Real-Time ROV Recognition Relying on Convolutional Neural Networks Employing Forward-Looking Sonar Images

V.Ramya
Asst. Professor, Department Of Computer Science And Engineering
QIS College of Engineering & Technology
Challa Jaswanth Sai [2], Sai Venkata Sumanth Katragadda [3], Venkata Ajay Kumar Reddy Thumu [4], Sreekanth Alagadri [5], Ajay Nallagorla [6]
Student, Department Of Computer Science And Engineering
[6]QIS College of Engineering & Technology

Abstract- To improve underwater manipulation, the agent system is used. A robot arm-based arrangement with solitary points is often used for conventional manipulation. As opposed to this armless manipulation, the agent system makes use of the agent vehicle as its end-effector. The end effector may be placed in any position if the agent's location can be determined. The technique of locating an agent vehicle has been offered as a means of implementing this system. Forward-looking sonar pictures of moving agents are used in the procedure. The convolutional neural network is used to locate the agent in the sonar images. The agent vehicle system was equipped with the most advanced object-detection algorithm available at the time. The neural network-based quick object-detection method may meet real-time detection requirements and demonstrate exceptional validity. In other words, the underwater robot may begin its journey under its own guidance. Using real-world testing, we've found that the approach we've developed is effective in locating the agent in subsequent sonar photos and following its progress.

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

I. INTRODUCTION

The deep sea is an intriguing field to investigate because of the many unknowns that exist there. Deep sea exploration has been made possible by the fast development of autonomous underwater vehicle technology (AUVs) and remotely operated vehicle (ROV) technology during the last several decades. Using a wide range of sensors, these robots may create maps of the ocean floor or examine specific undersea resources. Research on underwater robots' control theory, navigation systems, and sensors is ongoing. For certain physical procedures, the problem of manipulation in the water is becoming more important.

Some methods of underwater manipulation are available to underwater robots. One of the most common methods is to employ a robot arm. Using a robotic arm, the AUVs or ROVs are able to carry out certain underwater physical tasks. The angle of the joint may be accurately controlled. However, the joint construction, which is termed a single point, is hefty and takes up a lot of space and weight. As a result, the technique of manipulation without the use of arms is created.

The mini ROV is an endeffector robot that is attached to the larger AUV. Long tether ROVs may carry out a wide range of operations without the need for large batteries or specific embedded intelligence. The AUVs have a wide range of instrumentation and sensors because to their freedom of tether. Details may be accomplished with the help of the ROV's modest size. Agent docking and armless underwater manipulation are both possible with this device. A manipulative hand may use it to grasp something or do fine control activities. Using its position sensor and the main AUV's forward-looking sonar, we can determine the agent's location and assign it the name "Agent vehicle system" (Figure 1). The exact location data may lead to accurate manipulation.

An agent vehicle's location may be determined using neural network-based real-time object detection in this research. [4] The You Only Look Once (YOLO) object-detection technique demonstrates great speed and precise detection. Using data from our forward-looking sonar system, we tested this technique. As a consequence, we discovered that it might be used as a feed-back control input.

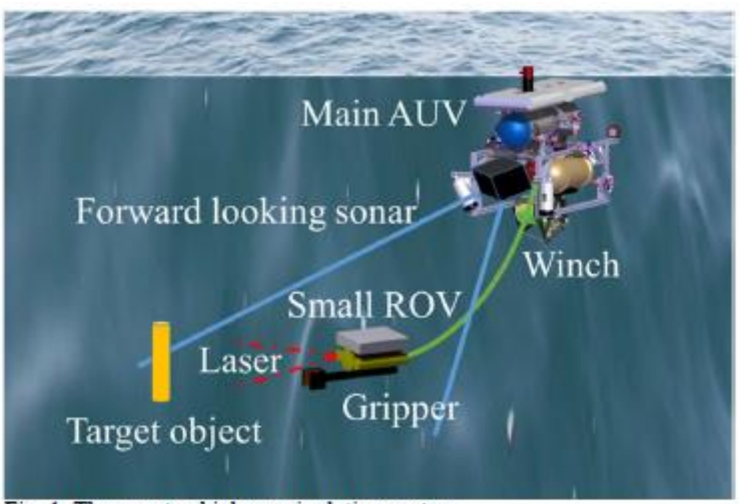

Fig. 1. The agent vehicle manipulation system.

