SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 1.1, (2026), pp. 4081-4092
Open Access. Online & Print

DOI: 10.5281/zenodo. 121126303

TWO-STAGE FRAMEWORK FOR OBJECT DETECTION IN SIDE SCAN SONAR IMAGES USING HOG AND RESNET FEATURES

Venkata Lakshmi Keerthi K[1], Vijayalakshmi P[2]*, Rajendran V[3]

[1]*Research Scholar, Department of ECE, Vels Institute of Science, Technology, and Advanced Studies (VISTAS), Chennai, India.*
[2]*Associate Professor, Department of ECE, Vels Institute of Science, Technology, and Advanced Studies (VISTAS), Chennai, India.*
[3]*Professor & Director, Department of ECE, Vels Institute of Science, Technology, and Advanced Studies (VISTAS), Chennai, India.*

Received: 01/12/2025
Accepted: 02/01/2026

Corresponding author: Vijayalakshmi P
(viji.se@velsuniv.ac.in)

ABSTRACT

Side-scan Sonars are being used in submarine search and rescue missions to detect drowning individuals, aircraft, and wrecks. Side scan object detection sonar imaging is complex and significantly affects military operations. The present approach involves subject-matter specialists examining sonar images to identify and determine the location of objects. In this paper, an underwater object detection framework through two-stage feature matching is developed using a Histogram of Gradient and pre-trained ResNet features. The proposed framework detects the object using patch feature matching. In the first stage, the Histogram of Gradient features is utilized for feature matching, and pre-trained ResNet features are used in the second stage. Feature matching is conducted in both phases by calculating the cosine similarity between the images in the database and the given side scan sonar image. Then, the maximum overlapping patches among the maximum scored patches in the two stages are the detected object patches. Finally, detected patches are merged, and their centroid is used to locate the object. The proposed object detection method achieves state-of-the-art efficiency in experimental findings, with a precision score of 72.67% and a detection speed of 29.4 frames per second.

KEYWORDS: Side-Scan Sonar, Feature Extraction, Feature Matching, Object Detection

1. INTRODUCTION

Side-scan sonars can provide detailed photographs of the ocean floor, even in water without visibility. These capabilities make autonomous underwater vehicles (AUVs) essential in various military applications, such as ocean mapping, mine countermeasures, underwater search and rescue, offshore oil exploration, and civilian applications [1-3]. Synthetic aperture sonar (SAS) and side-scan sonar (SSS) systems represent key sensor technologies in seafloor exploration and analysis. These systems are commonly employed to detect unexploded ordnance (UXO), locate shipwrecks, and delineate various seafloor compositions. These sensors, mounted on towed vehicles or AUVs, transmit acoustic signals and capture the resulting backscattered signals for analysis [4]. Extensive studies have been conducted on identifying objects in SSS images using manual examination and automated data processing. Manual data inspection is time-consuming, and determining the object is difficult since objects of interest may vary in real-time scenarios. Manual detection requires segmentation and annotation, which is a complex task [5]. Automatic analysis and object detection in SSS images involve feature extraction and matching algorithms. Several algorithms have performed exceptionally in this area [6-9].

The performance of object detection depends on robust feature extraction and matching. Traditional object detection techniques in SSS images perform feature matching by extracting features using handcrafted and deep features. Handcrafted feature-based object detection utilizes Discrete Wavelet Transform (DWT), Speeded-up Robust Features (SURF), Histogram of gradients (HOG), Scale-invariant feature transform (SIFT), and local binary pattern (LBP) features for feature matching [10]. Dong et al. (2019) introduced two-stage feature matching for object detection using convolutional neural networks (CNN) and SIFT features [11]. This approach utilizes LeNet5 (CNN) features for automatic object analysis, and for further recognition, feature matching is carried out using scale-invariant feature transform (SIFT). Hatanaka et .2010 proposed an algorithm based on discrete wavelet transform features to classify SSS images [12]. Quanhong et al. (2024) presented a novel methodology for object detection in SSS images, employing the YOLO (You Only Look Once) network. The dataset is subjected to preprocessing utilizing two-dimensional discrete wavelet decomposition to fortify the network's resilience and efficacy with reconstruction. This approach underscores the potential for heightened performance in object detection within SSS images [13]. Tao et al. (2010) proposed a SURF algorithm for image-matching and object detection in SSS images [14]. This method improves the narrow strip processing measures and edge-preserving filtering. Manonmani et al. (2021) developed a technique to detect mine-like objects underwater using HOGs and a Canny Edge Detector [15]. Patch feature matching is performed between the dataset and input image for object detection. Annalakshmi et al. (2019) introduced ocean sediment classification in SSS images by extracting texture features using LBP, Local Ternary Pattern (LTD), and Local Directional Pattern (LDP)[16]. Recent techniques, both hand-crafted features and deep feature techniques, have been developed for robust object detection.

Denos et al. (2017), addressed the problem of identifying objects in underwater images captured by Synthetic Aperture Sonar (SAS) [17]. To handle the complexity, the authors developed realistic synthetic image datasets for training the machine learning algorithm. Second, a deep learning approach was developed for automatic mine classification in underwater sonar images. Further, the heatmap is generated for an actual SAS image to extract the background snippets without an object using the reconstruction error of the autoencoder. The Visual Geometry Group (VGGCNN) trains the background and simulated target snippets. However, this method fails to detect objects due to autoencoder filtering without a background class. McKay et al. (2017) proposed objection detection in SSS images using AlexNet features, and the score is calculated using a Support Vector Machine (SVM) for large sonar image patches to identify the presence of objects [18]. Einsidler et al. (2018). Developed deep learning techniques for detecting objects and rocks on the seafloor using YOLOv2 trained by the collected dataset. No performance evaluation is carried out on experimental results [19]. Xu et al. (2019) developed a technique for shipwreck detection using YOLOv1 and Faster R-CNN [20] networks. The synthetic dataset is used for experimental performance due to the lack of a real-time data set; both YOLOv1 and Faster R-CNN have better performance in shipwreck detection in SSS images [20].

