Neural Network-Based Underwater Object Detection off the Coast of the Korean Peninsula

Won-Ki Kim, Ho Seuk Bae *, Su-Uk Son and Joong-Soo Park

Agency for Defense Development, Changwon 51678, Korea
* Correspondence: belfre@add.re.kr; Tel.: +82-55-540-6328

Abstract: Recently, neural network-based deep learning techniques have been actively applied to detect underwater objects in sonar (sound navigation and ranging) images. However, unlike optical images, acquiring sonar images is extremely time- and cost-intensive, and therefore securing sonar data and conducting related research can be rather challenging. Here, a side-scan sonar was used to obtain sonar images to detect underwater objects off the coast of the Korean Peninsula. For the detection experiments, we used an underwater mock-up model with a similar size, shape, material, and acoustic characteristics to the target object that we wished to detect. We acquired various side-scan sonar images of the mock-up object against the background of mud, sand, and rock to account for the different characteristics of the coastal and seafloor environments of the Korean Peninsula. To construct a detection network suitable for the obtained sonar images from the experiment, the performance of five types of feature extraction networks and two types of optimizers was analyzed. From the analysis results, it was confirmed that performance was achieved when DarkNet-19 was used as the feature extraction network, and ADAM was applied as the optimizer. However, it is possible that there are feature extraction network and optimizer that are more suitable for our sonar images. Therefore, further research is needed. In addition, it is expected that the performance of the modified detection network can be more improved if additional images are obtained.

Keywords: underwater object detection; sonar image; sea experiment; deep learning

1. Introduction

Sonar is a technique that allows for the detection of underwater objects by analyzing sound wave propagation in water [1]. Side-scan sonar devices continuously receive signals reflected from the seafloor, rocks, or objects while transmitting sound waves, and can therefore obtain two-dimensional images based solely on sound propagation. Therefore, data from side-scan sonars can be interpreted relatively easily and applied in various fields for different purposes. Side-scan sonars have been actively applied for various purposes, such as detecting underwater objects to maneuver and avoid the collision of unmanned surface vehicles or autonomous underwater vehicles [2,3].

Artificial intelligence has advanced considerably over the past decade, and various fields have taken advantage of this technology. Particularly, with the development of neural network-based deep learning techniques, object detection and recognition from images or videos have vastly improved, and performance is also rapidly improving as research progresses [4]. Moreover, studies on the detection and identification of underwater objects from sonar images are also being actively conducted. However, unlike optical images, acquiring sonar data is extremely time- and cost-intensive. Moreover, the characteristics of marine environments vary greatly in time and space, thus requiring constant monitoring to obtain accurate results. Collectively, these limitations make it very difficult to secure sufficient sonar data to apply deep learning techniques.
Previous studies have been conducted to detect and identify underwater objects from sonar images using machine learning techniques. Kang [5] detected and classified underwater objects by calculating similarities with previously obtained feature values after extracting object features from sonar images using a SIFT (Scale Invariant Feature Transform) algorithm. Moreover, Lee et al. [6] conducted a laboratory-level experiment to detect the outline of an object in sonar images and determine its similarity with a known set of features by applying the circle Hough transform. These previously developed machine learning techniques required humans to provide some information for object detection. Therefore, additional efforts would be required to apply this approach for the interpretation of sonar images with large temporal and spatial changes. For example, to apply this approach for the analysis of images acquired in a new environment, a person may need to provide additional information to determine similarity.

In contrast, neural network-based deep learning techniques do not require human input to train the network for object detection. Therefore, this approach may be more suitable for application to sonar images in which the signatures of objects change significantly and unpredictably. Due to this advantage, various deep learning techniques are being considered for various purposes as well as underwater object detection. Qin et al. [7] conducted a study for seabed classification using a CNN (Convolutional Neural Network)-based deep learning technique, and the authors concluded that ResNet was the most suitable among the existing public networks. Moreover, Wu et al. [8] developed a CNN algorithm for semantic segmentation of side-scan sonar data and demonstrated the applicability of the developed algorithm by comparing its performance with that of the existing U-Net, SegNet, and LinkNet algorithms.

