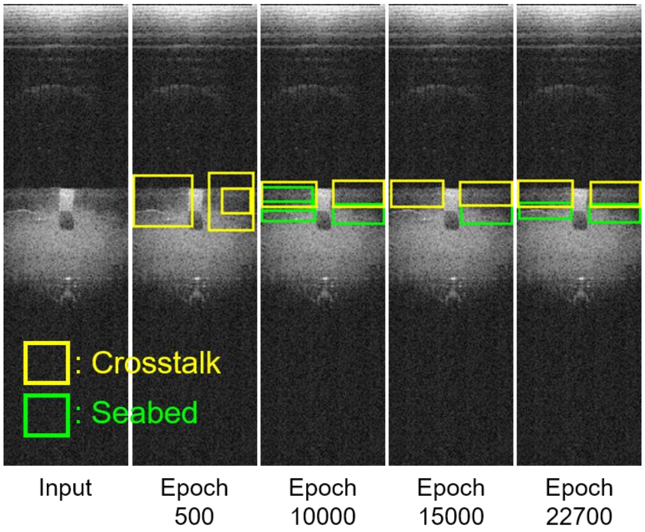


Fig. 20. Outputs of the CNN according to the training epochs.

into crosstalk noise, the floor of the water tank, and objects accurately, and extracted the bounding box precisely.

After completing the CNN training, we can detect the crosstalk noise and seabed using the trained CNN. In the images captured at the indoor water tank, the seabed indicates the floor of the water tank. Fig. 21 shows the detection result. The images used for the input were not included in the training dataset. As shown in Fig. 21a, the CNN can detect the floor of the water tank on the boundary between the effective and ineffective region. Fig. 21b shows the CNN detecting the crosstalk and floor simultaneously. As shown in Fig. 21c, the CNN can detect the crosstalk noise although the crosstalk overlapped with the highlight of the floor. Although both the crosstalk noise and floor have similar faint highlight, the CNN can classify them accurately.

We measured the detection accuracy and processing speed to verify the performance of the proposed CNN-based crosstalk detection method quantitatively. The receiver operating characteristic (ROC) curve in Fig. 22 and the measured detection accuracy in Table IV show that the trained CNN can distinguish the crosstalk noise and the seabed accurately from other types of highlights, such as underwater objects, in given FSS images. The error primarily occurred at the instant when the crosstalk noise appeared in the ineffective region, as shown in Fig. 21c. In this situation, the crosstalk noise overlapped with the floor of the water tank; therefore, distinguishing the crosstalk noise and floor is difficult.

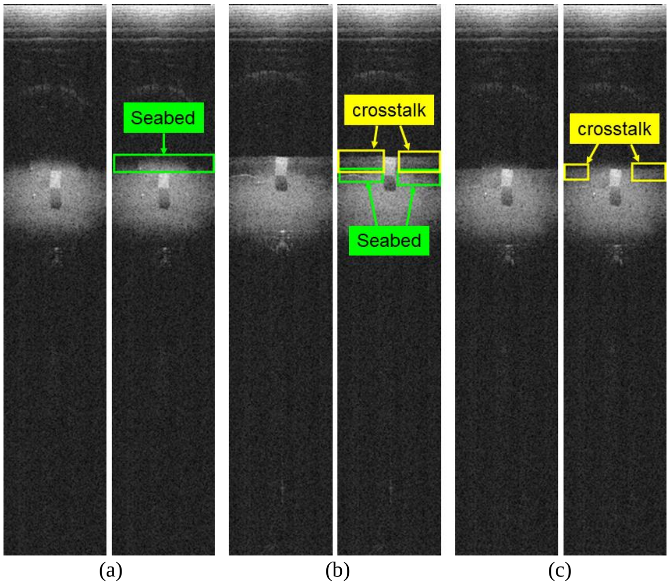


Fig. 21. Crosstalk and seabed detection result. (a) Detection of seabed, (b) Simultaneous detection of crosstalk and seabed, (c) Detection of crosstalk when it overlapped with seabed.

