Convolutional Neural Network-based Real-time ROV Detection Using Forward-looking Sonar Image

Juhwan Kim and Son-Cheol Yu
Dept. of Creative IT Engineering
Pohang University of Science and Technology
Pohang, South-Korea
robot_juhwan@postech.ac.kr

***Abstract*—Agent system is strategy to enhance the underwater manipulation. The conventional manipulation is generally robot arm-based configuration which has singular points. On the other hand, the agent system is an armless manipulation that the agent vehicle works as the end-effector. If the location of the agent can be measured, the end effector is able to be place to any position. To implement this system, the method of an agent vehicle localization is proposed. The method uses the sonar images of moving agent obtained by forward-looking sonar. To detect the location of the agent in the sonar images, the convolutional neural network is applied. We applied the state-of-art object-detection algorithm to the agent vehicle system. The fast object-detection algorithm based on neural network can fulfil the real-time detection and show the remarkable validity. It means the underwater robot can begin navigation under its feed-back. Through field experiment, we confirm the proposed method can detect and track the agent in the successive sonar images.**

Keywords—armless manipulation; agent vehicle; convolutional neural network; object detection; forward-looking sonar; sonar image processing.

I. Introduction

Exploring the deep sea is a fascinating field in that it has unknown environments. The autonomous underwater vehicles (AUVs) and the remotely operated vehicles (ROVs) were created for exploring deep seas where humans cannot explore directly, and this technology developed rapidly over the past few decades. These robots collect a variety of sensor data and it can make underwater maps or investigate the particular underwater resources. The underwater robots' control theory, navigation methods and sensing techniques are continuously studied by many researchers. The underwater manipulation issue is also rising field for some physical operations.

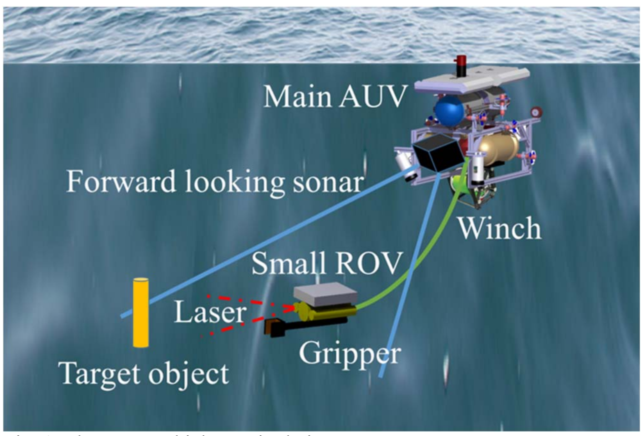

Fig. 1. The agent vehicle manipulation system.

Underwater robots have some ways of underwater manipulation. The robot arm method is the most widely used tools [1]. The AUVs or ROVs are equipped with a robotic arm and conduct some underwater physical works. It can precisely control the joint angle. However, it has limitation of moving, because joint structure which is called singular point takes heavy volume and weight [2]. For this reason, the armless manipulation method is developed.

The small ROV is a deployable and maneuverable as end-effector robot which is strap-on main AUV [3]. The ROV with long tether is free from heavy battery and conduct many missions without particular embedded-intelligences. The AUVs have extensive action radius related to freedom of tether, and they can use variety of instruments and sensors. The small ROV is beneficial to detail operations. It can perform armless underwater manipulation or execution of agent docking. It can play a part of manipulation hand to grip something or do precisely controlling works. We call its small ROV as "Agent vehicle" and its system as 'Agent vehicle system' (Fig. 1). For this system, we should detect agent's localization by its position sensor and main AUV's forward looking sonar. The precise location data can lead to accurate manipulation.

In this study, we proposed neural network based real-time object-detection for localization of the agent vehicle. The state-of-art and fast object-detection algorithm You Only Look Once (YOLO) shows the high-speed and exact detection [4]. We conducted this algorithm to our forward-looking sonar data. As a result, we found the possibility of using it as input of feed-back control.

