The convolution neural network based agent vehicle detection using forward-looking sonar image

Juhwan Kim, Hyeonwoo Cho, Juhyun Pyo, Byeongjin Kim, Son-Cheol Yu
Department of Creative IT Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, Gyeongbuk, South Korea
robot_juwan@postech.ac.kr

***Abstract*— We propose the convolution neural network (CNN) based underwater object recognition and detection solution using forward-looking sonar image for localization of agent vehicle (small ROV). We designed the multiple underwater vehicle system with the autonomous underwater vehicle (AUV) equipped with the forward-looking sonar and the agent vehicle linked to the AUV through a tether cable. For localization of agent vehicle, we use the forward-looking sonar images. The CNN based object recognition algorithm trained the agent vehicle's sonar image and detected it in the overall sonar images. We found that the CNN algorithm successfully recognized the agent vehicle in the sonar image. The goal of our research is to propose a solution to apply the CNN based recognition algorithm to the underwater robotics. Finally, it shows the elevated recognition rate in the underwater and we can get the agent vehicle's localization data.**

Keywords—Object recognition; Object detection; Forward-looking sonar; Machine learning; Neural network; Agent vehicle

I. INTRODUCTION

Underwater agent vehicle system is devised to solve the problem of performing underwater tasks utilizing only an autonomous underwater vehicle (AUV). It has multiple underwater devices which works as navigation sensors. In addition, it requires strong thrusters for exploring wide underwater area. To install equipment, the body size becomes larger, and also its position control accuracy becomes worse due to the high thruster power. Therefore, the AUV is not suitable for precise underwater tasks such as underwater docking and close-range investigation.

The AUV's manipulation method is important to find out some materials and conduct detailed works. Several researchers equipped the rigid manipulator on the AUVs [6]. With the underwater actuating hand, it can grip the objects or cut any obstacles. Its manipulation difficulty is somewhat low. However, its drifting body effects the end-effector of hand and heavy body changes the vehicle's orientation during the manipulation.

For these reasons, we designed the agent vehicle system that is a small-sized remotely operated vehicle (ROV) [8]. It is linked to the AUV through a tether cable [1]. Unlike the AUV, the agent vehicle has small thrusters for its accurate position control. Moreover, the agent vehicle has limited number of underwater navigation sensor to make its size small. Therefore, its position control requires the support of the AUV. One of the scenarios for agent position control is using the sonar vision system of the AUV [10]. It can detect and track the motion of agent vehicle by using its vision system. After that, it sends the position control commands to the agent vehicle through the tether cable to locate it to desired positions.

In this study, we concentrate on the agent vehicle's recognition on the vision system. We propose the convolution neural network based agent vehicle detection algorithm using forward-looking sonar image. Although the optical vision is one of candidates for the purpose, the optical vision is suffer from the irregularity of underwater lights and the turbid water condition. These environments dramatically reduce the visibility and detection success rate. On the other hand, the sonar image is based on the acoustic signal, so we don't matter the problem of lights and turbidity [2]. However, the sonar image-based real-