Feldens et al. (2019) used small patches from SSS images to train RetinaNet to detect rocks underwater and showed better performance [21]. Further, Feldens et al. (2020) extended this work by increasing the resolution of SSS images by a single-stage residual network instead of using small patches, improving the detection of rocks in the water [22]. Jiang et al. (2020) proposed an active learning approach using sampling and local information comparison for shipwreck detection using R-CNN. This framework employs the R-CNN and SSD

features, outperforming YOLOv1 [23]. Yu et al. (2021) developed an object detection mechanism using YOLOv5s with a multi-head self-attention module to increase performance and reduce computational complexity [24]. The main drawback is that the dataset used for training needs to be specified. Berthomier et al. (2019) analyzed this technique using CNN, which is trained to classify mine objects and clutter. However, these techniques require large SAS images and must be improved for complex object detection in the presence of ripples on the seafloor [25].

Le et al. (2020) developed an architecture for multi-scale object detection using Gabor CNN, similar to YOLOv3. Gabor filters are used for convolution to improve the detection performance. This technique performs better than the abovementioned techniques regarding false alarm rate and computational speed [26]. Fawcett et al. (2021) proposed a mine-like object detection framework using MiNet, similar to YOLO. For training, real-time and synthetic data sets are used [27]. This technique of quantitative evaluation performance is not included. Berthomier et al. (2019) work deals with SAS imagery object detection using CNNs [25]. The performance evaluations are compared with a small CNN, YOLOv2, and YOLOv3. Steiniger et al. (2021a) compared small CNN, YOLOv2, and YOLOv3-based algorithms for object detection in SSS images [28]. The performance analysis shows that YOLOv3 showed the best. Finally, the lack of SSS datasets and consistent metrics makes it difficult to identify the best method. In the literature, a standard metric like average precision is used to compare state-of-art techniques' performance fairly.

Many real-world applications need both top precision and real-time speed. So, an excellent appearance-based object detection should meet both needs. Handcrafted feature-based techniques are fast at computing and can detect real-time. But their detection precision is lower. This is because they are not robust to significant appearance changes and do not capture object meanings. Deep features-based object detectors excel at accuracy. This is because they have a high-level feature hierarchy. But, since it takes a lot of computing power to update CNN's many parameters, its detection speed is relatively slow. This work develops object detection in the SSS images framework. It is based on the HOG and pre-trained ResNet features. It uses two-stage similarity estimation. The highest-scored patches are obtained by performing feature matching between the dataset and the input SSS image. Finally, the detected object patch is obtained by maximizing overlap among the highest-scoring patches.

The main contributions of this work are summarized as follows.

1. Implementing a two-stage patch feature matching technique enhances object detection accuracy.
2. Two-stage feature matching, similarity estimation, and overlap maximization successfully identify comparable items and objects in complex seabed environments.
3. To decrease computational complexity, a method for object identification using two-stage patch feature matching and overlap maximization is developed.
4. The experimental findings on the SSS dataset demonstrate that the suggested approach is very proficient at detecting items under intricate seabed conditions.

The rest of the paper is organized as follows: Sect. 2 discusses the two-stage object detection technique and Sect. 3 illustrates the experimental results using an SSS dataset.

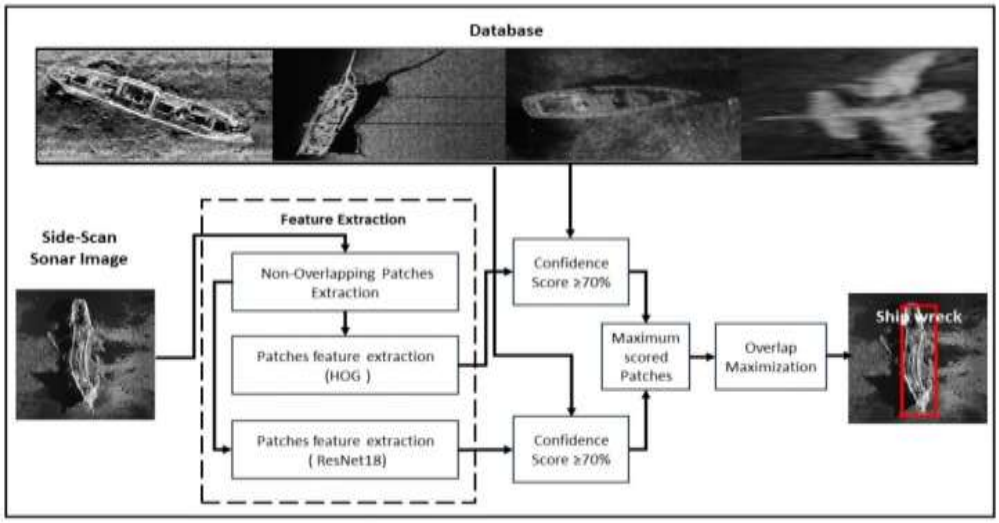

Figure 1: The proposed object detection in SSS images Framework

2. THE PROPOSED OBJECT DETECTION FRAMEWORK

Place The proposed framework for object detection in SSS images is shown in Figure 1. Detecting target objects using patch feature matching is performed in two stages. The first stage employs a Histogram of gradient (HOG) features. Feature matching is performed between the non-overlapping patch features in the SSS image and the database images using cosine similarity. The second stage of feature matching is performed using pre-trained ResNet18.The maximum similarity-scored patch greater than the threshold value is applied for overlap maximization. The patches with maximum overlap with the database image are merged and considered a detected object. The detected patch location bounds the object in the SSB image.

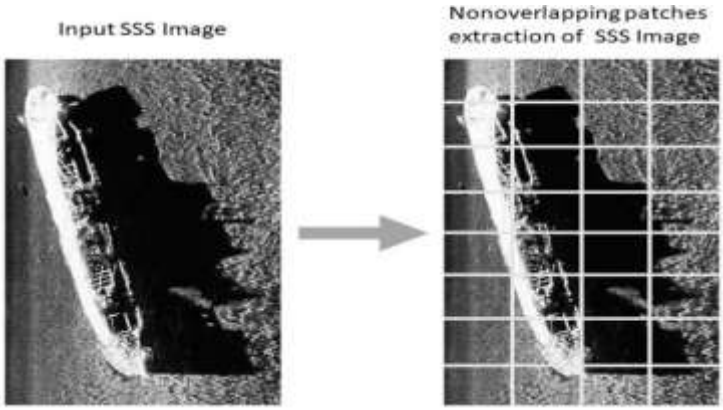

Figure 2: Nonoverlapping patches extraction.