The application of deep learning techniques in side-scan sonar images is also being used not only to identify the characteristics of the seafloor but also to generate and improve images. Sung et al. [9] and Bore and Folkesson [10] conducted research to produce more realistic sonar images based on real sonar images. Additionally, Jiang et al. [11] sought to generate multi-frequency sonar images by applying a Gaussian distribution-based feature extraction technique. Ye et al. [12] evaluated the performance of a generative adversarial network (GAN)-based compensation method instead of other existing methods (e.g., time-variant gain, histogram equalization, nonlinear compensation, function fitting) for compensating sonar signals. CNNs are known for their excellent performance in detecting objects in images and are therefore applied in various fields. Einsidler et al. [13] demonstrated the applicability of CNNs by training a network with a small amount of data using a pre-trained deep learning network. To mitigate the sensitivity of sonar images to variations in environmental conditions, Dura et al. [14] a study applying an active-learning algorithm. CNN-based deep learning techniques have previously been used to detect underwater objects for military or commercial purposes. For example, Kim et al. [15] and Palomeras et al. [16] successfully applied the CNN method to detect underwater objects and mines for military purposes. However, obtaining sufficient sonar data for training continues to be an important obstacle for the development of CNN-based recognition algorithms, and the characteristics of sonar images can vary greatly depending on the marine environment or sonar operating conditions, which further complicates object recognition.

As mentioned above, the quality of sonar images varies greatly depending on the marine environment and sonar operating conditions. Therefore, the deep learning technique can only be successfully implemented when sufficient sonar images are obtained under various marine environments and various operating conditions. This study applied a deep learning technique using side-scan sonar images in the vicinity of the Korean Peninsula. Therefore, it was essential to acquire sonar images that were representative of the characteristics of the coastal environments of the Korean Peninsula. Many studies have characterized the coastal environments of the Korean Peninsula. The coast of the Korean Peninsula is generally well-developed with continental shelves, exhibiting relatively shallow waters and a stable water depth. However, other
environmental factors such as water temperature can vary depending on the season and area [17]. Among the various marine environmental factors, sediment features are an important factor that greatly affects the interpretation of sonar images. The sediment features in the sea around the Korean Peninsula vary depending on the sea area. In the case of the Yellow Sea, the composition of sand and mud sediments of land origin is mainly influenced by the YSWC (Yellow Sea Warm Current) and JCC (Jiangsu Coastal Current). Moreover, the characteristics of the southeastern part of the Yellow Sea vary depending on the silt-to-clay ratio [18–22]. The Southern Sea differs depending on the sea area, but the sediments vary widely from silty to sandy surfaces [23]. Kim et al. [24] and Kim et al. [25] classified the southern coast into several regions according to sediment characteristics. According to the study, the sediments consist of mixtures of land-origin sediments descending along rivers, Holocene sediments, and floating sediments. In the case of the deep East Sea, the characteristics of the sediments are slightly different from those of the East Sea coast [26]. However, the East Sea coast shows similar characteristics to the southern part of the Southern Sea of Korea [27,28].

This study intended to apply a deep learning network to detect underwater object that may exist in the water off the Korean Peninsula. For this purpose, the algorithm widely used for target detection in the image was modified to be suitable for the sonar image containing the characteristics of the Korean Peninsula. To this end, we constructed a mock-up model similar in size, shape, material, and acoustic properties to the target object to be actually detected in the sea. Afterward, side-scan sonar images of the mock-up model were acquired. To account for the characteristics of the sediments in different areas of the seafloor, which is a factor that greatly affects side-scan sonar images, the experiment areas were specifically selected to be representative of the sediment features off the coast of the Korean Peninsula. Side-scan sonar images were obtained to compare the performance of the detection networks according to five types of feature extraction networks and two types of optimizers in the selected sea area. In Section 2, we compared the characteristics of the marine experiment areas, and Section 3 describes the design and acquisition of the sonar images. In Section 4, we explain the results of applying the deep learning technique to the acquired sonar images. Lastly, our discussion and conclusions are provided in Section 5.

2. Sea Environment

Representative sonar images were acquired around the Korean Peninsula to verify whether underwater objects could be successfully detected. Importantly, these images included a representative selection of sediment types that characterize the study area. Moreover, the experimental area was selected to be easily accessible by the research vessel and experimental equipment. After a preliminary investigation, we selected two experiment sites as shown in Figure 1. The selected sites are located in the southern sea of Korea, Site 1 is located near Geoje Island, and Site 2 is located near Busan. Site 1 is surrounded by islands and land, and therefore the waves are relatively weak and the water depth does not change much, mostly remaining at 20 m. Site 2 borders with land to the north and the open sea to the south. Therefore, the marine environment is quite different from Site 1. The waves in Site 2 are stronger than in Site 1, and the water depth is approximately 12 m. More importantly, the side-scan sonar can be easily operated in both experiment sites if the weather conditions are favorable.
Figure 1. Sea experiment area. The red arrows indicate the detailed locations.