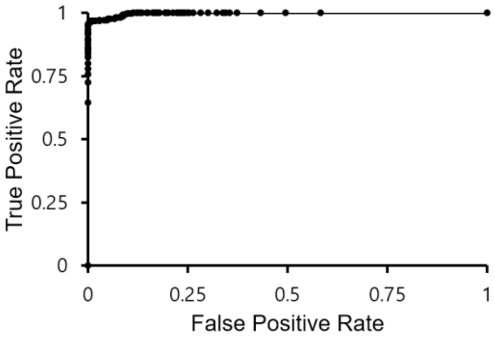


Fig. 22. ROC curve of the CNN.

TABLE IV.
DETECTION RESULT OF THE CNN

Detection Accuracy		Processing speed
Sensitivity	*Specificity*	
97.0 %	97.1 %	49.28 fps

Furthermore, the CNN can process 49.28 images per second when using the GPU Titan X. This processing speed was faster than the frame rates of the DIDSON. Therefore, if the proposed method is applied to the AUV, we can detect the crosstalk noise in real time.

Subsequently, we removed the crosstalk noise in the FSS image using the bounding box data of the crosstalk noise that the CNN detected. Because the crosstalk noise occurs near the underwater objects, it becomes difficult to recognize the exact highlight of the underwater objects or landmarks utilized in the sonar image-based algorithms. The crosstalk-free image generated by the proposed method can enhance the reliability of algorithms utilizing sonar images. We removed the crosstalk noise in the given sonar images by applying two image-processing algorithms, inpainting and intensity adjustment, on the region detected as crosstalk by the CNN.

Fig. 23 shows the result of the crosstalk noise removal. Figs. 23a and 23d are the input images for the crosstalk removal algorithms. Crosstalk noise occurred on both sides of the high-

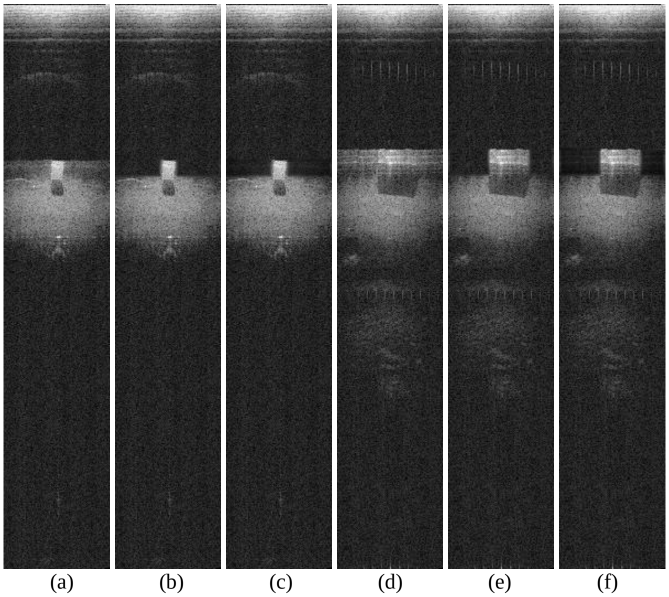


Fig. 23. Crosstalk removal result. (a) and (d) Input image, (b) and (e) Crosstalk removal result using inpainting, (c) and (f) Crosstalk removal result using intensity adjustment.

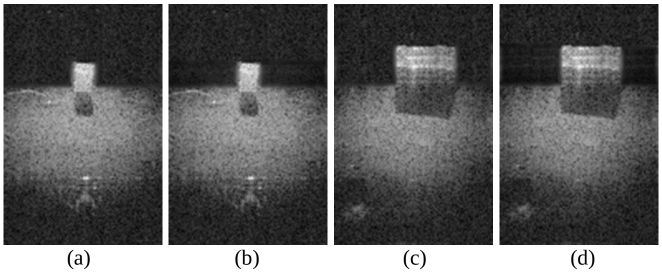


Fig. 24. Magnification of crosstalk removal result. (a) Magnification of Fig. 23b, (b) Magnification of Fig. 23c, (c) Magnification of Fig. 23e, (d) Magnification of Fig. 23f.

light of the underwater object. Figs. 23b and 23e are the results of removing the crosstalk noise using inpainting. Figs. 23c and 23f are the results of removing crosstalk noise using intensity adjustment. Both methods remove crosstalk noise well in the given image. Fig. 24 is a magnification of Fig. 23 around the object. It illustrates that the proposed method removes crosstalk noise well and the resulting images appear natural.