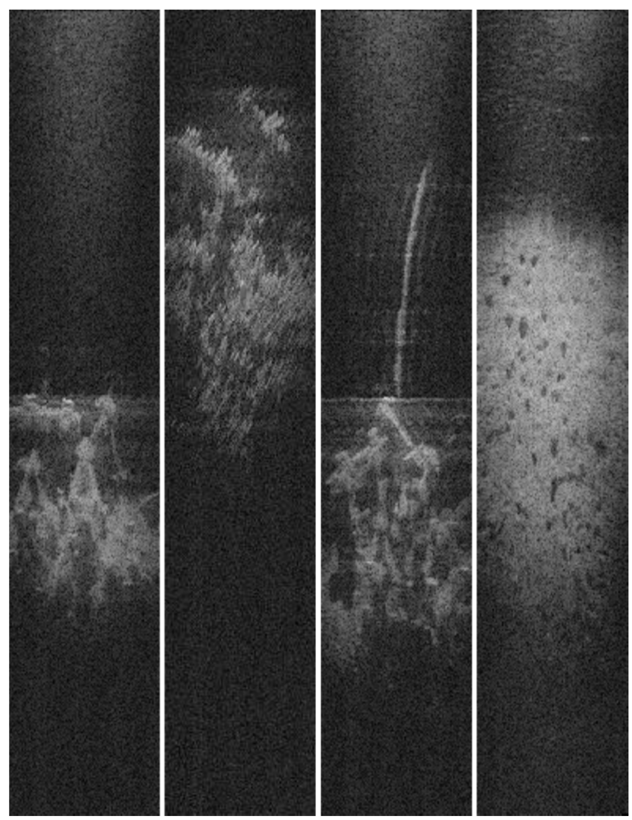


Fig. 1. Sample Images of high resolution forward looking sonar. They were taken by the hovering type AUV 'Cyclops' at the Jangil Bay, South Korea [7].

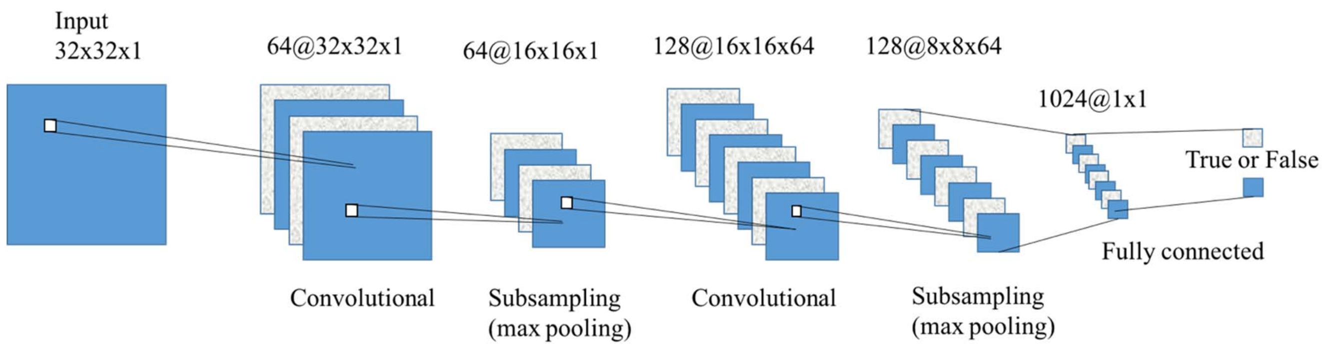


Fig. 2. Structure of convolution neural network in the study.

time detection and tracking with high reliability are not accomplished by using the convent ional image processing algorithms for optical vision such as template matching and feature tracking, because the sonar images have low resolution and high noise.

II. Proposed Method

A. Convolution neural network (CNN) based object recognition

To recognize the object on sonar image is difficult, because it has low resolution and shows the mixed three parts of image: shadow, background and highlight (Fig. 1) [9]. To solve the problem, we developed the supervised machine learning algorithm CNN for object recognition in sonar images. It can have the invariances which is irrelevant to topology changes by the local connectivity and shared weights [5]. The agent vehicle can have enormous sonar images, because its shape is fixed and we operated it at several missions. In the end, it can defectively train CNN, and have high recognition rates in low resolution images.

We trained the sonar image of a small agent vehicle to CNN, and attempted the object recognition. The properly cropped images are labeled as matching region. The regions which in snot properly fitted are labeled as non-matching region. Those labeled images are pushed in input nodes and output nodes are the value of label. The neural network structure is Le cun's basic model that consists of convolution and max-pooling layers (Fig. 2) [3]. Total ten thousands of sonar images are learned and those images include the real sea backgrounds.