Figure 2 shows the photographs of each experimental site and the sediment samples. Site 1 is surrounded by islands and land, whereas Site 2 is in the open sea toward the south side of the Korean peninsula. Therefore, the geographical characteristics of the two sites are rather distinct. Furthermore, the sediments of both sites are also quite different. As shown in Figure 2, the sediment sample of Site 1 has a darker color and smaller particles than that of Site 2. To quantitatively confirm the characteristics of the sediment features in the two sites, we analyzed the collected samples taken from each site.
Figure 2. Photographs of the experimental sites and sediment samples; (a) Site 1, (b) Site 2.

Figure 3 illustrates the ternary diagrams of sand-silt-clay for each site. These diagrams represent the proportions of sand, silt, and clay contained in sediment samples, where the black dots are analysis results. Therefore, it can be used to characterize the sediment with respect to particle size. Although the small samples obtained in our surveys cannot represent the entire study area from which they were obtained, the characteristics of the two experiment sites were significantly different according to our analyses. Figure 3a shows the results of the grain size analysis of Site 1, and we found that the sediment mainly consisted of clay and silt, which are fine-grained sediments. The particle size analysis results of Site 2 in Figure 3b show that the sand was very dominant, indicating that this site was primarily dominated by coarse sediments. Table 1 shows the average seabed features of each experimental site from the particle size analysis. The MGS (Mean Grain Size), which is reported in $-\log$ 2 scale, was 8.07 for Site 1 and 3.87 for Site 2, indicating that the average particle size of the sediments at Site 1 was relatively small. Site 1 contains very little sand (4.14%), whereas the silt and clay contents were very high. Site 2 has a gravel content of 1.36% and a sand content of 73.03%. Therefore, the two sites exhibited distinct sediment compositions. Site 1 is a seabed composed predominantly of mud, and Site 2 is a seabed composed predominantly of sand. Given that the reflectivity of sound waves in the sand with large particles is larger than that of mud, it is presumed that the acoustic properties of both sites are considerably different. These differences in acoustic characteristics are also reflected in the side-scan sonar images, and therefore we inferred that the conditions of the acquired images would likely be different. Particularly, some rocks are exposed at Site 2. Because exposed rocks have different acoustic properties from mud and sand, side-scan sonar images of the rock background were also obtained to improve detection performance.
Table 1. Averaged sediment properties.

<table>
<thead>
<tr>
<th>Area</th>
<th>MGS (φ)</th>
<th>Gravel</th>
<th>Sand</th>
<th>Silt</th>
<th>Clay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>8.07</td>
<td>0.00</td>
<td>4.14</td>
<td>39.10</td>
<td>56.76</td>
</tr>
<tr>
<td>Site 2</td>
<td>3.87</td>
<td>1.36</td>
<td>73.03</td>
<td>19.90</td>
<td>5.71</td>
</tr>
</tbody>
</table>

3. Data Acquisition

As mentioned above, sonar images with coastal characteristics are essential for detecting underwater objects in the waters off the Korean Peninsula. However, obtaining these data requires a lot of time and cost. For this reason, there has been no sonar image for the underwater object with the coastal characteristics of the Korean Peninsula. Therefore, some sea experiments are required to secure the data.

In order to obtain various side-scan sonar images containing underwater objects, we performed sea experiments against the background of representative sediment types in the vicinity of the Korean Peninsula. The size, shape, material, and acoustic properties of the mock-up underwater object were meant to be similar to those of the object that we wished to detect. Furthermore, to prevent artificial distortion of the sonar images, all ropes and buoys for installation and recovery of the object were removed by divers. The side-scan sonar for marine experiments was selected on account of it being the most widely used model in Korea, and attempts were made to obtain as many images as possible by varying the azimuth of the towing line and the separation distance from the underwater object.