The proposed method can process sonar images captured at various environments. Fig. 25 shows more results for the crosstalk noise detection and removal. Fig. 25a is a brick captured at a short distance, which is different from the sonar settings of Fig. 23a. Furthermore, the CNN detects the crosstalk noise and the floor of the water tank, as shown in Fig. 25b, and the detected crosstalk noise is removed accurately as shown in Fig. 25c. Furthermore, the proposed method detects and removes crosstalk noises occurred near another types of object, aluminum cylinder, such as in Figs. 25d–f.

We applied the proposed crosstalk detection and the removal algorithm to the 3D point cloud generation algorithm. First, we extract the true highlight of the underwater object following Fig. 26. Given the input image and the position of the AUV, we first calculate the r_{emin} and crop the ineffective area. Subsequently, the difference filter extracts the highlight extension in the ineffective region. However, these highlights includes the crosstalk noise and seabed. Therefore, the CNN detects the crosstalk noise and the seabed simultaneously. Finally, we identify the true highlight of the underwater object by excluding pixels included in the detected bounding boxes.

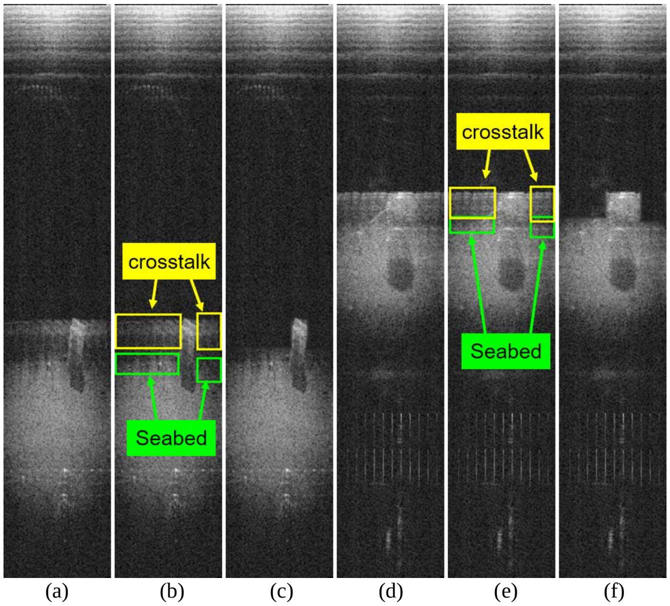


Fig. 25. Crosstalk noise detection and removal result. (a) and (d) Brick and cylinder in different conditions, (b) and (e) Crosstalk and seabed detection results, (c) and (f) Crosstalk removal results.