II. Background

A. Forward-looking Sonar

The forward-looking sonar obtains the acoustic video images in real-time [5]. It has longer visual range than that of optical imaging, so it is a prospective solution for underwater object detection. However, the image quality of the forward-looking sonar is lower than that of optical images. Because of the

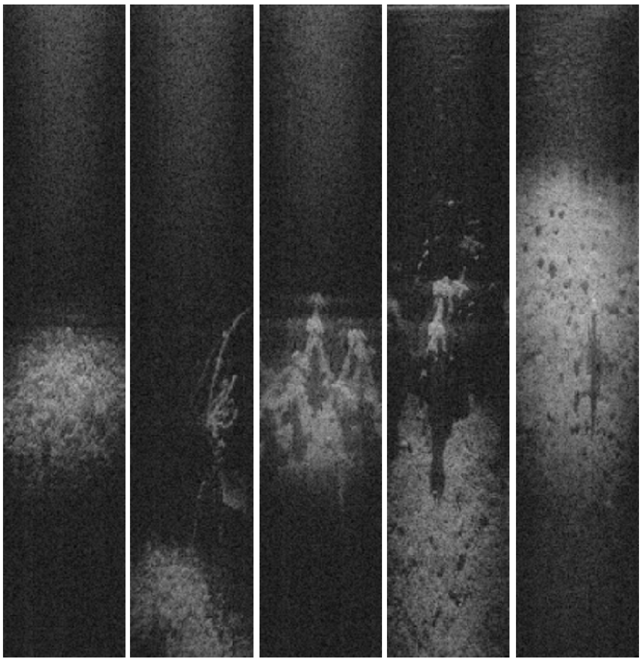

Fig. 2. The forward-looking sonar images. They were taken by AUV 'Cyclops' [17].

intrinsic limitation of acoustic beam, the forward-looking sonar delivers the low quality of acoustic images that only human's eyes can distinguish objects (Fig. 2). The images have low resolution and high noise. Moreover, the single image shows the three parts: shadow, background, highlights. Its image topology shows the different shape by taking view heights and angles [6]. These characteristics make it difficult to extract valid information by image processing techniques. Therefore, we cannot use the ordinary algorithms for processing sonar images.

B. Convolutional neural network based image classifier

Convolutional neural network is supervised machine learning algorithm that conduct convolution on neural network, and have locality and shared weights [7]. With increasing computing power of Graphics Processing Unit (GPU)'s parallel architecture, neural network modeling is the currently the hottest trend in image processing area [8]. The researchers can train and test the massive neural network model in short time [9]. Ordinary image processing used the feature matching to find interesting area. The low-level features are particular fixed shapes or post-processing algorithms. However, the convolutional neural network used high-level features that are determined by training [10]. The deeper with the heavy layers, the feature-levels increase and the model cannot be analyzed logically [8] [22]. In the end, the supervised machine learning generates the black-box function of accurate classifier.

C. Object-detection

The image classifier can only show the possibility of existence. Therefore, we cannot easily detect the location of target object in the image. Most of all, setting the proper ROI is important for object-detection. Several algorithms can find the ROI examples. Scale Invariant Feature Transform (SIFT) or Histogram of Gradient (HOG) algorithms were used for object detection which was low-level feature based [11]. However, they had limitation of validation performance and neural network based object-detection algorithms were emerged.

Regions with CNN features (R-CNN) algorithm increased the detection validity over twice of the best algorithm before [12]. After that, Spatial Pyramid Pooling in deep convolutional Networks for visual recognition (SPPNet) algorithm quicken the speed 24 ~ 104 times of R-CNN's [13]. In addition, fast R-CNN and faster R-CNN improve the detection of validity and speed [14] [15]. However, their detecting speeds are somewhat slow to integrate in embedded computing systems. Their models are