II. RELATEDWORKS

The acoustic video pictures are captured in real time by the forward-looking sonar. Underwater detection might benefit from this technology since it has a greater visual range than optical imaging. However, forward-looking sonar's picture quality is inferior than optical images. Because the auditory pictures are of such poor quality, only the human eye can discriminate between different things (Fig. 2). A lack of detail and excessive levels of noise mar the photos. Aside from that, the single picture clearly displays the three distinct elements of the composition: shadow, background, and highlights. When looking at it from various angles and heights, its image topology reveals the multiple shapes it has.

It's tough to get useful information out of images with these features. As a result, standard image processing techniques are ineffective for analysing sonar pictures. Neural networks are used in the convolutional neural network, which is a supervised machine learning method [7]. Neural network modelling is now the trendiest topic in image processing, thanks to GPU parallel architecture's rising computational power [8]. The large neural network model may be trained and tested in a short period of time [9].

Image processing techniques such as feature matching are standard in most applications. Specific preset forms or post-processing techniques serve as the low-level elements of the system. The convolutional neural network, on the other hand, made advantage of training-level features [10]. The more complex the model becomes, the more difficult it is to examine it rationally [8] [22]. The black-box function of an accurate classifier is generated via supervised machine learning. Classification using an image classifier is limited to showing the potential of something existing. Because of this, it is difficult for us to locate the target item in the picture. For object detection, it is critical to have the right return on investment (ROI). In order to discover ROI instances, there are many algorithms. To identify low-level features, SIFT or HOG algorithms were utilised [11]. However, their validation performance was limited, leading to the development of object-detection algorithms based on neural networks.

In the R-CNN algorithm, the detection validity was double that of the best algorithm previously [12]. [10,11] Spatial Pyramid Pooling in deep convolutional networks for visual identification (SPPNet) accelerates the speed by 24 104 times over R-method CNN's after that[13]. R-CNN and R-CNN increase the detection of validity and speed [14,15]. They are, however, a little sluggish to incorporate into embedded computer systems. CNN has repeatedly recalculated its models. As a result, finding the ROI of targets took a long time.

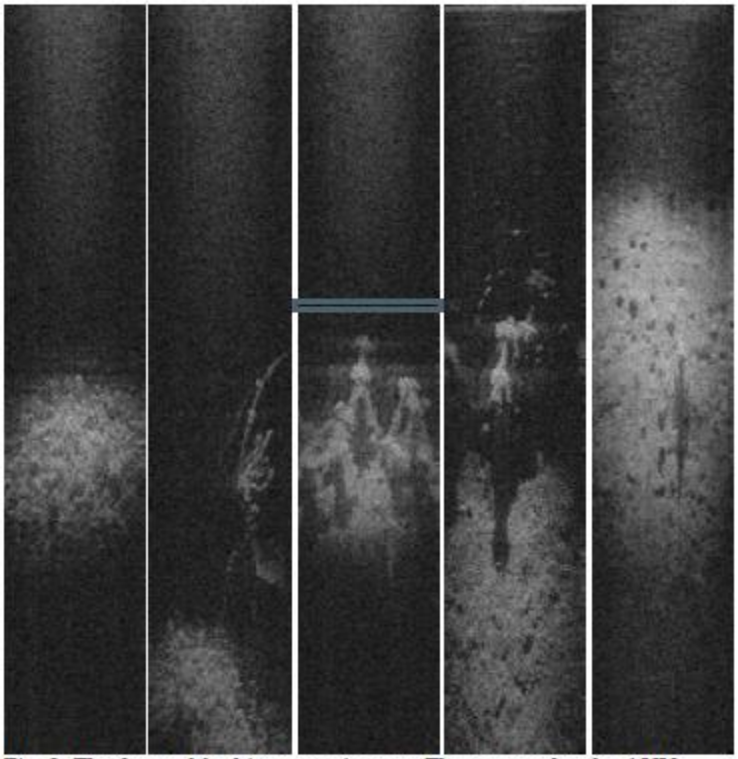

Fig. 2. The forward-looking sonar images. They were taken by AUV 'Cyclops' [17].