B. Overall operation strategy

The forward-looking sonar's background variation makes difficult to find particular objects. And many different image shapes of target object make the computer vision algorithm not to find correct feature. Therefore, our proposed method uses machine learning algorithm to detect agent vehicle in the changing environment. Eventually, we need plenty of image data to train our algorithm.

CNN based agent vehicle detection strategy has four steps (Fig. 3).

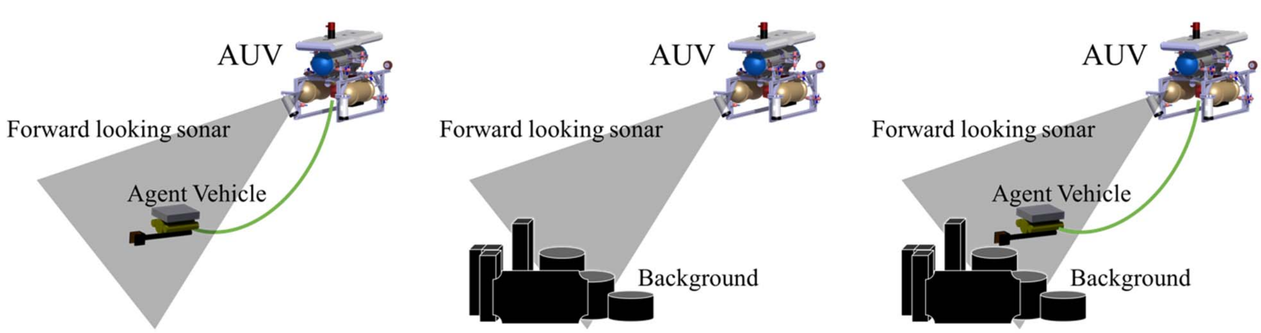
Step 1) *Take sonar images of agent vehicle and make data-set as shown in Fig. 3, (1).*

Step 2) *Explorer the place and take sonar images in advance and make data-set as shown in Fig. 3, (2).*

Step 3) *Train the data-set to the designed neural network.*

Step 4) *Operate the mission in the place and do object detection of the agent vehicle as shown in Fig. 3, (3).*

Before the agent vehicle operates in the particular region, we gathered the background sonar images in advance. Once the neural network trained the background images, it does not detect the part of background as target object.



(1) Agent image taking **(2)** Environment image taking **(3)** Agent control for the manipulation

Fig. 3. Scenario of the proposed manipulation system.

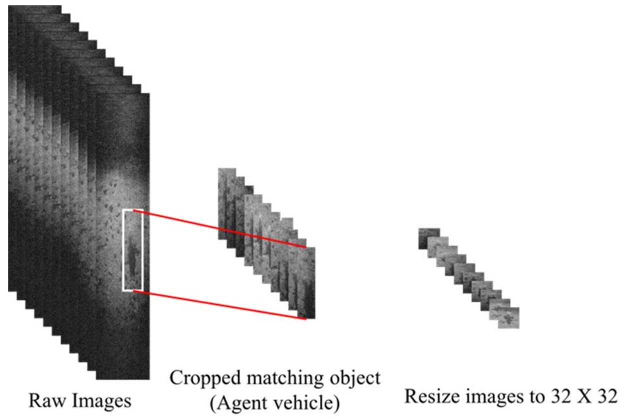


Fig. 4. Generation of the target images from the raw image. They are the matching data set, the agent vehicle image crops.

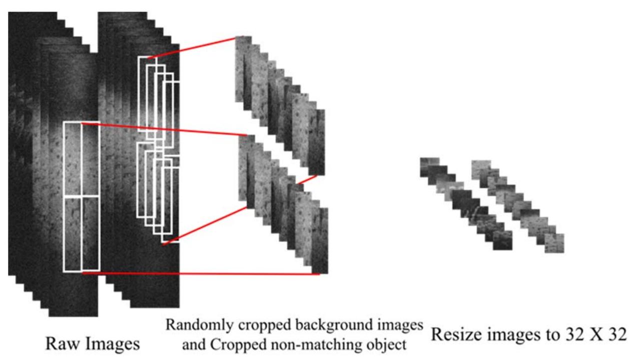


Fig. 5. Generation of background images, They are the non-matching object, the background crops.