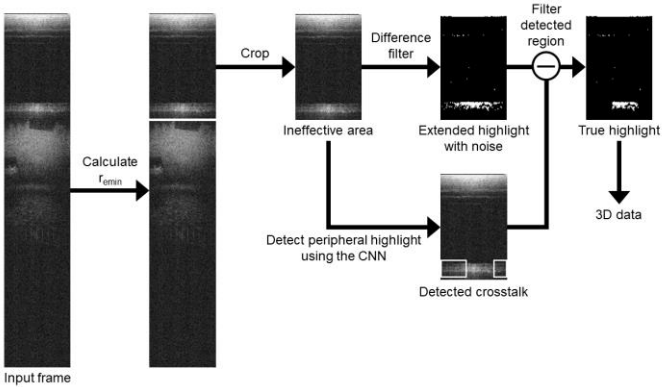
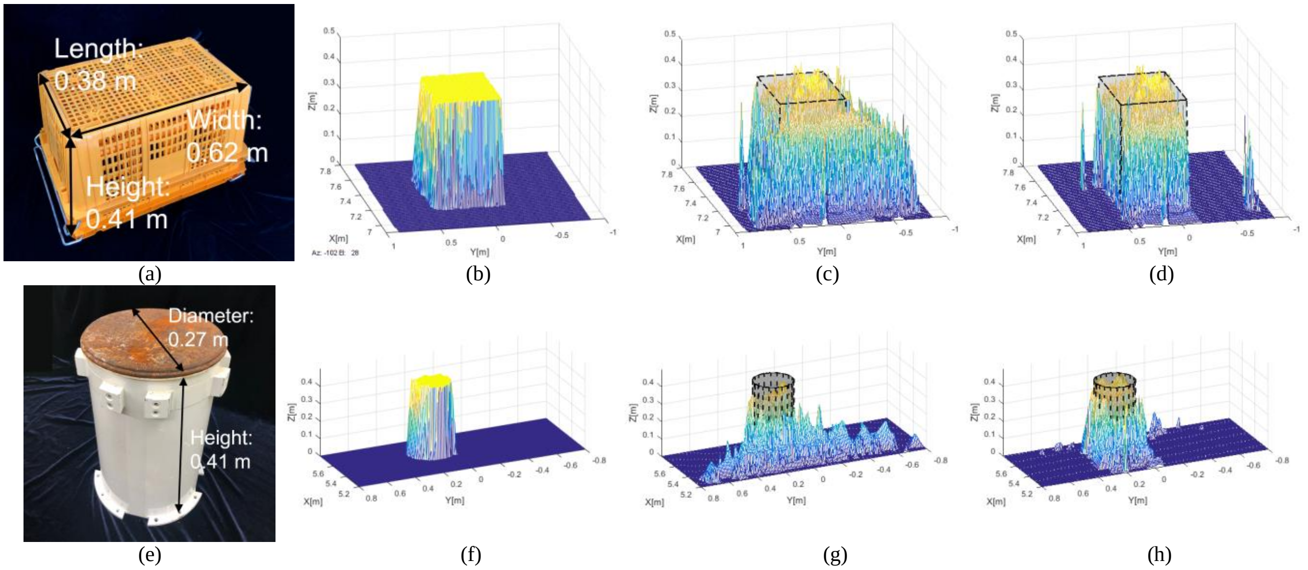


Fig. 26. Processing pipeline of the proposed method.

Consequently, we can generate more accurate 3D point cloud of the underwater objects using the proposed method, as shown in Fig. 27. The AUV scans the object and captures the sequential FSS images. Subsequently, we calculate the 3D coordinate values of the extracted highlight for every frame in the sequential images and map the calculated 3D values according to the AUV position. Sonar images scanning the basket shown in Fig. 27a were the inputs of the proposed method. Fig. 27b shows the ground-truth point cloud and it appears as a gray box in Figs. 27c and 27d. Fig. 27c shows the point cloud generated using the sonar images without the proposed method. Owing to the other highlights of the crosstalk noise and the seabed near the object, the generated point cloud was different with the ground truth. Meanwhile, the proposed method can eliminate the peripheral highlights such as the crosstalk noise and seabed. As shown in Fig. 27d, we can generate a more accurate 3D point cloud of the underwater object. Furthermore, we generated the 3D point cloud of the aluminum cylinder of Fig. 27e. Fig. 27f is the ground-truth point cloud of the cylinder, and it is illustrated as gray cylinder



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 27. Comparison of generated 3D point cloud. (a) and (e) Object used in the experiment, (b) and (f) Ground truth point cloud, (c) and (g) Point cloud generated without the proposed method, (d) and (h) More accurate point cloud by applying the proposed method.