2.1. Non-Overlapping Templates Extraction

The location and characteristics of the identified object are to be ascertained through SSS object detection. The patch feature matching technique finds the target object in the input SSS image by comparing it with the identified object patch in the SSS database. The detected object's location is determined by comparing the identified object patch with similarity measures between SSS database features and non-overlapping template features of the input SSS image. The non-overlapping patches are extracted by partitioning the image equally in horizontal and vertical directions, as shown in Figure 2.

The input image nonoverlapping patches set Pn is given by

$$P_n = [P_1, P_{2.}, P_3, P_4 \dots \dots P_k] \quad (1)$$

where $k = \frac{m}{8} \times \frac{n}{4}$ =number of overlapping patches, $P_1, P_{2.}, P_3, P_4 \dots \dots P_k$ are the nonoverlapping patches extracted, m=Number of rows, and n=Number of columns of SSS input image.

2.2. Feature Extraction

The effectiveness of object recognition techniques in computer vision relies heavily on feature extraction. Extracting features is a challenging task in visual applications. The determination of similarity structures between templates, segments, and patches is achieved through feature matching. The object detection framework utilizes a two-stage feature extraction process. During the initial phase, feature vectors are extracted from a histogram of gradient features. In 2005, Dalal and Triggs et al. introduced the representation of HOGs for detecting humans [29]. This feature is founded on the concept that the description of an object's appearance and shape can be achieved by organizing local intensity gradients or edge orientations, even without precise information regarding their specific locations. The HOG feature allows for visualizing an area, focusing on its distribution rather than the entire image. To ascertain if the object target region lies within the local HOG feature extraction range, it is necessary to assess its inclusion. The window was divided into smaller sections called blocks and cells. The creation of the fundamental orientation histogram involves the combination of 1-D histograms at each cell level. Extracting features from an object using the Histogram of Gradients (HOG) method involves five steps: Implementing global image normalization equalization to mitigate illumination effects. Gamma compression typically requires the computation of the square root for each color channel. Next, to calculate the each image window f(x,y) gradient vector, follow these steps [29]:

$$G_h(x, y) = f(x + 1, y) - f(x - 1, y) \quad \forall\, x, y \quad (2)$$

$$G_v(x. y) = f(x, y + 1) - f(x, y - 1) \quad \forall\, x, y \quad (3)$$

where $G_h(x, y)$ is the horizontal gradient of the image, and $G_v(x, y)$ is the vertical gradient of the image.

The gradient strength $M(x. y)$ of the image is given by

$$M(x.y) = \sqrt{G_v(x.y)^2 + G_h(x.y)^2} \quad (4)$$

The gradient direction $\theta(x.y)$of the image is given by

$$\theta(x,y) = \left(\frac{G_h(x.y)}{G_v(x.y)}\right) \quad (5)$$

These gather information regarding contours and other textures, making them even more resistant to changes in illumination. The locally dominant color channel is utilized to achieve a significant level of color invariance.

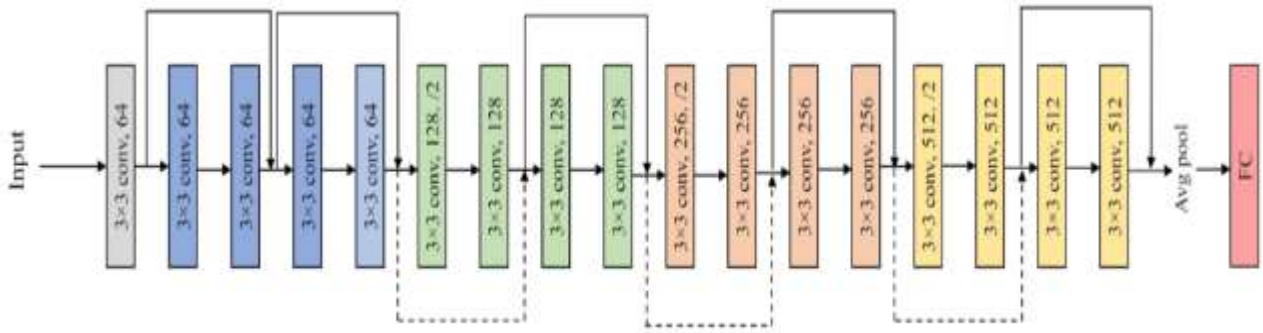

Figure 3: ResNet18 Architecture.

Additionally, it is essential to gather gradient orientation information localized, similar to SIFT. The window is partitioned into cells. The gradient across each cell pixels generates a local one-dimensional histogram. Create an encoding method that considers specific details within an image while handling slight variations in appearance and pose. The gradient angles range was divided into predetermined bins. The magnitudes of the pixel gradients in the cell determine the orientation histogram. Additionally, the normalization process includes the crucial step of comparing the responses of local groupings of cells and making necessary adjustments before proceeding further. The consistency of illumination, shadowing, and edge contrast is enhanced through normalization. The method involves collecting a local histogram over blocks. The results normalize every cell within the block. Cells are often shared among multiple blocks, but the normalizations may vary from block to block—the numerous types of cells in the final output vector, each with its unique normalization. HOG descriptors are widely recognized as normalized block descriptors. The HOG descriptors are collected from each block within a dense overlapping grid that covers the track window. The descriptors are merged to create a feature vector for further analysis. Furthermore, the HOG descriptors from each block should be combined in a deep overlapping grid of blocks that extends the track window. This will result in a single feature vector that can be used. In the initial stage, the feature vector f_h of the patch is represented as F_{HOG}.

$$f_h = \{F_{HOG}\} \quad (6)$$

In the second stage, patch features are extracted using pre-trained ResNet18 [30,31,32]. Several pre-trained CNNs, such as InceptionV3, ResNet18, Xception, and DenseNet121, have demonstrated their robustness in object detection in various standard datasets. The features of ResNet18 have shown their strength in multiple applications, such as object detection, image retrieval, and object classification techniques. ResNet18 features are optimal for patch feature extraction because they maintain the same image dimension size during training. Deep neural networks typically employ multiple layers to extract intricate features, improving accuracy. Still, errors in the network layers degrade the training process, leading to an overfitting problem. This problem is solved by residual blocks in the architecture of ResNet18, as shown in Figure 3. For Resnet18 to work correctly, the input image size should be at least 224 × 224. The model comprises layers of two down-sampling and 16 convolutions and is fully connected. The first convolution kernel size is 7 × 7, while the following are 3 × 3. After performing average pooling, the fully connected layer produces a feature map of eigenvectors. The figure illustrates two convolution layers with identical colors; combining two components creates a residual block. Figure 2 showcases the curved shortcut connections bypassing two layers and increasing the dimensions with dotted shortcuts.