III. PROPOSED SYSTEM ARCHITECTURE

The undersea mini ROV has a real-time object-detection technique provided by us. However, prior to detecting an item, we must first recognise an object. Using sonar pictures, it displays the likelihood of the target being present. Classifier models are used to distinguish between "positive" and "negative" pictures. For the 'positive' photographs, we used appropriately cropped ROV images; for the 'negative,' we used images that were incorrectly cropped or had background images. Our model was evaluated and the calculation weights were recalculated after we developed the classifier. Our prior forward-looking sonar scans that did not include the target were also used for training. Post-processing reduces the likelihood of things in the background being missed. We can use object-detection techniques when the model has a high classification rate. For example, it is able to identify the target object's "region of interest" (ROI). The sliding window method and the neural network approach were both put to the test. It was necessary to train a huge number of images using the machine learning method that incorporates image training. Classical Convolutional Neural Networks (CNNs) have a model called "Darknet Reference Model" (CNNs). It's a little model, but it packs a tremendous punch. Six max-pooling and seven convolution layers are included. A real-world experiment was used to acquire the training data set. The forward-looking sonar photos were captured by the hovering-type AUV 'Cyclops' in South Korea in 2016 [17] [19]. When they were launched together, the AUV snatched the little ROV up. We were able to get all 2,000 photos. The bulk of our effort was spent creating a data set. As labeldata, it contains ROIs and class numbers. The photos were manually cropped, and the label data was coded. In each shot, there are two different ROIs. The 'positive' and the 'negative' are both distinct. Images were manually moved with the mouse for exact tiny ROV ROI and cropped at random. We were able to create the bogus data for the model revision using this information as well. Using a random selection of 1,000 random sonar pictures, we identified two ROIs and then labelled them. In order to improve the classifier's identification accuracy, we need to collect additional fake-background photos. The 'positive' ROIs in the fakebackgrounds were discovered by the model without any revisions being made to it. No additional item was spotted in the photos after retraining using phoney data. There, we discovered that the data-set helps us plan our underwater explorations. The first step is to get enough photos of the little ROV. The little ROV was photographed 1,152 times and its photographs were tagged. Pre-scan the target area's surroundings next. There were 455 photos of backgrounds in all. Robust models may be trained using them. A powerful calculation machine, such as a desktop computer equipped with a graphics processing unit (GPU), may then be used to train the model. Once the workout is over, the weight-training data is safely archived. anything that can be controlled in real time. We discovered a novel way to object identification called the You Only Look Once (YOLO) algorithm [4]. Both the bounding box and the class probability are predicted using a single CNN model. Eleven-by-eleven region is sliced up and connected to the classifier model. The classifier model is completely coupled to the split ROIs and class probabilities at the conclusion of the process (Fig. 3). Our bespoke data-set was applied to their open-source software. Class number and ROI are included in the data set format. In order to use the pretrained classifier weights in our YOLO model, we retrained it. After that, we ran the 2,413 forward-looking sonar pictures through their paces and estimated their ROIs and trajectory.

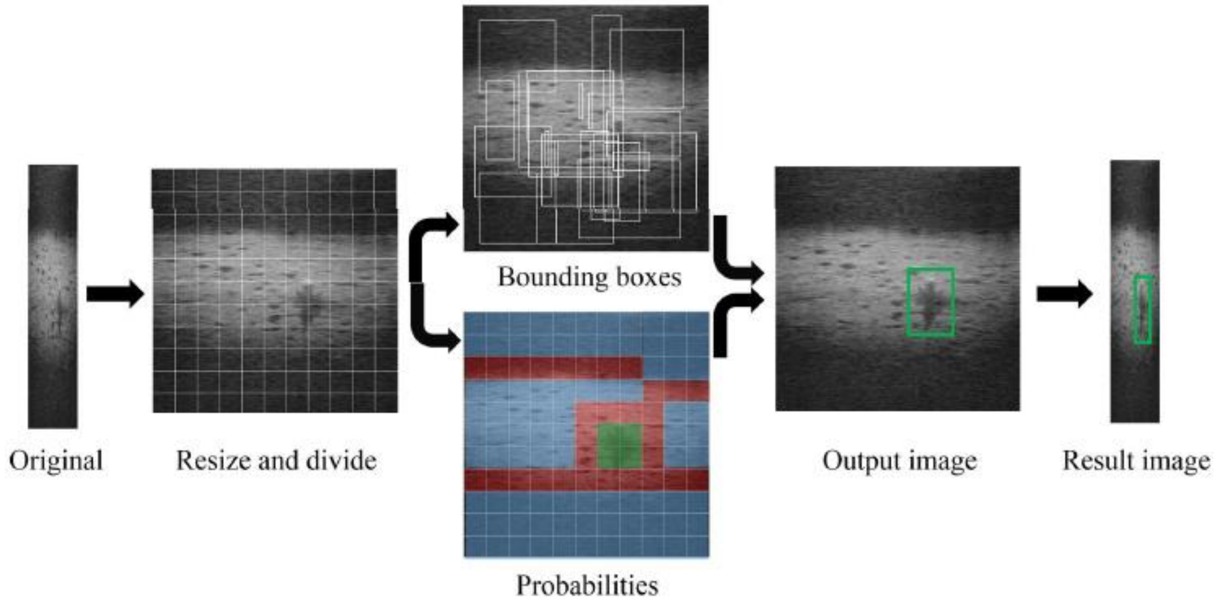

Fig. 3. The YOLO algorithm structure conducting our custom data-set [4].