3.1. Side Scan Sonar

The side-scan sonar used to image the seafloor or objects in our experiments was the ‘SeaView400S’ model, which is produced by a Korean manufacturer and is widely used by several research institutions in Korea. The images obtained with this sonar device are easy to analyze because post-processing programs such as attenuated signal correction and spatial compensation procedures for this equipment are well established. Figure 4 shows the side-scan sonar equipment used in our marine experiments including the towfish, cable, and depressor wing. The towing body is sized for human operation and transmits the acquired data to the main controller through a transmission cable. To stably tow the sonar at the desired depth, a depressor wing weighing approximately 15 kg was manufactured and operated by mounting it on the upper part of the towfish to minimize acoustic interference.

![Side scan sonar with towfish, cable, and depressor wing used in experiments.](image)

Table 2 shows the major specifications of our equipment. The sonar uses a single beam of 455 kHz, and the maximum swath is 300 m. The maximum towing speed is approximately eight knots, and the equipment can be operated up to a depth of 500 m. The transmission beam has a beam width of 40 degrees in the vertical direction and 0.2 degrees in the horizontal direction. In our experiments, we adjusted the swath range up to 120 m according to the sea environment conditions and set the towing speed to 3.5–5.5 knots depending on the weather conditions. For safety, the operating depth of the towing
body was set to approximately 10 m and 5 m, which was approximately half of the depth of experiment sites 1 and 2, respectively.

**Table 2.** Acquisition specifications of our side scan sonar (SeaView400s).

<table>
<thead>
<tr>
<th>Frequency (kHz)</th>
<th># of Beam Swath (m)</th>
<th>Tow Speed (Knots)</th>
<th>Operation Depth (m)</th>
<th>Beam Width (°)</th>
<th>Pulse Length (μs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>455</td>
<td>1</td>
<td>~300</td>
<td>~8</td>
<td>~500</td>
<td>0.2 (H)/40 (V)</td>
</tr>
</tbody>
</table>

3.2. **Underwater Mock-Up Model**

The underwater object used in the sea experiment was manufactured based on the size, shape, and material of the object to be detected in a real situation (Figure 5). The object was cylindrical with a 2400 mm length and a 500 mm diameter. Several eyebolts were included in the upper part for easy installation and recovery with ropes, and there was a separate space at the lower part in which weights can be mounted for stable settlement in water. The weights were designed to be detachable as needed. The inside of the mock-up model was filled with air with a complete watertight structure to mimic the acoustic properties of the underwater object to be detected in a real situation. To withstand deep water pressure, bearing walls were installed inside the object. The object was orange to facilitate its detection by our divers.

![Mock-up cylindrical model](image)

**Figure 5.** Mock-up cylindrical model. The interior of the mock-up was filled with air to resemble the acoustic properties of the underwater object to be detected in a real situation.

3.3. **Data Acquisition**

Sonar images were obtained using the side-scan sonar and manufactured mock-up described above at two experiment sites with different sediment features shown in Figures 1 and 2. Figure 6 shows the procedure for installing the underwater mock-up and the schematic diagram of the survey lines towing the sonar. The mock-up was installed by moving it to the sea surface using a crane and then dropping it to settle on the sea floor. Once the mock-up reached the seafloor, divers approached it and visually checked the settling condition on the seafloor, after which they removed the equipment used during the installation of the mock-up including ropes and buoys. When recovering the equipment, the divers connected the rope in the reverse order of installation, and then safely recovered it using a crane.
To obtain sonar images in various conditions, sea experiments were carried out by setting towing lines with various azimuths and distances from the underwater mock-up. A total of eleven exploration lines were set. The sixth line was located on top (center) of the mock-up, and five lines to the left and right were set at a 10 m distance from the sixth line. This enabled the acquisition of sonar images at various distances with respect to the mock-up model. Moreover, images of various directions could be obtained by rotating the exploration lines 30 degrees with respect to the true north. The length of each exploration line was set to ensure that the underwater mock-up could be fully depicted and characterized.

Sonar, which transmits sound waves and receives the reflected signals from an object, can be expressed as an active sonar equation as shown in Equation (1) [1]:

\[
SE = SL - 2TL + TS + DI - DT - NL
\]

where \(SE\) denotes a signal excess (dB) and \(SL\) represents a transmission source level (dB). \(TL\) is the transmission loss (dB), \(TS\) is the target strength (dB), \(DI\) is the detection index (dB), and \(DT\) is the detection threshold (dB). \(NL\) is the ambient noise level in the water (dB). As in the sonar equation, the signals in the sonar image are affected by propagation loss and the reflection characteristics of the sea bottom and the detected object, as well as sonar performance parameters. In other words, the characteristics of the sonar image can vary depending on the acoustic changes of the acquisition environment.