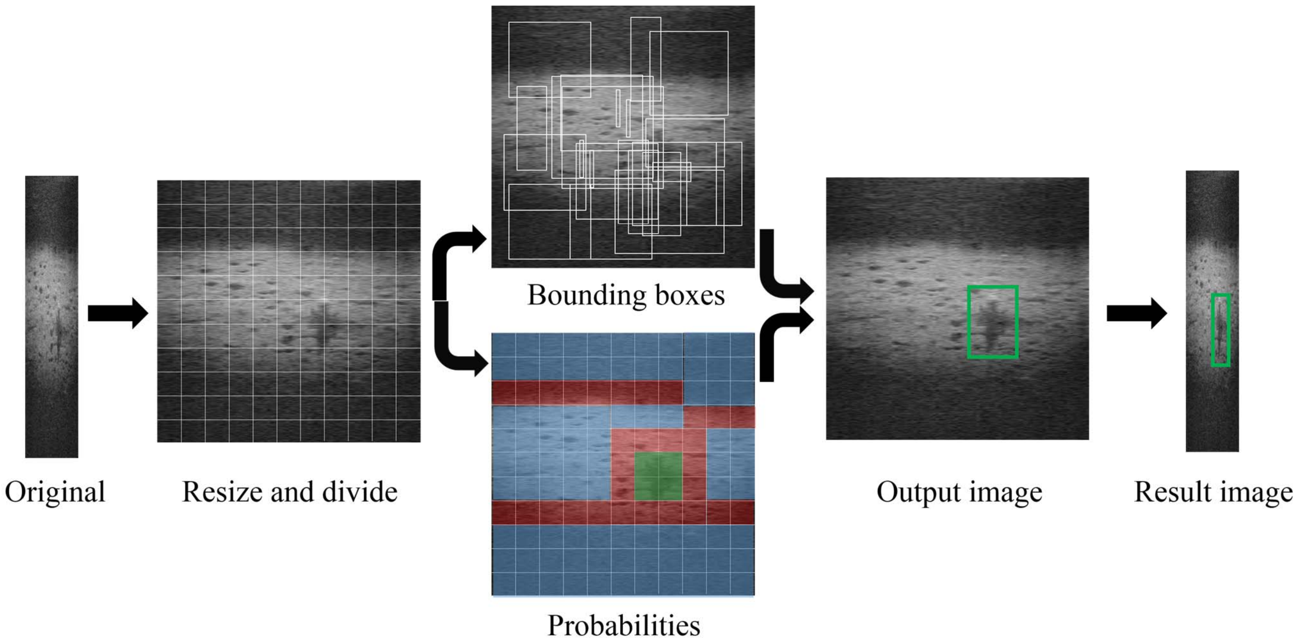

Fig. 3. The YOLO algorithm structure conducting our custom data-set [4].

recurrently calculated CNN several times. Then it took much time to find the ROI of targets

III. Proposed Method

We proposed the real-time object-detection strategy of the underwater small ROV. However, we must proceed the object-recognition before object-detection. It shows the existence probabilities of the target in the sonar images. For this reason, we designed the classifier model to separate the 'positive' images and the 'negative' images. We set the correctly cropped ROV images as the 'positive' and miss-cropped images and background images as the 'negative'.

After designed the classifier, we validated the model and revised the calculation weights. Therefore, we also trained the previous collected forward-looking sonar images that not taken the target as the 'negative'. The post-processing lowers the miss-detection of background's any objects.

When the model showed the high classifying rate, we can apply the object-detection algorithms. It can find the Region Of Interest (ROI) that bounds the target object. We tested the sliding window algorithm and neural network based algorithm.

A. Convolutional Neural Network-based Object-Classifier

We used the machine learning algorithm that includes the training of a large number of image data. The model is 'Darknet Reference Model [16]' in the classical Convolutional Neural Networks (CNNs). It is small but powerful model. It has seven convolution layers and six max-pooling layers.

The training data-set was gathered from real-sea experiment. The hovering-type AUV 'Cyclops' took the forward-looking sonar images at the Jangil Bay, South Korea, 2016 [17] [19]. The AUV took the small ROV when they were launched together [18]. We got the roughly 2,000 images. We spent most of time for making data-set. It includes ROIs and class numbers as label-data. We manually cropped the images and coded label data.