IV. RESULTS AND DISCUSSION

We performed a field experiment to confirm the suggested strategy. As the primary AUV, we employed a hovering model named 'Cyclops' [17]. Station-keeping control was used to maintain the experiment's location. We disregard the station-keeping inaccuracy since it was just a few centimetres. Our AUV was outfitted with a DIDSON system, which allowed us to record sonar pictures of the agent [19]. There are 5 frames per second in DIDSON. The sonar pictures have a resolution of 512 x 96. This meant that the forward-facing sonar was able to see the ROV. After then, the ROV's location was manually adjusted. The suggested approach was used to rectify and evaluate the sonar pictures collected during the ROV operation. There are 1,000 sonar pictures. The YOLO model was trained on 1,607 pictures. The loss feature may be used to monitor training progress (Fig. 4). We can see from the saturation of the graph tendency that training has a favourable effect. The training was completed in approximately an hour. Using the total of 1,000 photos, we were able to identify the location in each one. The YOLO neural network model was able to correctly identify the agent cars. ROI boxes defined the exact boundaries of each agent vehicle (Fig. 5). We used the forward-looking sonar pictures database to inject a random selection of images. Negative ROIs were all that could be found after that, while positive ROIs were completely absent. In addition, each image's trajectory was recorded. It keeps track of the x and y axes in the photos. We removed the area that wasn't identified and joined the photos. The path of the agent vehicle may be seen in this graph. Because we must utilise it for real-time underwater operations, the speed of the procedure is critical. In off-line processing by the GPU, the YOLO objectdetection technique showed 107.7 FPS [20]. While this method was 0.20 FPS (Table 1), it was still too slow. Forward-looking sonar pictures are captured at a frame rate of five frames per second in the actual world. Consequently, if the object-detection result speed is greater than 5 FPS, it can be used in real-time controls or missions. Using forward-looking sonar images, we developed a real-time object detection method for locating agent vehicles. The scanty sonar images show that an agent vehicle system is possible. The more data we have, the more reliable the system will be. The next step is to conduct a real-world test.

The detection speed would be slowed if we used embedded systems instead of powerful PCs. Use a mobile GPU-powered embedded system to solve the problem [21].

Neural networks can be processed by state-of-the-art embedded boards, which have enough processing power and can run for a long time. We can now move forward with the agent vehicle system and verify its benefits with these newly designed systems.

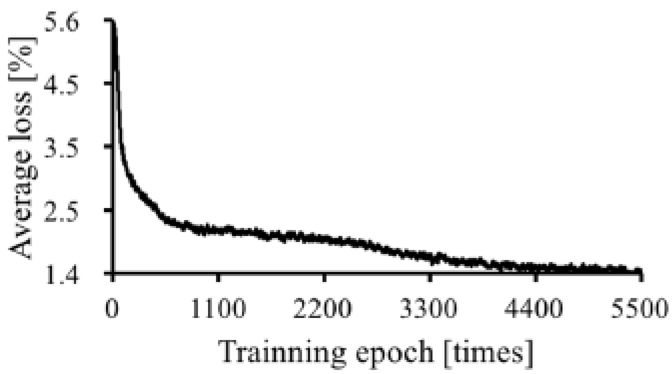

Fig. 4. The average loss function value of trainning.