Through several sea experiments, we acquired various side-scan sonar images of the mock-up object at sites 1 and 2. We obtained 277 images containing the mock-up object with a mud background at Site 1. Moreover, 91 and 83 images containing the object with a sand and rock background were also obtained, respectively. The images were only counted when the object could be visually identified. Figure 7 shows examples of object images and traces with mud, sand, and rock sediments obtained through our experiments. The traces represent normalized magnitude in the image. These values are directly used as characteristics at each point for network training. The extracted positions of the traces presented in red and black solid lines are indicated in red and black dashed lines in the image, respectively. The left side of Figure 7a is a mock-up image obtained in a mud background, and a cylinder-shaped object is clearly observed. In the line plot on the left side of Figure 7a, the reflected signal from the mock-up (blue arrow) was identified, and its signal was stronger than that of the background. After strong reflection, a weak signal was clearly observed. The strong reflected signals are caused by the stronger reflection strength of the underwater mock-up object compared to the background (mud). In contrast, the signal weaker than the background represents a shadow zone generated by the underwater mock-up. Figure 7b is a sonar image in a sand background, which exhibited distinctly different characteristics from those in the mud background in Figure
In the sonar image, the mock-up and shadows are clearly observed. However, in the line plot, the reflected signal from the mock-up is clearly observed, whereas the shadow zones are barely visible. This means that the strength scattered from the sand and the shadow strength generated by the mock-up were similar. Figure 7c shows a mock-up image against the rock background. Both the shape and shadows of the mock-up are well observed in the image; however, only shadows are clearly distinguished from the background. The highlights of the mock-up cannot be easily distinguished because the reflection strength of the mock-up and rock is similar. However, the shadows can be clearly seen.

![Sonar images and traces of the selected line.](image)

**Figure 7.** Sonar images and traces of the selected line; the backgrounds of the images are (a) mud, (b) sand, and (c) rock. The red line indicates the object signals (blue arrow), and the black line represents an arbitrary background signal.

In summary, our findings indicated that the reflected strength of the object varies depending on the sediment type. Therefore, the sediment features may affect the training performance of the CNN. Given that highlights and shadows vary according to the sediment characteristics, we considered both highlights and shadows as training features for underwater object detection.

From the sea experiments, sonar images including the characteristics of the waters off the Korean Peninsula and underwater object to be detected were acquired. It means that deep learning network training is possible because the sonar image data for training of the deep learning network for underwater object detection in Korean Peninsular have been secured. In addition, it was confirmed that the characteristics of the highlights and shadows of underwater objects in the sonar image were changed as the different acoustic
characteristics of the sediments. From the results, it was identified that not only highlights of underwater objects but also shadows of underwater objects should be considered for deep learning network learning.

4. Deep Learning Application

In this study, we applied the YOLO (You Only Look Once) algorithm, which is known for its fast processing speed and excellent accuracy among various object detection algorithms. Unlike other two-stage algorithms such as R-CNN, YOLO is a one-stage algorithm [29]. This algorithm is still being upgraded continuously and multiple versions of YOLO have been released. In this study, we applied YOLO v2 to construct a YOLO-based network [30].

Figure 8 shows a schematic diagram of the YOLO-based detection network applied in this study. When the underwater object image was used as an input, the features were extracted from the feature extraction network. An existing CNN can be used as the feature extraction network, and an appropriate network can be selected according to the data. Afterward, object detection and identification are conducted based on the extracted features. In this study, images of the mock-up obtained from marine experiments with representative sediment types in the vicinity of the Korean Peninsula were labeled. Additionally, the performance was compared by changing the feature extraction network to establish a network with optimal performance for underwater object detection in the coastal waters of Korea. Furthermore, an appropriate optimizer was selected by comparing ADAM and SGDM among the optimization techniques.