The second stage patch features vector f_r extracted using pre-trained Resnet18, denoted as $F_{ResNet18}$.

$$f_r = \{F_{ResNet18}\} \quad (7)$$

2.3. Similarity Score

In this work, feature matching uses cosine similarity between the overlapping patches features in the database image features and the input SSS image. The maximum scored patches greater than the threshold are utilized to extract object patch location using overlap maximization. The metric cosine

similarity measures the feature vectors' similarity irrespective of size. In high-dimensional spaces, due to the curse of dimensionality, the Euclidean distance between vectors chances to lead closer to other high-dimensional feature vectors. Cosine similarity is unaffected and less sensitive to the dimensionality of the vectors. Cosine similarity is more advantageous than Euclidean distance regarding dimensionality, magnitude independence, document similarity, and sparsity. The cosine similarity calculation divides two vectors' dot product by Euclidean norms product. The two feature vectors, P and Q, cosine similarity, are expressed in Equation 7.

$$S = \frac{P \,.\, Q}{\|P\|\|Q\|} = \frac{\sum_{k=1}^{m} P(k)\, Q(k)}{\sqrt{\sum_{k=1}^{m}(P(k))^2}\sqrt{\sum_{k=1}^{m}(Q(K))^2}} \quad (8)$$

where m = the total number of elements in the vector P; S represents the cosine similarity; $P \,.\, Q$= the dot product of two vectors, P and Q; $\|P\|\|Q\|$ = product of the Euclidean norms of vectors P and Q;

2.4. Overlap Maximization

The object in the SSS image is determined by the highest intersection over union (IOU) overlap, which refers to the overlap ratio among the patches with the highest scores. Calculating the overlapping ratio involves dividing the area of overlap by the area of union. The identified database object patch and bounding box locations are then applied.

$$IOU(i) = \frac{Area\ of\ overlap(groundtruth(DOP), groundtruth(Somax(i)))}{Area\ of\ Union(groundtruth(DOP), groundtruth(Somax(i)))} \quad (9)$$

where $Somax(i)$ denotes ith maximum scored patch in the set Smax; $IOU(i)$ denotes the IOU of the maximum scored patch of the set Somax, and DOP denotes the identified database object patch.

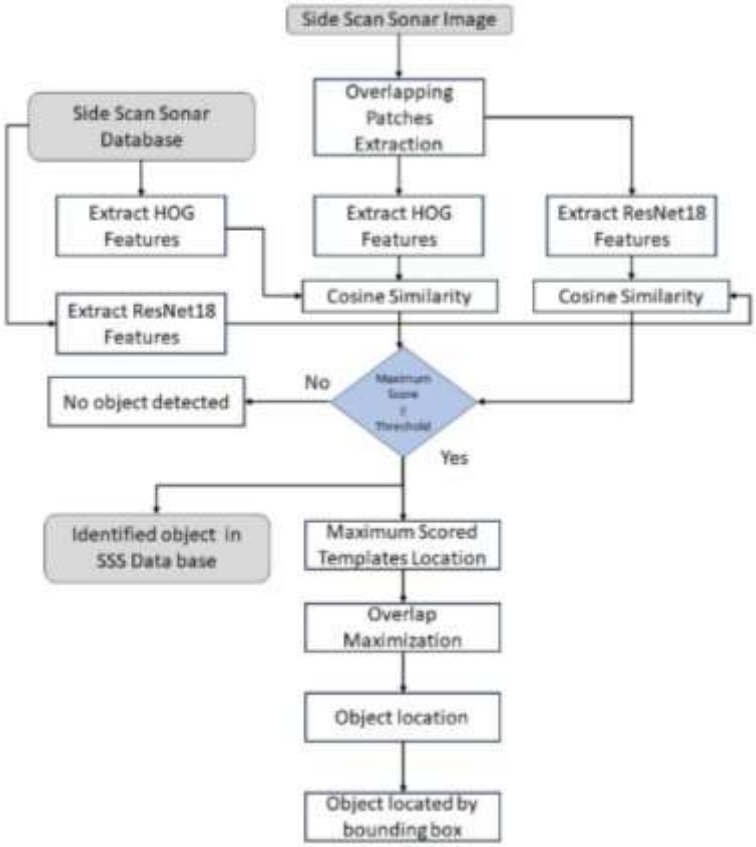

Figure 4: The proposed two-stage object detection framework flow chart.

2.5. Object Detection

The proposed technique for object detection involves conducting a maximum score search in the feature space by performing patch matching with the abovementioned features. The process of patch matching consists of calculating a cosine similarity score between the feature vectors of the database object patch and the overlapping patch features from the SSS image. During the initial stage of the framework, features are extracted using HOG features. In the subsequent stage, pretrained ResNet18 features are employed for further feature extraction. Figure 4 depicts the flowchart of the SSS object detection.

Using Equations 1 and 6, the first stage feature vector set of the overlapping patches using HOG features is formed as shown in Equation 10:

$$F_{hn} = [f_{h1}, f_{h2}, f_{h3}, \ldots . f_{hk}] \quad (10)$$

where F_{hn} represents the input image nonoverlapping patches feature vector set, and fh1 to f_{hk} represents the nonoverlapping patches feature vectors.

Using Equations 1 and 7, the second stage feature vector set of the overlapping patches using ResNet18 features is formed as shown in Equation 11:

$$F_{rn} = [f_{r1}, f_{r2}, f_{r3}, \ldots . f_{rk}] \quad (11)$$

where F_{rn} represents the input image nonoverlapping patches feature vector set, and f_{r1} to f_{rk} represents the non-overlapping patches feature vectors. The similarity score, which uses the cosine similarity between the dataset feature vector F_{DH} using HOG and the l^{th} overlapped patch feature vector $F_h(l)$ of the given SSS image using Equation 8, is expressed by Equation 12:

$$Sh(k,l) = cosinesim(FDH(k), Fh(l)) \quad (12)$$

where FDH(K) is the K^{th} image feature vector in a HOG dataset. Sh(k,l) represents the similarity score between a dataset's Kth image feature vector and the l^{th} overlapped patch feature vector.