C. Data-sets of sonar images

The data-set is the input data for neural network trainning. The amount of trainning data is somewhat large, because roughly over ten thousands of images are stacked in the data-set. In addition, making the strategic data-set increases the neural network's accuracy and lowers the error rates. The original sonar image size is 96x512 pixels. We cropped the images as 32x128 pixels and resize to 32x32 pixels.

We gathered the two kinds of data-sets.

i. Cropped matching object (Agent vehicle)

ii. Randomly cropped background images and Cropped non-matching object

Firstly, matching object data-set is the clearly cropped agent vehicle images (Fig. 4). The target object images are changed by its positions, scales, shadows and movements. For precise recognition of its various shapes, a plenty of matching object images are needed. We used the usual image tools to crop the images manually. In the 2,500 forward-looking sonar images of agent vehicle, we cropped the 1,156 images and labeled them as true.

Secondly, we made randomly cropped background images and non-matching object data-set (Fig. 5). The randomly cropped background images increase the durability of detection when the unexpected objects emerge. Thus, we used the ten different regions of background sonar images to make data-set. One original background sonar image can generate the 45 cropped images. The other important data-set is non-matching object. It is inaccurately cropped agent vehicle image in purpose. This data-set can increase the precise object recognition. Total second data-sets include 18,956 images and labeled them as false.

All data-sets randomly rearrange their orders and divided into training-set and test-set. The test-set is to test network's accuracy after training process. We set the portion of test-set as

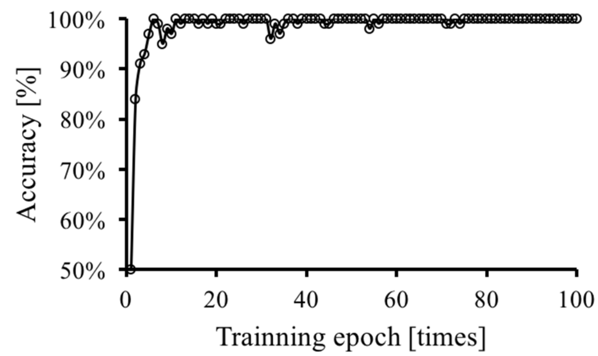


Fig. 6. The training-set test result of CNN Training, Accuracy.

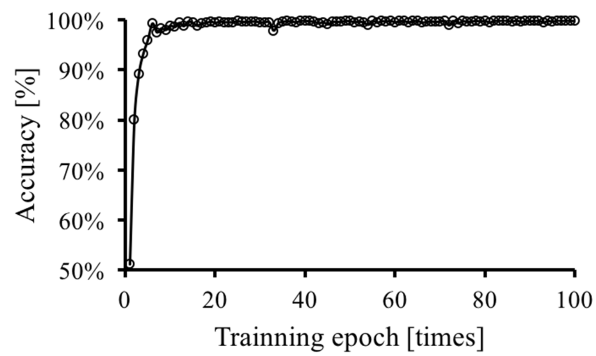


Fig. 7. The test-set test result of CNN Training, Accuracy.

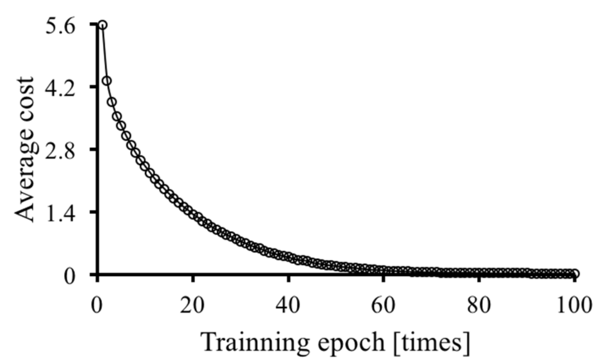


Fig. 8. The CNN average training cost value.