![Figure 8. Schematic diagram of YOLO (You Only Look Once)-based detection network for a side scan sonar image.](image)

In this study, a deep learning algorithm (YOLO), which has been widely used for object detection in images, was applied. However, it may not be suitable for sonar images, and training was not performed on the cylindrical underwater object to be detected in this study. As for the sonar image, the characteristics of the image change greatly because the sea varies greatly in time and space. Therefore, even though the basic algorithm is the same, it needs to be modified to fit the sonar image, and training for a cylindrical underwater object must be newly performed. For these reasons, we tried to build a deep learning network for detecting cylindrical underwater objects in sonar images by comparing the performance of five types of feature extraction networks and two types of optimizers instead of using the existing pre-trained network.
4.1. Data Preprocessing and Network Setting

We first performed data preprocessing for network training for the detection of our underwater object. The input images were resized to 256 × 256 pixels to avoid overloading the memory of the computer in which the analyses were performed. In order to train the network, it is necessary to provide object characteristics information by indicating the object to be detected. To this end, by boxing an underwater object in each image, the coordinates are provided during training. The labeling results such as area and aspect ratio are useful to understand the characteristics of object in the image. An in-house developed labeling tool was used for labeling images for training. Figure 9a is an example of labeling for training, which illustrates the label of the mock-up against the sand background. Figure 9b shows the results of labeling for all data as a graph of the labeling box area and aspect ratio. The labeling boxes of the images acquired from the mud background (black circle) are distributed in the specific box area under the 1000 value. In contrast, the labeling boxes of the images obtained from the sand (red circle) and rock (blue circle) backgrounds are widely scattered. The change in the box area and aspect ratio is caused by various factors such as the operating conditions of the sonar and the weather. In fact, sonar images, unlike optical images, can vary widely depending on the imaging conditions. This is because the sound is refracted in water and several parameters such as pitch, roll, and sonar equipment can vary widely according to the weather and other factors. The area of the box is mainly distributed between 500 and 1000. Particularly, sonar images with a mud background were concentrated in that area. Regarding the aspect ratio of the labeling box, most of the boxes did not exhibit a ratio of 1.0 due to the shape of the cylinder.

![Figure 9. (a) Labeling example, (b) labeling result, and (c) anchor boxes applied to the network.](image)

The YOLO algorithm uses an anchor box to detect suspicious objects in images. Anchor box serves as a reference window when searching for objects in the image in the YOLO network. The anchor box should cover the object to be detected with a variety of sizes and aspect ratios. Figure 9c shows the 20 anchor boxes selected based on the size of the labeling box. As illustrated in the figure, the anchors come in a variety of sizes and aspect ratios. The YOLO algorithm extracts features using convolutional layers. Therefore, it is necessary to apply layers that can effectively extract features from the training data. To compare the adequacy for our sonar images, we analyzed the performance of five published networks (AlexNet, ResNet50, GoogleNet, DarkNet-19, and SqueezeNet) for the feature extraction network that are actively used for image classification [29,31–34].

We next sought to identify a proper optimizer by applying two methods, after which we compared the performance of the feature extraction network. Here, we selected ADAM and SGDM, which are known to have excellent performance in deep learning research [35,36]:

- ADAM:
  \[
  \min \left\{ \frac{1}{n} \sum_{i=1}^{n} L(y_i, \hat{y}_i) : w \in \mathbb{R}^d \right\}
  \]
  where \( L(y_i, \hat{y}_i) \) is the loss function, \( y_i \) is the true label, and \( \hat{y}_i \) is the predicted value.

- SGDM:
  \[
  \min \left\{ \frac{1}{n} \sum_{i=1}^{n} L(y_i, \hat{y}_i) : w \in \mathbb{R}^d \right\}
  \]
  where \( L(y_i, \hat{y}_i) \) is the loss function, \( y_i \) is the true label, and \( \hat{y}_i \) is the predicted value.
\[ M(t) = \beta_1(t - 1) + (1 - \beta_1) \frac{\partial}{\partial W(t)} \text{Cost}(w(t)) \]  
\[ V(t) = \beta_2 V(t - 1) + (1 - \beta_2) \left( \frac{\partial}{\partial W(t)} \text{Cost}(w(i)) \right)^2 \]  
\[ W(t + 1) = W(t) + \alpha \frac{\tilde{M}(t)}{\sqrt{V(t)} + \varepsilon} \]  