The similarity score, which uses the cosine similarity between the dataset feature vector F_{DR} using ResNet18 and the lth overlapped patch feature vector $F_r(l)$ of the given SSS image Equation 8, is expressed by Equation 13:

$$Sr(k,l) = cosinesim(FDR(k), Fr(l)) \quad (13)$$

where FDR(K) is the Kth image feature vector in a dataset using ResNet18. Sr(k,l) represents the similarity score between the Kth image feature vector in a dataset and the l^{th} overlapped patch feature vector,

The set of maximum scores Smax using Equation 12 and Equation 13 is obtained as in Equation 14;

$$Smax = (Sh \geq T, Sr \geq T,) \quad (14)$$

where T=threshold score.

Further, if maximum-scored patches exist, the object location is obtained using overlap maximization. Overlap maximization is applied for every maximum-scored patch using equation (9). The patch with maximum IOU is considered an object patch. The patches with maximum scores after overlap maximization are merged to locate the object.

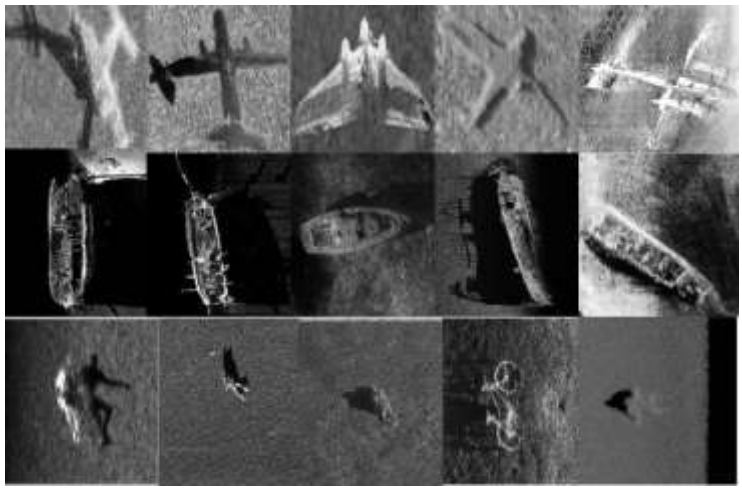

Figure 5: The collected dataset from various sources.

3. EXPERIMENTAL RESULTS

This section provides a comprehensive assessment of the performance of the two-stage object detection approach utilizing the collected dataset. Since no public datasets are available for the side scan sonar dataset. The collected data set consists of 1123 images from various sources, including AS-dataset [33], QD dataset [34], SO-KLSG dataset [35], and public sharing websites. For performance evaluation, each SSS image is annotated with a bounding box using a label box [36]. Figure 4 shows the collected dataset. The qualitative and quantitative results of the two-stage object detection were evaluated using MATLAB software on a system equipped with an Intel® Core™ Ultra 9 processor, NVIDIA® GeForce RTX™ 4070 8GB, and 32 GB LPDDR5x RAM. A cosine similarity score over the threshold extracts the mostly detected patches. The threshold score is 70, which extracts the most similarity patches.

Figure 5 displays the qualitative results of the proposed object detection in SSS images on the collected database. In each SSS image, the proposed object detection technique successfully detects the target object in a completed scene with a ground truth overlap more significant than 0.5. Furthermore, the object detection method may be enhanced with multi-feature extraction to increase detection performance and minimize computational complexity.

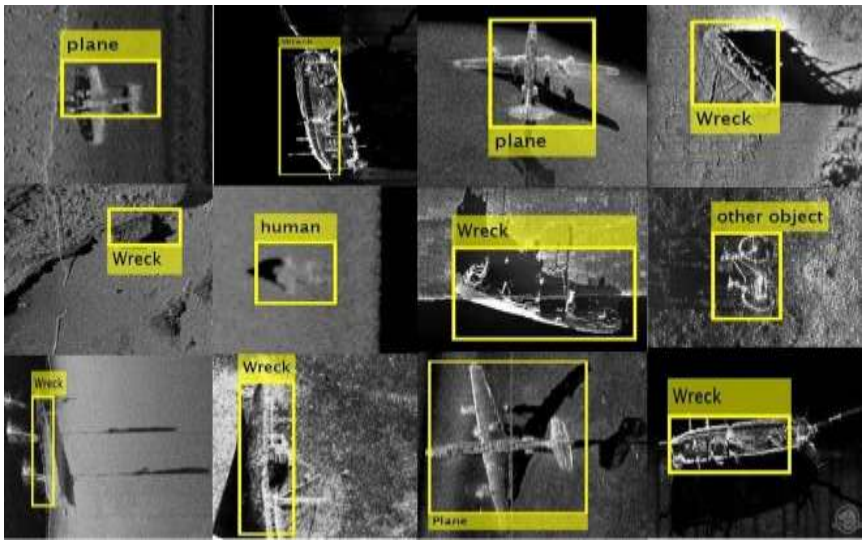

Figure 6: Proposed object detection results collected dataset.

The performance of the proposed object detection approach with the estimate was quantitatively assessed using a confusion matrix. Precision, recall, average precision, and detection speed are the performance measures used for qualitative evaluation [50]. The performance of two-stage object detection results in Tables 1, 2, and 3 demonstrate that compared with HOG features and ResNet18 features, the proposed technique results in a robust precision score of 77.47%, accuracy of 67.59%, F1 score of 0.8. The proposed approach meets the most recent visual object detection standards regarding overall performance metrics.

Table 1: Performance of proposed object detection technique using first-stage HOG features.

Performance measure	HOG features				
	Aircraft	Ship	Human Body	others	Overall
Number of SSS images	467	398	203	55	1123
True Positive	251	214	108	29	602
True Negative	102	87	42	12	243
False Positive	74	61	32	9	176
False Negative	41	37	19	5	102
Precision	71.1	71.1	72	70.7	71.24
Accuracy	62.4	62.9	63.2	61.8	62.69
F-1 Score	0.744	0.051424	0.74	0.734	0.74

Table 2: Performance of proposed object detection technique using Second stage Resnet18 features.

Performance measure	ResNet18				
	Aircraft	Ship	Human Body	others	Overall
Number of SSS images	467	398	203	55	1123
True Positive	257	219	112	31	619
True Negative	97	83	43	11	234
False Positive	69	58	30	8	165
False Negative	44	37	19	5	105
Precision	72.6	72.5	72.2	73.8	72.57
Accuracy	64.5	64.4	64.2	65.4	64.47
F-1 Score	0.755	0.756	0.754	0.765	0.76

Table 3: Performance of proposed two stage object detection technique.