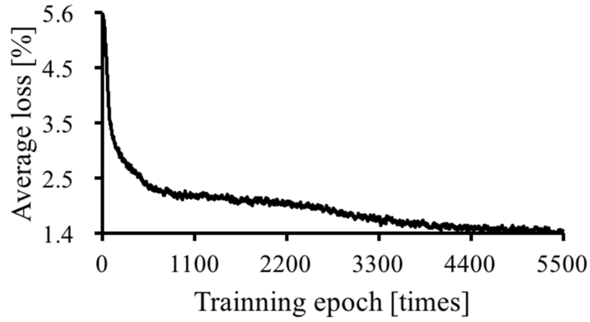

Fig. 4. The average loss function value of trainning.

Each image has two classes of ROIs. One is the 'positive' and the other is the 'negative'. We manually dragged the mouse in the images for precise small ROV's ROI and randomly cropped ROI. With this data, we also made the fake-data for revising the model. We randomly cropped two ROIs in the 1,000 random sonar images and labeled them.

The more we gather the fake-background images, the classifier can lower the recognition error rates. Without revising the model, it detected the 'positive' ROIs in the fake-backgrounds. After retraining of fake-data, it never detected the any other object in the images. In that case, we found that the data-set makes strategy to explore particular underwater regions.

First, take sufficient images of small ROV. We gathered 1,152 images of the small ROV and labeled them. Next, pre-scan the target region's backgrounds. We used 455 images of background. They can be used for training robust model. Then, make the data-set form and do the training by the powerful computation machine such as desktop computer with GPU [20]. After training is finished, we saved the training-weight data.

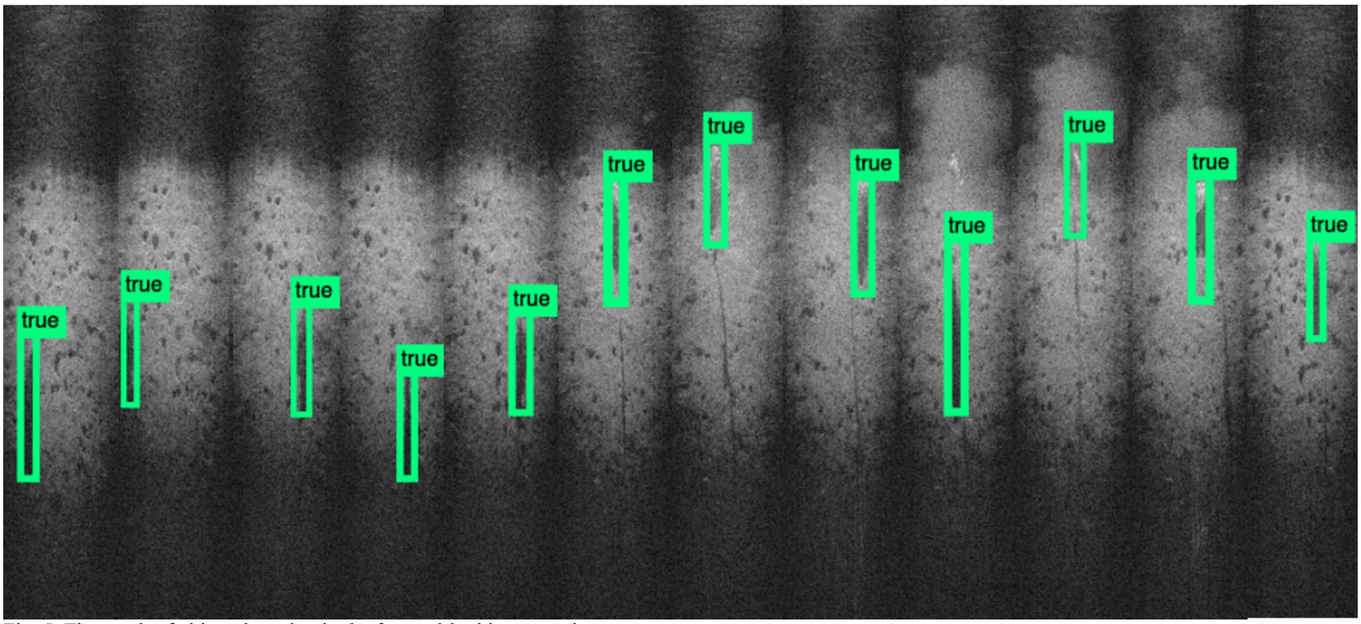

Fig. 5. The result of object-detection in the forward-looking sonar images.