Equations (2)–(4) describe the ADAM optimizer. \( W \) denotes a weight to be updated, whereas \( \alpha \) denotes a learning rate. \( M \) in Equation (2) and \( V \) in Equation (3) are terms related to the Adagrad and RMSProp techniques, respectively. \( \beta_1 \) and \( \beta_2 \) are weight factors. This is one of the main advantages of the ADAM, as it is able to consider both methods at the same time. The SGDM is a kind of SGD that considers the momentum and facilitates the detection of the global minimum by considering the momentum \( m \) in the general SGD formula as shown in Equation (5). The weight factor can be updated via Equation (6):

\[ V(t) = m \ast V(t - 1) - \alpha \frac{\partial}{\partial W} \text{Cost}(w) \]  
\[ W(t + 1) = W(t) + V(t) \]  

4.2. Network Training

We applied five types of feature extraction CNNs and two types of optimization techniques to the YOLO-based deep learning algorithm to detect underwater objects in the side scan sonar image, after which we compared their performance. The detailed setting parameters are summarized in Table 3. The data used for training consisted of a total of 451 images, which were divided into 316 (70%) for training data, 45 (10%) for validation data, and 90 (20%) for test data. The learning rate was set as low as 0.00005 to avoid local minima, and the batch size was set to 16 to avoid overloading the computer memory and to maximize efficiency. Moreover, the epoch number was set to 600 to ensure sufficient convergence.

<table>
<thead>
<tr>
<th># of Images</th>
<th>Feature Extraction Network</th>
<th>Optimizer</th>
<th>Learning Rate</th>
<th># of Epoch</th>
<th>Batch Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>316 (training)</td>
<td>AlexNet</td>
<td>ADAM</td>
<td>0.00005</td>
<td>600</td>
<td>16</td>
</tr>
<tr>
<td>45 (validation)</td>
<td>ResNet50</td>
<td>SGDM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90 (test)</td>
<td>GoogleNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DarkNet-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SqueezeNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In order to estimate the trained network, the performance was evaluated using test data that were not used for training. Figure 10 shows a graph comparing the average precisions according to the feature extraction networks and optimizers. In the figure, the ADAM and SGDM optimizers are indicated in blue and red, respectively. When we used AlexNet as the feature extraction network, the performance change according to the optimizer was relatively low. In contrast, in the case of the GoogleNet and SqueezeNet networks, the average precision was relatively high when SGDM was used as the optimizer rather than ADAM. When we selected the ResNet50 and DarkNet-19 networks, the performance was better with the ADAM optimizer. Among them, when we used DarkNet-19 as the feature extraction network and ADAM as the optimizer, the average
accuracy was approximately 0.77, showing the best performance. Conversely, the lowest precision was obtained when we used the ResNet50 network and SGDM optimizer. Therefore, among the five networks, DarkNet-19 is evaluated to be the most suitable as a feature extraction network, and ADAM is more suitable than SGDM as an optimizer for our sonar images obtained in the coastal sea of Korea.

![Figure 10. Average precision according to feature extraction networks and optimizers.](image)

Figure 10. Average precision according to feature extraction networks and optimizers.

Figure 11a shows the precision–recall curve when the threshold is 0.5. Here, we compared only three cases with excellent precision in Figure 10. The performance of the precision–recall curve can be evaluated by the area, and our findings confirmed that the best performance was achieved when the DarkNet-19 network and ADAM optimizer were selected together (black line). In the figure, the recall means the ratio of detected underwater objects among the test data used for analysis, and the precision means the ratio of actual underwater objects among the predicted underwater objects. When the recall is 0.2, the solid black line stays 1.0 while the other solid lines show about 0.86. As the recall increases, it can be seen that the DarkNet-19 with ADAM has the best performance based on the precision. This means that the false alarm rate is also the lowest. Moreover, when we used DarkNet-19 and ADAM as shown in Figure 11b, the training loss and RMSE (root-mean square error) stably converged to the global minimum as the training progressed.
Figure 11. (a) Precision–recall curves at a 0.5 of threshold and (b) training RMSE and loss curves for DarkNet-19 network and ADAM optimizer.

Table 4 shows the specification of hardware that has been trained for deep learning and the training time for three networks with excellent performance shown in Figure 11. The training time will vary depending on the training options in Table 3 and computer performance. Here, a relative comparison of the obtained training time under the same conditions is possible. Comparing the training time calculated, it was similar to 3777 s and 3686 s for training of networks equipped with GoogleNet and DarkNet-19, respectively. On the other hand, when SqueezeNet is applied, training time is 2957 s, which is about 800 s shorter. The reason is considered to be that the network structure of SqueezeNet is relatively shallow compared to those of GoogleNet and DarkNet-19. However, considering the performance, it is not fast enough to consider SqueezeNet.