TABLE V.
3D DATA CALCULATION RESULT OF THE BASKET

	Ground truth			Without the proposed method			The proposed method		
	w	*l*	*h*	*w*	*l*	*h*	*w*	*l*	*h*
Value [m]	0.62	0.38	0.41	0.98	0.41	0.42	**0.60**	0.41	0.43
Error rate [%]	-	-	-	58.1	7.9	2.4	**3.2**	7.9	4.9

TABLE VI.
3D DATA CALCULATION RESULT OF THE CYLINDER

	Ground truth		Without the proposed method		The proposed method	
	D	*h*	*D*	*h*	*D*	*h*
Value [m]	0.27	0.41	0.53	0.39	**0.53**	0.39
Error rate [%]	-	-	96.3	4.9	**7.4**	4.9

in Fig. 27g and 27h. We can generate the accurate cylindrical point cloud as shown in Fig. 27h by removing the crosstalk noise with the proposed method.

To evaluate two generated point clouds quantitatively, we measured the typical 3D data: width (*w*), length (*l*), and height (*h*) of the basket, and diameter (*D*) and height (*h*) of the cylinder. Tables V and VI show the results. The proposed method allowed to reconstruct the 3D data of the underwater object more accurately. Particularly, by filtering the crosstalk noise near the object, the proposed method can decrease the error significantly with respect to the horizontal data of the object such as the width and diameter. Because we can measure the accurate 3D data with the proposed method, we can recognize the underwater target object by comparing with the ground truth.

C. Field experiment

Furthermore, we applied the proposed method to the sonar images captured at the sea to verify the robustness of the proposed method. Similar to the indoor water tank experiment, we attached DIDSON to the AUV and captured sonar images with the AUV moving in a lawn mower trajectory. We scanned a site containing a rocky seabed and acquired sonar images.

The proposed method can process sonar images captured at sea, as shown in Fig. 28. As shown in Figs. 28a and 28f, the sonar images captured at sea are more complex, exhibiting highlights and shadows that are more diverse because the seabed has many natural terrains such as coral reefs, seaweeds, and rock. Among the diverse highlights, the CNN detects the crosstalk noise occurring near a rock and seabed, as shown in Figs. 28b and 28g. Subsequently, we can generate crosstalk-free images for sonar images captured at sea, as shown in Figs. 28c and 28h. Furthermore, we can extract the true highlight of the rock by eliminating the highlights of the crosstalk noise and seabed using the bounding box data detected by the CNN as shown in Figs. 28e and 28j. Compared to Figs. 28d and 28i, the extended highlights can be extracted more accurately using the proposed method. We can generate the point cloud of the rock using the extracted true highlights in Figs. 28e and 28j.

Although the training of the CNN used only images captured in the indoor water tank, the proposed method can detect the crosstalk noise and seabed in sonar images of the sea. The gradation pattern of crosstalk noise was similar to those of sonar images at sea although other characteristics such as intensity and shape were different. Therefore, the CNN can use the feature map trained by the sonar images at the indoor water tank to detect the crosstalk noise and seabed in the images of real sea. In summary, the proposed method can handle FSS images captured at the various environments. Furthermore, the proposed method can be transferred to other sonars if the sonar uses the similar imaging mechanism; thus, crosstalk occurs based on similar causes and has similar characteristics.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2925830, IEEE Sensors Journal