12.5 %. In the all data-set of 20,112, the 2,514 are test-set and 17,598 are training-set.

D. Object detection and localization

We used simple object detection algorithm to find location of target object in the sonar image. The sliding reconition region is shifted in two pixels and scores the each region to find agent vehicle. When it finds the correct object, the rectangular shape overwraps the original sonar image on the founded agent vehicle location. After processing all portion of images, we can get the x, y average coordinate values of every true labeled regions. The transformation expression can convert these values to relational position values.

III. Recognition Test

A. CNN based object recognition

We trained the CNN's weights and biases by inserting total 20,112 32x32 data-set and label data. It takes an hour to train repetition of 400 epochs. At the sixth epoch, it shows the over 99% trainning accuracy (Fig. 6). It means our classifyer can recognize training data as correct answer over 99%. In each epoch, the CNN also calculates the test-set accuracy. It also converged to over 99% after sixth epoch (Fig. 7). The average of training cost value lowers during the training epochs (Fig. 8).

The proposed CNN algorithm has higher recognition accuracy than general template-matching algorithm (Fig. 9). The first image shows the agent vehicle and the unusual object on the top of the image. The second image is general underwater constructions. Both algorithms catch the agent vehicle. However, the template-matching catches the odd object and underwater structures as agent vehicle. The proposed CNN can catch only the agent vehicle.