Table 4. Hardware specification and training time.

<table>
<thead>
<tr>
<th>HW Specification</th>
<th>CPU</th>
<th>GPU</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intel(R) Xeon(R) Silver 4210 @ 2.20 GHz (2 processor)</td>
<td>NVIDIA GeForce RTX 2080 Ti</td>
<td>32 GB</td>
</tr>
<tr>
<td>Training Time (s)</td>
<td>GoogleNet (SGDM)</td>
<td>DarkNet-19 (ADAM)</td>
<td>SqueezeNet (SGDM)</td>
</tr>
<tr>
<td></td>
<td>3777.3</td>
<td>3685.9</td>
<td>2957.3</td>
</tr>
</tbody>
</table>
Figure 12 shows examples of the detection results of mock-up with mud, sand, and rock background in the side-scan sonar images through the trained network by applying the DarkNet-19 and ADAM optimizer. Notably, our network exhibited acceptable detection performance and precision for the target underwater object in each background. Therefore, we confirmed that our modified YOLO network with DarkNet-19 and ADAM optimizer is available as a network for detecting cylindrical underwater objects in side-scan sonar images acquired around the Korean Peninsula.

![Figure 12](image)

**Figure 12.** Examples of the detection results of the mock-up object with (a) mud, (b) sand, and (c) rock background through the YOLO network by applying the DarkNet-19 network and ADAM optimizer.

However, these positive results do not necessarily mean that the deep learning network proposed herein has an optimal training and overall performance. The detection accuracy with the mud, sand, and rock backgrounds were approximately 0.86, 0.95, and 0.94, respectively. The reason why the detection accuracy in the mud background was lower than that in the other backgrounds was because there was a large number of training images with a mud background, which enabled the algorithm to make more conservative predictions based on these conditions. In contrast, far fewer images with sand and rock backgrounds were obtained, resulting in higher detection accuracies due to overestimation. This means that the amount of learning data was insufficient, and training was therefore biased for specific conditions. That is, a high detection performance cannot be guaranteed for sonar images having different conditions. Therefore, it is necessary to acquire additional sonar images under a wider diversity of conditions. Moreover, an equal number of images should ideally be acquired for each condition to avoid biases. Furthermore, the modified YOLO network was constructed by comparing five kinds of extraction networks and two types of optimizers. This means that more diverse comparative studies are needed because more efficient feature extraction networks and optimizers may exist.

5. Conclusions

Our study evaluated the applicability of deep learning techniques to detect underwater objects in side-scan sonar images around the Korean Peninsula. We selected marine experiment areas with mud, sand, and rock to obtain a representative depiction of the characteristics of the seafloor off the coast of Korea, and conducted sea experiments to acquire sonar images. To obtain more realistic sonar images, we fabricated a mock-up model similar in size, shape, material, and acoustic characteristics to the underwater object we wished to detect.

Through several sea experiments, we acquired a variety of mock-up sonar images with mud, sand, and rock backgrounds. In the sonar image, we confirmed that the highlight signal and shadow features from the object were different depending on the
background of the seafloor, and we considered both highlights and shadows as features for the detection of the underwater object.

The YOLO-based deep learning network was modified and trained using the acquired sonar images. To build an acceptable object detection network, we compared five types of feature extraction CNN networks and two types of optimizers. From the results, the DarkNet-19 network coupled with the ADAM optimizer had the best performance in identifying the underwater object in the side-scan sonar images used in this study. Therefore, the modified detection network could be applied to detect cylindrical objects off the coast of Korea.

However, the number of sonar images obtained from the sea experiments was somewhat insufficient for training purposes. Particularly, the number of sonar images with the sand and rock backgrounds was very small, and therefore the network training was biased toward specific conditions. Therefore, sufficient sonar images under various conditions must be acquired to improve the performance of the proposed deep learning network for the detection of underwater objects off the coast of the Korean Peninsula. In addition, since there is a possibility that a network more suitable for our sonar data exists, it is necessary to conduct a performance analysis for more various networks and optimizers.

**Author Contributions:** W.-K.K., H.S.B., and S.-U.S. conducted the experiments, analyses, and wrote the manuscript under the supervision of J.-S.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Korean Government.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**