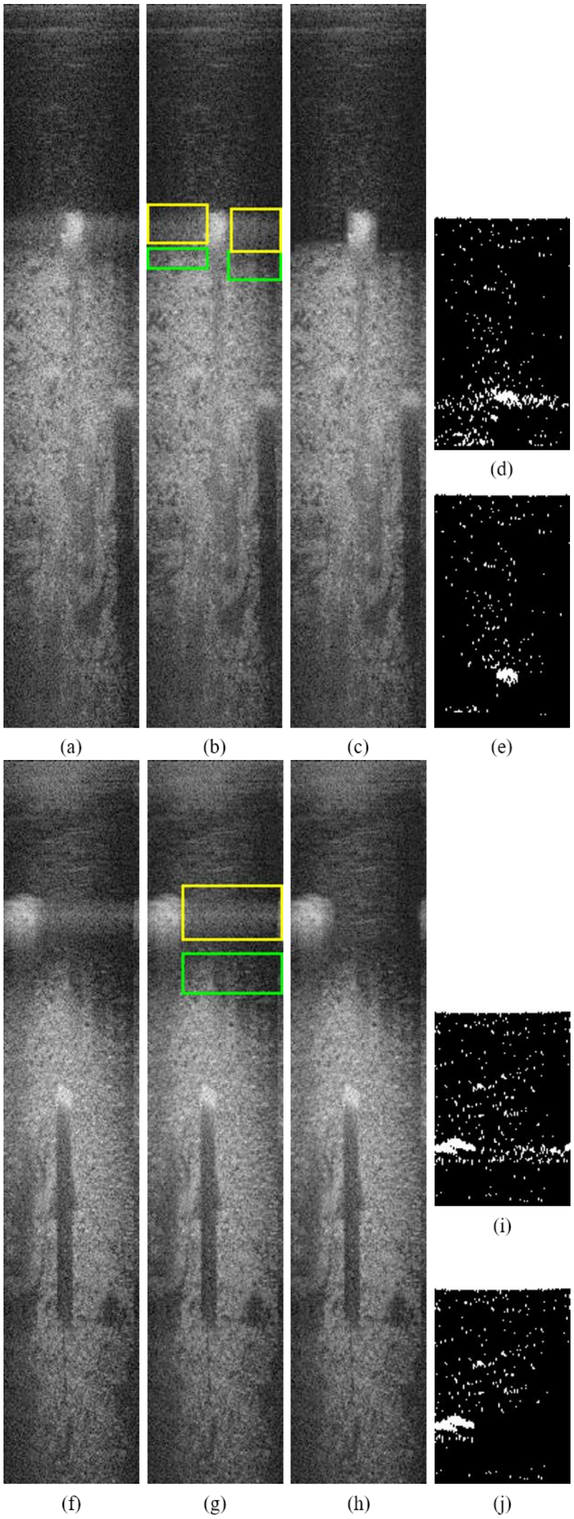


Fig. 28. Results of the proposed method in field FSS image. (a) and (f) Input images, (b) and (g) Detection results, (c) and (h) Crosstalk removal results, (d) and (i) Extracted highlight extensions without the proposed method, (e) and (j) More accurate highlights extracted by the proposed method.

VI. Conclusion

In this paper, we proposed a method to detect and remove the crosstalk noise using the CNN in the given sonar images. Because the crosstalk noise occurred in similar form regardless of the underwater object and environments, obtaining training images is relatively easy, and the trained CNN detects the crosstalk noise accurately in the given sonar images captured in various environments. Then, the proposed method removes the crosstalk noise preserving other important information from a single given image by applying the image processing algorithms on the detected region.

We applied the proposed method to the 3D point cloud generation-based object detection method to verify the performance of the proposed method. With the proposed method, we extracted the true highlight of the object and generated a more accurate 3D point cloud. Then, it is possible to recognize the underwater object by comparing the calculated 3D data and the ground truth of the target object.

Because the crosstalk noise occurs near the underwater object and distorts the highlight of the object, the crosstalk noise makes recognizing the underwater objects and landmarks difficult. The crosstalk-free sonar images generated by the proposed method can be applied to other sonar-image-based applications and enhance the reliability of those applications.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2017R1A5A1014883).

References

[1] B. Bingham, B. Foley, H. Singh, R. Camilli, K. Delaporta, R. Eustice, A. Mallios, D. Mindell, C. Roman, and D. Sakellariou, “Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle,” *Journal of Field Robotics*, vol. 27, no. 6, pp. 702–717, 2010.

[2] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, “Imaging sonar-aided navigation for autonomous underwater harbor surveillance,” in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 4396–4403,

[3] M. F. Fallon, J. Folkesson, H. McClelland, and J. J. Leonard, “Relocating underwater features autonomously using sonar-based slam,” *IEEE Journal of Oceanic Engineering*, vol. 38, no. 3, pp. 500–513, 2013.