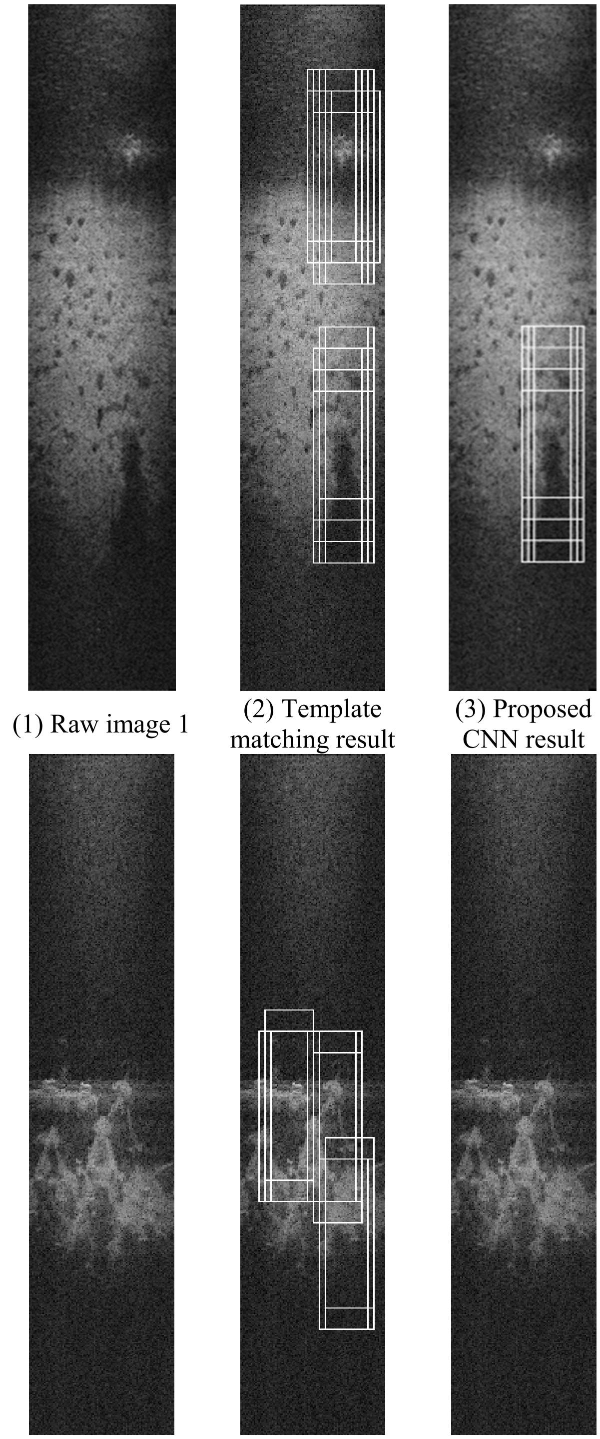
B. Object detection and localization

We conducted real-sea experiment to find agent vehicle in the forward-looking sonar images. We processed two hundreds of sonar images to find location of agent vehicle. The agent vehicle detection was successfully conducted (Fig. 10). Among the 221 images, only three images have a detection error.

The agent vehicle's trajectories were recorded (Fig. 11). The different shapes of trajectories show separated routes of sequential images. The agent vehicle was disappeared in the images between the routes.

IV. Discussion

Our proposed solution successfully detected the target object. However, the sliding window recognition is quite slow to process in real-time. Reducing CNN node's depth size or input image size would be helpful to raise speed of recognizing image. However, it has possibility to lower the accuracy. Most effective method of improving speed is applying state of the art object detection algorithm based on CNN. Regions with convolutional neural network (R-CNN) algorithms or "You only look once (YOLO)" algorithm are typical examples. After localization of using sonar image in real-time, we can precisely control the agent vehicle by the feedback of AUV's real-time image processing.



(1) Raw image 1 (2) Template matching result (3) Proposed CNN result

(4) Raw image 2 (5) Template matching result (6) Template matching result

Fig. 9. The result of object detection by template matching algorithm and proposed CNN algorithm.

The other problem of our research is that we did not verify the non-shadow agent vehicle. If our solution is hard to recognize that object, we have to gather the non-shadow sonar images and train the data.

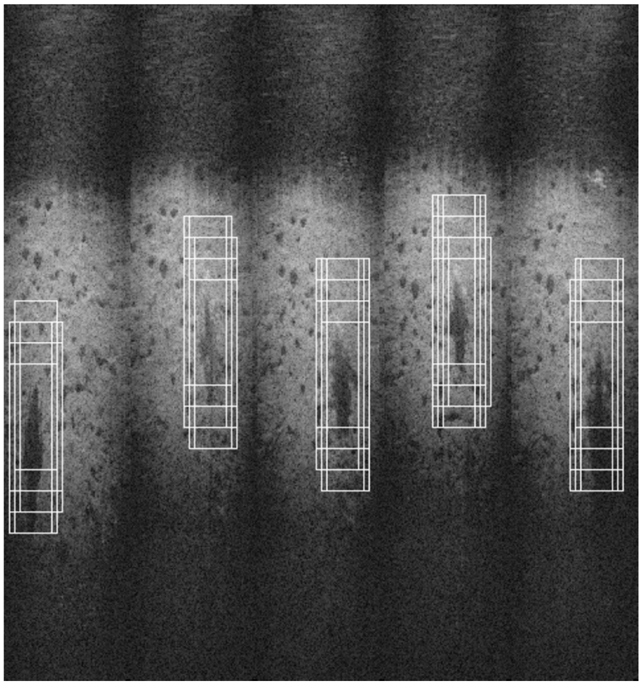


Fig. 10. The result of agent vehicle detection in the sonar images.