B. Object-detection

Recently, some state-of-art algorithms focused on speed up for real-time processing and control something. We found the You Only Look Once (YOLO) algorithm that is new approach to object-detection [4]. It has single CNN model and predicts both bounding box and class probability. It divides the image 11 by 11 area on the first layer, and connects to classifier model. At the end of classifier model, it is fully connected to divided ROIs and class probabilities (Fig. 3). We used their open source to apply our custom data-set. The form of data-set includes class number and ROI. We retrained the YOLO model with the pre-trained classifier weights. After that, we tested the 2,413 forward-looking sonar images and calculated ROIs and trajectory.

IV. RESULT

A. Experimental Set-up

To verify proposed method, we conducted field experiment. We used a hovering-type AUV 'Cyclops' as the main AUV [17]. While the experiment, its position was fixed with the help of station-keeping control. The station-keeping error was few centimeter, so we neglect the error.

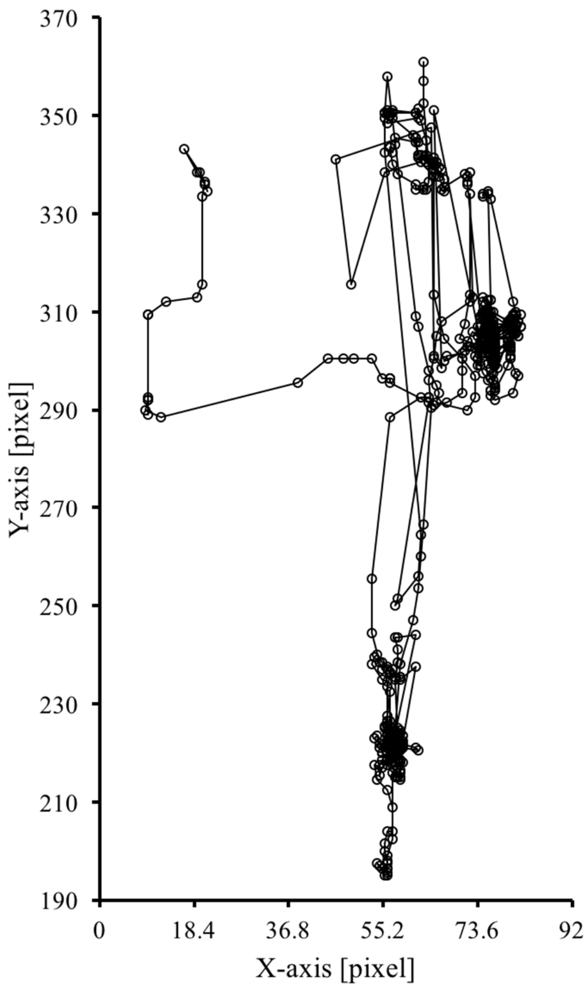

Fig. 6. The trajectories of agent vehicle on the forward-looking sonar images.

We used dual-frequency identification sonar (DIDSON) and it is equipped to the AUV in order to capture the sonar images of the agent [19]. The frame rate of DIDSON was set to 5 frames per second. The resolution of the sonar images was 512 x 96.

Under the settings, the ROV was located in the field of view of forward-looking sonar [18]. Then, the position of the ROV was manually changed. The sonar images obtained while the ROV operation were corrected and tested by proposed method. The number of sonar images are 1,000.

B. Data-set Trainning

We trained the total 1,607 images in YOLO model. We can check the training progress by the loss function (Fig. 4). With the saturation of graph tendency, we can learn that the training leads to positive way. It took about one hour to finish the training.

C. Real-time Object-detection

After the training finished, we tested the total 1,000 images to find location in the images. The YOLO neural network model successfully detect the agent vehicles. Each agent vehicle was precisely bounded by ROI boxes (Fig. 5). We inserted the certain images that is randomly picked within the forward-looking sonar images database. After that, it only showed the negative ROIs and positive ROIs were not existed.

In addition, we recorded the trajectories of each image. It stores x and y axis locations in the images. We deleted the non-detected location and connected the series of images. In this graph, we can figure out the route of agent vehicle.