[4] D. P. Williams, “On optimal AUV track-spacing for underwater mine detection,” in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 4755–4762.

[5] M. Couillard, J. Fawcett, and M. Davison, “Optimizing constrained search patterns for remote mine-hunting vehicles,” *IEEE Journal of Oceanic Engineering*, vol. 37, no. 1, pp. 75–84, 2012.

[6] K. J. DeMarco, M. E. West, and A. M. Howard, “Sonar-based detection and tracking of a diver for underwater human-robot interaction scenarios,” in *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2013, pp. 2378–2383.

[7] J. Pyo, H. Cho, H. Joe, T. Ura, and S.-C. Yu, “Development of hovering type auv cyclops and its performance evaluation using image mosaicing,” *Ocean Engineering*, vol. 109, pp. 517–530, 2015.

[8] R. P. Stokey, A. Roup, C. von Alt, B. Allen, N. Forrester, T. Austin, R. Goldsborough, M. Purcell, F. Jaffre, G. Packard, *et al.*, “Development of the remus 600 autonomous underwater vehicle,” in *OCEANS, 2005. Proceedings of MTS/IEEE*. IEEE, 2005, pp. 1301–1304.

[9] T. Nakatani, T. Ura, Y. Ito, J. Kojima, K. Tamura, T. Sakamaki, and Y. Nose, “Auv “tuna-sand” and its exploration of hydrothermal vents at kagoshima bay,” in *OCEANS 2008-MTS/IEEE Kobe Techno-Ocean*. IEEE, 2008, pp. 1–5

[10] H. Singh, A. Can, R. Eustice, S. Lerner, N. McPhee, and C. Roman, “Seabed auv offers new platform for high-resolution imaging,” *Eos, Transactions American Geophysical Union*, vol. 85, no. 31, pp. 289–296, 2004.

[11] H. Cho, J. Gu, and S.-C. Yu, “Robust sonar-based underwater object recognition against angle-of-view variation,” *IEEE Sensors Journal*, vol. 16, no. 4, pp. 1013–1025, 2016.

[12] B. Kim and S.-C. Yu, "Imaging sonar based real-time underwater object detection utilizing adaboost method," in *Underwater Technology (UT), 2017 IEEE*. IEEE, 2017, pp. 1–5.

[13] A. A. Bennett and J. J. Leonard, "A behavior-based approach to adaptive feature detection and following with autonomous underwater vehicles," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 2, pp. 213–226, 2000.

[14] S.-C. Yu, T.-W. Kim, G. Marani, and S. K. Choi, "Real-time 3d sonar image recognition for underwater vehicles," in *Underwater technology and workshop on scientific use of submarine cables and related technologies, 2007. Symposium on*. IEEE, 2007, pp. 142–146.

[15] A. Lorenson and D. Kraus, "3d-sonar image formation and shape recognition techniques," in OCEANS 2009-EUROPE. IEEE, 2009, pp. 1–6.

[16] H. Cho, B. Kim, and S.-C. Yu, "Auv-based underwater 3-d point cloud generation using acoustic lens-based multibeam sonar," *IEEE Journal of Oceanic Engineering*, 2017.

[17] X. Cao, R. Togneri, X. Zhang, and Y. Yu, "Convolutional neural network with second-order pooling for underwater target classification," *IEEE Sensors Journal*, 2018.

[18] S. W. Perry and L. Guan, "A recurrent neural network for detecting objects in sequences of sector-scan sonar images," *IEEE Journal of oceanic engineering*, vol. 29, no. 3, pp. 857–871, 2004.

[19] K. Denos, M. Ravaut, A. Fagette, and H.-S. Lim, "Deep learning applied to underwater mine warfare," in *OCEANS 2017-Aberdeen*. IEEE, 2017, pp. 1–7.

[20] J. Kim and S.-C. Yu, "Convolutional neural network-based real-time rov detection using forward-looking sonar image," in *Autonomous Underwater Vehicles (AUV), 2016 IEEE/OES*. IEEE, 2016, pp. 396–400.