Performance measure	HOG+ResNet18 Features				
	Aircraft	Ship	Human Body	others	Overall
Number of SSS images	467	398	203	55	1123
True Positive	301	257	131	36	725
True Negative	88	75	38	10	211
False Positive	64	54	27	8	153
False Negative	14	12	6	2	34
Precision	77.3	77.4	77.5	78.26	77.46
Accuracy	67.4	67.6	67.8	67.86	67.59
F-1 Score	0.798	0.799	0.8	0.8	0.8

Table 4: The proposed technique performance comparison with state-of-art techniques.

Method	Precision				Average Precision
	Aircraft	Ship	Human Body	Other objects	
Mask R-CNN [37]	63.9	71.2	61.4	58.7	63.80
MS R-CNN [38]	65.7	73.9	61.5	62.21	65.83
SOLO [39]	69.8	70.3	66.2	63.4	67.43
YOLOACT [40]	71.3	72.8	63.7	63.8	67.90
LW R-CNN [41]	74.3	74.3	68.2	66.7	70.88
Proposed	77.3	75.4	68.7	69.26	72.67

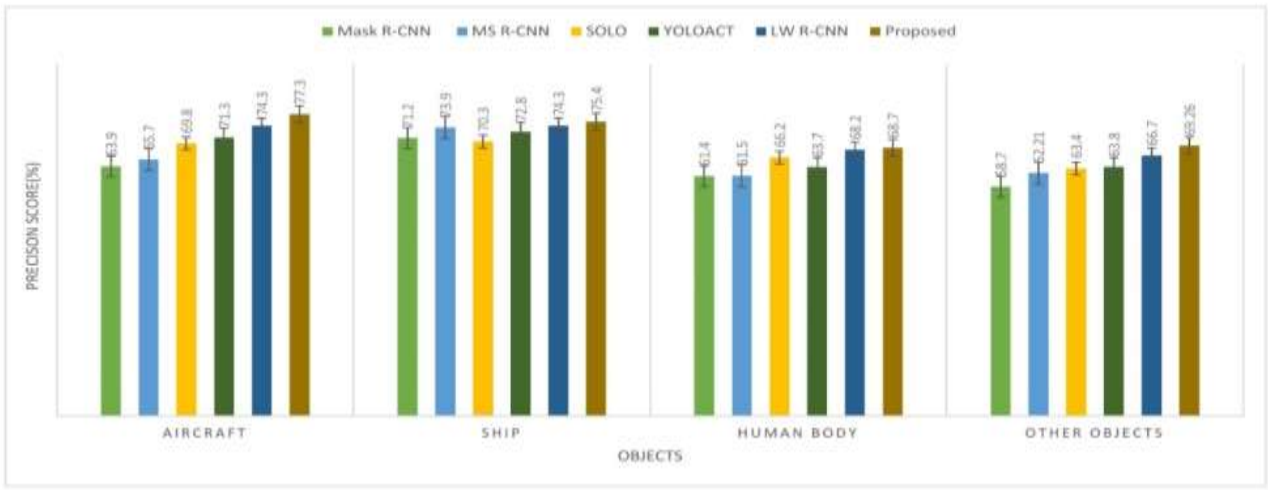

Figure 7: The proposed technique performance compared with the state-of-the-art techniques on collected datasets.

Various object detection techniques, including Mask R-CNN [37], MS R-CNN [38], SOLO [39], YOLOACT [40], and LW R-CNN [41], were used to evaluate and compare the overall performance results. Table 4 compares the proposed technique's performance with other state-of-the-art methods.

The proposed technique effectively identifies the target object in the SSS image with improved precision compared to the object detection technique Mask R-CNN [37], MS R-CNN [38], SOLO [39], YOLOACT [40], and LW R-CNN [41]. The average precision of each detection technique on the collected dataset is evaluated using a confusion matrix.

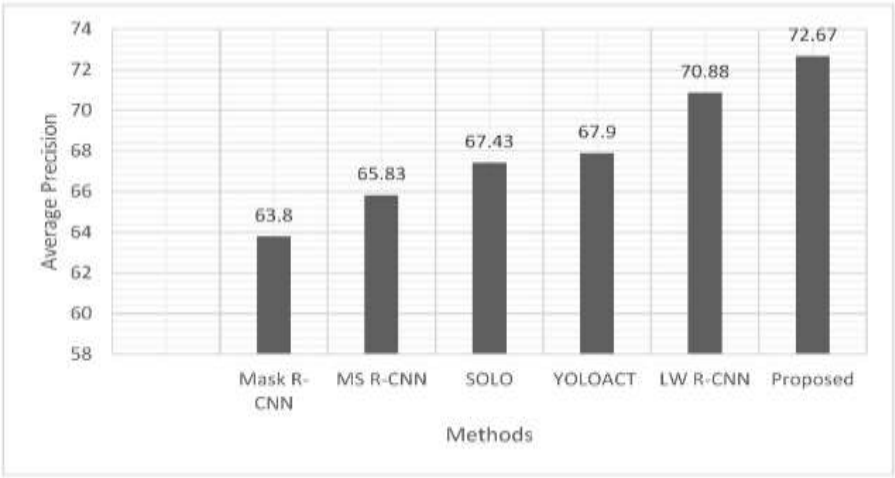

Figure 8: The overall performance of the proposed technique compared with the state-of-the-art methods.

The precision score of the proposed technique is improved in the detection of aircraft objects, with 77.3% compared with techniques precision scoresMask R-CNN of 63.9%, MS R-CNN of 65.7%, SOLO of 69.8%, YOLOACT of 71.3%, and LW R-CNN of 74.3%. Figure 6 demonstrates the overall performance of the proposed technique in various object detection scenarios compared to state-of-the-art techniques. The proposed object detection achieved a precision score of 75.4% for Shipwreck

detection, surpassing the 74.3% achieved by LW R-CNN. Similarly, the proposed technique achieved a precision score of 68.7% for human body detection, outperforming the 68.2% achieved by LW R-CNN. These results demonstrate the robustness and effectiveness of the proposed technique. Figure 7 illustrates the proposed technique's overall performance compared to the state-of-the-art methods. The average precision score of the proposed technique of 72,67% shows robust performance compared with techniques average precision scores Mask R-CNN of 63.8%, MS R-CNN of 65.83%, SOLO of 67.43%, YOLOACT of 67.90%, and LW R-CNN of 70.88%.