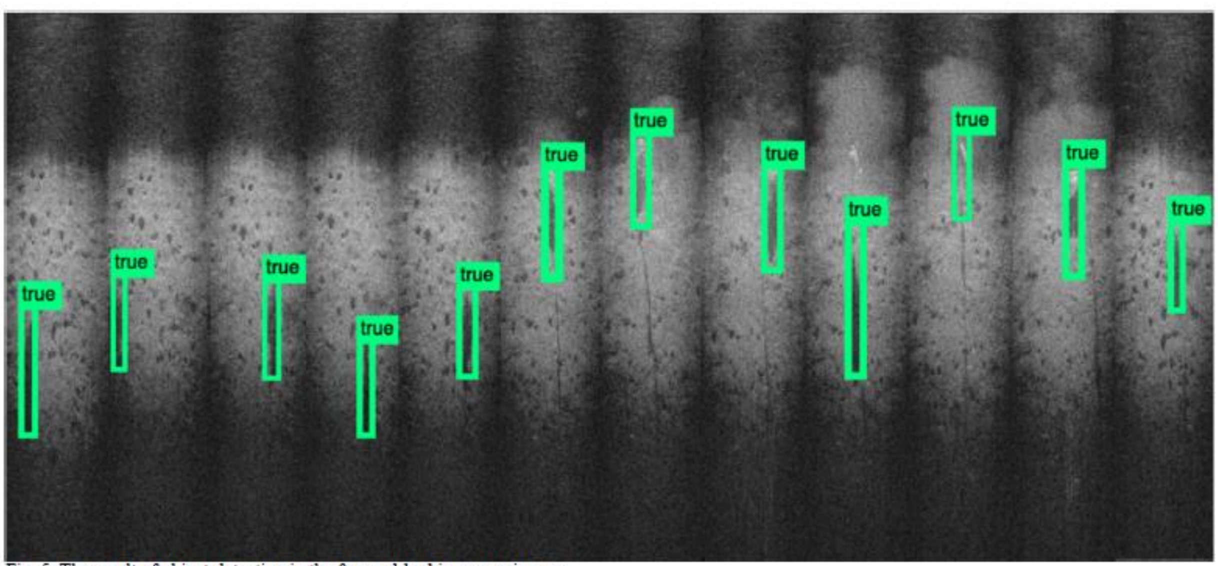

Fig. 5. The result of object-detection in the forward-looking sonar images.

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

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

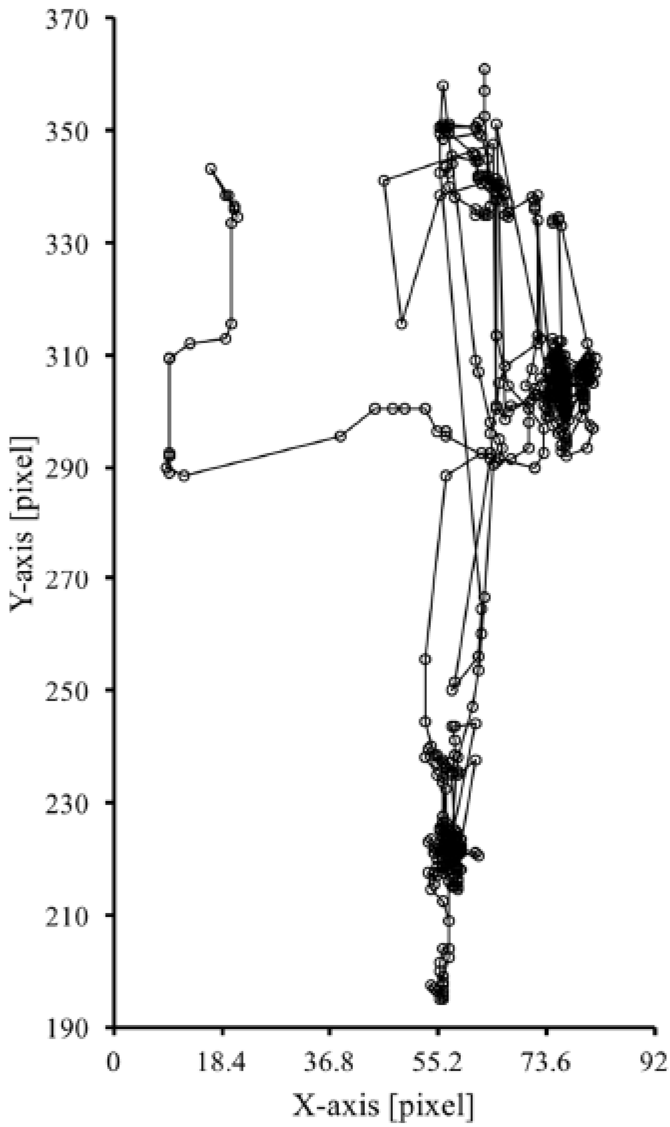

Fig. 6. The trajectories of agent vehicle on the forward-looking sonar images.

V. FUTURE SCOPE AND CONCLUSION

The forward-looking sonar image based on CNN and YOLO was used in this study to verify real-time object detection. The object-detection algorithm is based on a custom data set that we created. After that, we realised that small ROVs could be located. We discovered that processing forward-looking sonar images with the YOLO algorithm was much more efficient. Finally, this study shows that using machine learning algorithms to process sonar images is far more effective.

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 backpropagation 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 realtime 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).