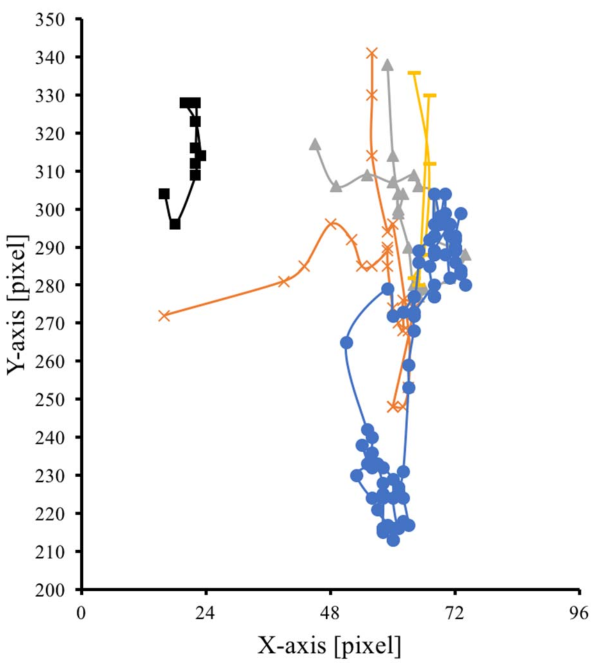


Fig. 11. The agent vehicle's trajectories on the sonar images.

The oceanic background has limited environments and fewer obstacles around the floating object. Therefore, if the large amount of underwater data-set is gathered, the machine learning algorithms can demonstrate the huge power in the underwater sonar image processing academic world.

V. CONCLUSION

This study verified the solution of processing forward-looking sonar image that is based on CNN and background data-set strategy. We designed the CNN and object detection algorithm. Then, we conducted it for localization of agent vehicle. We found that CNN algorithm is much effective to process forward-looking sonar images and we can detect the agent vehicle. Finally, it shows applying machine learning algorithms on processing sonar image is much more useful.

VI. ACKNOWLEGEMENT

This research was supported by the Office of Naval Research Global, US Navy (Grant No. N62909-14-1-N290)and the project titled "Gyeongbuk Sea Grant Program", funded by the Ministry of Oceans and Fisheries, Korea, Ministry of Science, ICT and Future Planning, Korea, under the "ICT Consilience Creative Program" (IITP-R0346-16-1007) supervised by the Institute for Information & communications Technology Promotion.

REFERENCES

[1] Son-Cheol Yu, "A Preliminary Test On Agent-based Docking System For Autonomous Underwater Vehicles", International Journal Society of Offshore and Polar Engineers, vol. 19, no. 1, 2009, pp. 52-59.

[2] H. Cho, J. Gu, H. Joe, A. Asada, S. -C. Yu, "Acoustic beam profile-based rapid underwater object detection for an imaging sonar," J. of Marine Sci. and Tech., vol. 20, no. 1, pp. 180-197, 2015

[3] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a backpropagation neural network", in Advances in Neural Information Processing Systems 2, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, 1990, pp. 396-404.

[4] S. C. Yu, T. W. Kim, A. Asada, S. Weatherwax, B. Collins, and J. Yuh, "Development of High-Resolution Acoustic Camera based Real-Time Object Recognition System by using Autonomous Underwater Vehicles,", IEEE OCEANS 2006, September, 2006, pp. 1-6.

[5] H. Noh, P. H. Seo, B. Han, "Image Question Answering using Convolutional Neural Network with Dynamic Parameter Prediction", eprint arXiv:1511.05756, 11. 2015

[6] G. Marani, S. K. Choi, and J. Yuh, "Underwater autonomous manipulation for intervention missions AUVs," *Ocean Eng.*, vol. 36, no. 1, pp. 15–23, Jan. 2009.

[7] Juhyun Pyo, Han-Gil Joe, Jin-Hwan Kim et al "Development of hovering-type AUV "cyclops" for precision observation", Oceans-San Diego, pp.1-5, 2013.

[8] VideoRay, LLC The Global Leader In MicroROV Technology, http://www.videoray.com

[9] J. Pyo, J. H. Gu, H. Cho, S. C. Yu, "Agent based localization method for underwater docking.", 2014 Oceans-St. John's. IEEE, 2014.

[10] DIDSON sonar, Sound Metrics Corp, http://www.soundmetrics.com