Table 5: The proposed object detection speed performance with state-of-the art-techniques,

Method	Mask R-CNN	MS R-CNN	SOLO	YOLOACT	LW R-CNN	Proposed
Speed (Frames/second)	18.5	24.4	41.7	39.6	30.3	29.4

To assess object detection speed, we measure the frame per second on the collected dataset. Table 5 displays the speed comparison of the proposed object detection method with state-of-the-art techniques. SOLO has an object detection speed of 41.7 frames/second. The proposed object detection speed of 29.4 frames per second with multi-stage feature extraction and matching demonstrates its resilience compared to the speeds of the state-of-the-art techniques. Furthermore, a faster feature extraction technique may enhance the proposed object detection approach's performance and detection speed.

4. CONCLUSIONS

This paper presents a two-stage object detection framework with HOG and Pre-Trained ResNet features. The process consists of four steps: patch extraction, feature extraction, similarity estimation, and overlap maximization. The framework proposed in this study can detect objects through patch feature matching. Feature matching is performed using HOG features in the initial stage, while pre-trained ResNet features are employed in the subsequent stage. Feature matching is conducted using cosine similarity in both stages. The detected object patch is determined by selecting the maximum overlapping patch from the highest-scored patches in two stages. The experimental results indicate that the proposed technique has achieved a precision score of 72.67% while maintaining a detection speed of 29.4 frames per second. This demonstrates the effectiveness and robustness of the object detection technique being proposed. However, the experiment results indicate that multiple objects in the scene impact object detection. Hence, it is imperative to consider these aspects in future research. The suggested algorithm for object detection can be adapted to identify multiple objects in the seabed by utilizing distinctive features.

REFERENCES

S. Maznev, S. Ogorodov, A. Baranskaya, A. Vergun, V. Arkhipov, and P. Bukharitsin, “Ice-Gouging Topography of the Exposed Aral Sea Bed,” Remote Sensing., vol. 11, no. 2, pp. 113–121, Jan. 2019, doi: 10.3390/rs11020113.

A. Grzadziel, “Results from Developments in the Use of a Scanning Sonar to Support Diving Operations from a Rescue Ship,” Remote Sensing., vol. 12, no. 4, pp. 693–705, Feb. 2020, doi: 10.3390/rs12040693.

L. Character, A. Ortiz JR, T. Beach, and S. Luzzadder-Beach, “Archaeologic Machine Learning for Shipwreck Detection Using Lidar and Sonar,” Remote Sensing., vol. 13, no. 9, pp. 1759–1768, Apr. 2021, doi: 10.3390/rs13091759.

Steiniger, Yannik, Dieter Kraus and Tobias Meisen. “Survey on deep learning based computer vision for sonar imagery.” Eng. Appl. Artif. Intell. 114 (2022): 105157.

Langner, F., Christian Knauer, Wolfgang Jans and Alfons Ebert. “Side scan sonar image resolution and automatic object detection, classification and identification.” OCEANS 2009-EUROPE (2009): 1-8.

Jayanthi, N.; Indu, S. Comparison of Image Matching Techniques. Int. J. Latest Trends Eng. Technol. 2016, 7, 396–401.

Alam, M.; Morshidi, M.; Gunawan, T.; Olanrewaju, R. A Comparative Analysis of Feature Extraction Algorithms for Augmented Reality Applications. In Proceedings of the 2021 IEEE 7th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Bandung, Indonesia, 23–25 August 2021; pp. 59–63

Ma, J.; Sun, Q. Image Recognition Method based on Artificial Intelligence Technology. In Proceedings of the 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 27–29 May 2022; pp. 971–974.

Fakiris, E.; Blondel, P.; Papatheodorou, G.; Christodoulou, D.; Dimas, X.; Georgiou, N.; Kordella, S.; Dimitriadis,

C.; Rzhanov, Y.; Geraga, M.; et al. Multi-Frequency, Multi-Sonar Mapping of Shallow Habitats—Efficacy and Management Implications in the National Marine Park of Zakynthos, Greece. Remote Sens. 2019, 11, 461.

Yang, Dianyu, Jingfeng Yu, Can Wang, Chensheng Cheng, Guang Pan, Xin Wen, and Feihu Zhang. 2024. "Side-Scan Sonar Image Matching Method Based on Topology Representation," Journal of Marine Science and Engineering 12, no. 5: 782. https://doi.org/10.3390/jmse12050782

C. Dong, L. Guo, K. Hu, J. Yin and X. Sheng, "Side-scan Sonar Image Rough Recognition and Feature Matching Based on CNN and SIFT," 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 2019, pp. 1-5, doi: 10.1109/ICSIDP47821.2019.9172987.

K. Hatanaka and M. Wada, "An algorithm based on the wavelet transform for the classification of seabed textures," OCEANS 2010 MTS/IEEE SEATTLE, Seattle, WA, USA, 2010, pp. 1-6, doi: 10.1109/OCEANS.2010.5664485. keywords: {Classification algorithms;Wavelet coefficients ;Discrete wavelet transforms; Algorithm design and analysis;Waveletanalysis;Software;Reflection},

Ma, Quanhong, Shaohua Jin, Gang Bian, and Yang Cui. 2024. "Multi-Scale Marine Object Detection in Side-Scan Sonar Images Based on BES-YOLO" Sensors 24, no. 14: 4428. https://doi.org/10.3390/s24144428

W. Tao, J. Zhao, J. Liu and H. Zhang, "Study on the Side-scan Sonar Image Matching Navigation Based on SURF," 2010 International Conference on Electrical and Control Engineering, Wuhan, China, 2010, pp. 2181-2184, doi: 10.1109/iCECE.2010.537.

M. S, A. L, A. S. L, A. S. L and S. Rangaswamy, "Underwater Mine Detection Using Histogram of oriented gradients and Canny Edge Detector," 2021 International Carnahan Conference on Security Technology (ICCST), Hatfield, United Kingdom, 2021, pp. 1-6, doi: 10.1109/ICCST49569.2021.9717404.

G. Annalakshmi, S. S. Murugan and K. Ramasundaram, "Side Scan Sonar Images Based Ocean Bottom Sediment Classification," 2019 International Symposium on Ocean Technology (SYMPOL), Ernakulam, India, 2019, pp. 138-144, doi: 10.1109/SYMPOL48207.2019.9005290.