[21] J. Kim, H. Cho, J. Pyo, B. Kim, and S.-C. Yu, "The convolution neural network based agent vehicle detection using forward-looking sonar image," in *OCEANS 2016 MTS/IEEE Monterey*. IEEE, 2016, pp. 1–5.

[22] B. Kim, H. Cho, H. Joe, and S.-C. Yu, "Optimal strategy for seabed 3d mapping of auv based on imaging sonar," in *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*. IEEE, 2018, pp. 1–5.

[23] B. Kim, H. Cho, S. Song, and S.-C. Yu, "Imaging sonar based navigation method for backtracking of auv," in *OCEANS–Anchorage, 2017*. IEEE, 2017, pp. 1–5.

[24] H. Cho, J. Pyo, and S.-C. Yu, "Drift error reduction based on the sonar image prediction and matching for underwater hovering," *IEEE Sensors Journal*, vol. 16, no. 23, pp. 8566–8577, 2016.

[25] J. Pyo, H. Cho, and S.-C. Yu, "Beam slice-based recognition method for acoustic landmark with multi-beam forward looking sonar," *IEEE Sensors Journal*, vol. 17, no. 21, pp. 7074–7085, 2017.

[26] Y. Petillot, I. T. Ruiz, and D. M. Lane, "Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 2, pp. 240–251, 2001.

[27] M. D. Aykin and S. Negahdaripour, "Forward-look 2-D sonar image formation and 3-D reconstruction," in *Oceans-San Diego, 2013*. IEEE, 2013, pp. 1–10.

[28] H. Medwin and C. S. Clay, *Fundamentals of acoustical oceanography*. Academic press, 1997.

[29] K. G. Foote, "Acoustic methods: brief review and prospects for advancing fisheries research," in The future of fisheries science in North America. Springer, 2009, pp. 313–343.

[30] J. Hsieh, J. Olsonbaker, and W. Fox, "A screening application for image data collected by an acoustic lens sonar," Washington Univ Seattle Applied Physics Lab, Tech. Rep., 2005.

[31] K. G. Foote, D. Chu, T. R. Hammar, K. C. Baldwin, L. A. Mayer, L. C. Hufnagle Jr, and J. M. Jech, "Protocols for calibrating multibeam sonar," *the Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2013–2027, 2005.

[32] J. Benesty and D. R. Morgan, "Multi-channel frequency-domain adaptive filtering," in *Acoustic Signal Processing for Telecommunication*. Springer, 2000, pp. 121–133.

[33] Y. Kim, O. Deille, and P. Nelson, "Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners," *Journal of Sound and Vibration*, vol. 297, no. 1-2, pp. 251–266, 2006.

[34] A. Asada, H. Kunishima, T. Igarashi, T. Nagase, T. Matsuda, and K. Shibata, "Advanced mosaic techniques of acoustic video images for underwater surveillance and diagnosing degradation levels of harbor structures," in *Proceedings of Underwater Acoustic Measurements: Technologies and Results*, vol. 1. Citeseer, 2009, pp. 227–234.

[35] A. Borsdorf, R. Raupach, T. Flohr, and J. Hornegger, "Wavelet based noise reduction in ct-images using correlation analysis," *IEEE transactions on medical imaging*, vol. 27, no. 12, pp. 1685–1703, 2008.

[36] B. Kim, J. Kim, M. Lee, M. Sung, and S.-C. Yu, "Active planning of auvs for 3d reconstruction of underwater object using imaging sonar," presented at the 2018 IEEE/OES AUV, Porto, Portugal, Nov. 6-9, 2018, Paper 112.

[37] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[38] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *European Conference on Computer Vision*. Springer, 2012, pp. 214–227.

[39] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[40] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[42] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[43] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[44] E. Belcher, W. Hanot, and J. Burch, "Dual-frequency identification sonar (DIDSON)," in *Underwater Technology, 2002. Proceedings of the 2002 International Symposium on*. IEEE, 2002, pp. 187–192.

[45] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.