The speed of process is greatly important because we must use it for real-time underwater missions. The YOLO object-detection algorithm showed the 107.7 Frames per second (FPS) on the off-line processing by GPU [20]. However, simple sliding window algorithm was 0.20 FPS which means too slow (Table 1). In the real-sea, we take forward-looking sonar images as 5 FPS. Therefore, if object-detection result speed is over the 5 FPS, it can be used for real-time controls or missions.

TABLE I. THE COMPARISON BETWEEN TWO ALGORITHMS ABOUT FRAMES PER SECOND.

	Object-detection Algorithms	
	YOLO	Sliding Window
Frames / s	107.7	0.20

D. Discussion and Future work

We proposed the real-time object detection of forward-looking sonar images for localization of agent vehicle. The limited sonar images show the possibility of conducting agent vehicle system. If we gather the more data-set, we can get the reliability of the system. Then, it will conduct the real-sea trial.

If we use the embedded system rather than powerful PCs, the detection speed would be slow. The solution is using the specialized embedded system that includes mobile GPU [21]. State-of-art embedded boards has enough capacity of processing neural networks and can assure long operation time. With these designed system, we will proceed with the agent vehicle system and validate the system's advantages.

V. Conclusion

This study verified the real-time object-detection using forward-looking sonar image that is based on CNN and YOLO. We generated custom data-set and conduct the object-detection algorithm. Then, we realized the localization of small ROV. We found that YOLO algorithm is much effective to process forward-looking sonar images. Finally, it shows applying machine learning algorithms on processing sonar image is much more useful.

Acknowledgment

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] G. Marani, SK. Choi, and J. Yuh, "Underwater autonomous manipulation for intervention missions AUVs." *Ocean Engineering* 36.1 (2009): 15-23.

[2] P. Song, M. Yashima, and V. Kumar, "Dynamic simulation for grasping and whole arm manipulation." *Robotics and Automation*, 2000. Proceedings. ICRA'00. IEEE International Conference on. Vol. 2. IEEE, 2000.

[3] SC. Yu, "A Preliminary Test on Agent-based Docking System for Autonomous Underwater Vehicles." *International Journal of Offshore and Polar Engineering* 19.01 (2009).

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection." arXiv preprint arXiv:1506.02640 (2015).

[5] CD. Loggins, "A comparison of forward-looking sonar design alternatives." *OCEANS*, 2001. MTS/IEEE Conference and Exhibition. Vol. 3. IEEE, 2001.

[6] H. Cho, J. Gu, H. Joe, A. Asada, SC. Yu, "Acoustic beam profile-based rapid underwater object detection for an imaging sonar." *Journal of Marine Science and Technology* 20.1 (2015): 180-197.

[7] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network." *Advances in neural information processing systems*. 1990.

[8] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning." *Nature* 521.7553 (2015): 436-444.

[9] A. Krizhevsky, I. Sutskever, GE. Hinton, "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

[10] S. Lawrence, CL. Giles, AC. Tsoi, "Face recognition: A convolutional neural-network approach." *IEEE transactions on neural networks* 8.1 (1997): 98-113.

[11] DG. Lowe, "Object recognition from local scale-invariant features." *Computer vision,* 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee, 1999.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.

[13] K. He, X. Zhang, S. Ren, J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition." *European Conference on Computer Vision*. Springer International Publishing, 2014.

[14] R. Girshick, "Fast r-cnn." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

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

[16] J. Redmon, "Darknet: Open source neural networks in c." http://pjreddie.com/darknet/, 2013–2016.

[17] J. Pyo, HG. Joe, JH. Kim, A. Elibol, and SC. Yu, "Development of hovering-type AUV "cyclops" for precision observation." *2013 OCEANS-San Diego*. IEEE, 2013.

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

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

[20] GTX 970, *Nvidia Corp*, http://www.nvidia.com

[21] Jetson TX1, *Nvidia Corp*, http://www.nvidia.com

[22] H. Noh, PH. Seo, and B. Han. "Image question answering using convolutional neural network with dynamic parameter prediction." arXiv preprint arXiv:1511.05756 (2015).