K. Denos, M. Ravaut, A. Fagette and H. -S. Lim, "Deep learning applied to underwater mine warfare," OCEANS 2017 - Aberdeen, Aberdeen, UK, 2017, pp. 1-7, doi: 10.1109/OCEANSE.2017.8084910.

McKay, J., Gerg, I., Monga, V., Raj, R.G., 2017. What's mine is yours: Pretrained CNNs for limited training sonar ATR. In: OCEANS 2017 MTS/IEEE Anchorage. IEEE, pp.1–7.

Einsidler, D., Dhanak, M., Beaujean, P.-P., 2018. A deep learning approach to target recognition in side-scan sonar imagery. In: OCEANS 2018 MTS/IEEE Charleston. IEEE, pp. 1–4. http://dx.doi.org/10.1109/OCEANS.2018.8604879.

Xu, Yichao; Wang, Xingmei; Wang, Kunhua; Shi, Jiahao; Sun, Wei: 'Underwater sonar image classification using generative adversarial network and convolutional neural network', IET Image Processing, 2020, 14, (12), p. 2819-2825, DOI: 10.1049/iet-ipr.2019.1735

Feldens, P., Darr, A., Feldens, A., Tauber, F., 2019. Detection of boulders in side scan sonar mosaics by a neural network. Geosciences 9 (4), 159. http://dx.doi.org/10. 3390/geosciences9040159.

Feldens, P., 2020. Super-resolution by deep learning improves boulder detection in side scan sonar backscatter mosaics. Remote Sens. 12 (14), 2284. http://dx.doi.org/10. 3390/rs12142284.

Jiang, L., Cai, T., Ma, Q., Xu, F., Wang, S., 2020. Active object detection in sonar images. IEEE Access 8, 102540–102553. http://dx.doi.org/10.1109/ACCESS.2020. 2999341.

Yu, Y., Zhao, J., Gong, Q., Huang, C., Zheng, G., Ma, J., 2021. Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5. Remote Sens. 13 (18), 3555. http://dx.doi.org/10.3390/rs13183555.

Berthomier, T., Williams, D.P., d'Alès de Corbet, B., Dugelay, S., 2020. Exploiting auxiliary information for improved underwater target classification with convolutional neural networks. In: Global Oceans 2020: Singapore — U.S. Gulf Coast. IEEE, pp. 1–10. http://dx.doi.org/10.1109/IEEECONF38699.2020.9389138.

Le, H.T., Phung, S.L., Chapple, P.B., Bouzerdoum, A., Ritz, C.H., Tran, L.C., 2020. Deep Gabor neural network for automatic detection of mine-like objects in sonar imagery. IEEE Access 8, 94126–94139. http://dx.doi.org/10.1109/ACCESS.2020.2995390.

Topple, J.M., Fawcett, J.A., 2021. MiNet: Efficient deep learning automatic target recognition for small autonomous vehicles. IEEE Geosci. Remote Sens. Lett. 18 (6), 1014–1018. http://dx.doi.org/10.1109/LGRS.2020.2993652.

Steiniger, Y., Stoppe, J., Meisen, T., Kraus, D., 2020. Dealing with highly unbalanced side scan sonar image

datasets for deep learning classification tasks. In: Global Oceans 2020: Singapore — U.S. Gulf Coast. IEEE, pp. 1–7. http://dx.doi.org/10. 1109/IEEECONF38699.2020.9389373.

N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings of the Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 886–893, 2005.

Aklak, A.F., Vadamala, P.R. Visual object tracking via adaptive deep feature matching and overlap maximization. Pattern Anal Applic 26, 889–906 (2023). https://doi.org/10.1007/s10044-023-01157-9

[Vadamala, P.R., Aklak, A.F. Discriminative appearance model with patch spatial adjustment for visual object tracking. Soft Comput 27, 9787–9800 (2023). https://doi.org/10.1007/s00500-023-07820-x

Vadamala, Purandhar Reddy, and Annis Fathima Aklak. "Adaptive Patch Feature Matching and Scale Estimation for Visual Object Tracking." Journal of Electronic Imaging, June 2019. SPIE - International Society for Optical Engineering, https://doi.org/10.1117 /1.jei.28.3.033037.

L. Jiang, T. Cai, Q. Ma, F. Xu, and S. Wang, "Active Object Detection in Sonar Images," IEEE Access., vol. 8, no. 12, pp. 102540–102553, May. 2020, doi: 10.1109/ACCESS.2020.2999341.

Z. Wang, S. Zhang, W. Huang, J. Guo, and L. Zeng, "Sonar Image Target Detection Based on Adaptive Global Feature Enhancement Network," IEEE Sensors Journal., vol. 22, no. 2, pp. 1509–1530, Jan. 2022, doi: 10.1109/JSEN.2021.3131645

G. Huo, Z. Wu, and J. Li, "Underwater Object Classification in Sidescan Sonar Images Using Deep Transfer Learning and Semisynthetic Training Data," IEEE Access., vol. 8, no. 12, pp. 47407–47418, Mar. 2020, doi: 10.1109/ACCESS.2020.2978880.

B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," International Journal of Computer Vision., vol. 77, no. 3, pp. 157–173, May. 2008, doi: 10.1007/s11263-007-0090-8.

K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in Proc. IEEE Conf. Comp. Vis. Patt. Recogn (ICCV), Oct. 2017, pp. 2980–2988. doi: 10.1109/ICCV.2017.322.

Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask Scoring R-CNN," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 6402–6411. doi: 10.1109/CVPR.2019.00657.

X. Wang, T. Kong, C. Shen, Y. Jiang, and L. Li, "SOLO: Segmenting Objects by Locations," in Proc. Eur: Conf. Comput. Vis. (ECCV), Dec. 2020, pp. 649–665, doi: 10.1007/978-3-030-58523-5 38.

D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-Time Instance Segmentation," in Proc. IEEE Conf. Comp. Vis. Patt. Recogn (ICCV), Oct. 2019, pp. 9156–9165. doi: 10.1109/ICCV.2019.00925.

Z. Fan, W. Xia, X. Liu, and H. Li, "Detection and Segmentation of Underwater Objects from Forward-Looking Sonar based on a Modified Mask RCNN," Signal, Image and Video Processing., vol. 15, no. 6, pp. 1135–1143, Sep. 2021, doi: 10.1007/s11760-020-01841-x.