

Thoughts on object detection using convolutional neural networks for forward-looking sonar

Abstract

This work reviews the problem of detection in Forward-Looking Sonar images. In the underwater realm, most of the imaging is done by acoustic means, i.e. sonar. The Forward-Looking Sonar usually has a very low Signal to Noise Ratio therefore object detection in Forward-Looking Sonar images is still an open issue. The article will introduce our database and some conclusions that were gathered from working with it. It will also show results from a Convolutional Neural Network designed for Forward-Looking Sonar Images.

Keywords: convolutional neural networks, forward-looking sonar, autonomous underwater vehicle, machine learning

Volume 4 Issue 3 - 2018

Assaf Livne, Alon Baruch, Hugo Guterman

Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel

Correspondence: Assaf Livne, Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel, Email assafliv@post.bgu.ac.il

Received: October 20, 2017 | **Published:** May 23, 2018

Abbreviations: AUV, autonomous underwater vehicle; SNR, signal to noise ratio; FLS, forward looking sonar; CNN, convolutional neural networks; RPN, region proposal network; LAR, laboratory for autonomous robotics; FPS, frame per second

Introduction

The Hydrocamel II Autonomous Underwater Vehicle (AUV) was developed and built at the Laboratory for Autonomous Robotics (LAR),¹ at Ben-Gurion University. The AUV has a Forward-Looking Sonar (FLS), which is used for different tasks such as obstacle avoidance,² target tracking, infrastructure inspection³ and navigation.⁴ Today in the underwater world it is crucial to detect and classify objects such as pipes, rocks, corals, mines, etc. However, object detection and classification in sonar images is a difficult task due to the low Signal to noise ratio (SNR). Furthermore, due to the nature of the application, the image processing has to be done in real-time using the AUV's available hardware. A FLS can capture high-resolution images of underwater scenes, but their interpretation is complex. Generic object detection in such images has not been solved, most of the research that has been done on object detection was not robust to a variety of objects and has a low real-time factor.^{3,5-10} Since 2012's Image Net challenge, the use of classic image processing tools in classification and detection tasks in regular colored images, has decreased dramatically. Today the academic community uses deep neural networks for those tasks. Detection and classification algorithms based on CNN have produced top performing object detectors and classifiers in real-world color images.¹¹ The main purpose of this article is to introduce some findings that have been gathered over the years about object detection using convolutional neural networks (CNN) for FLS images (Figure 1). Any system that is designed to detect objects in FLS images should meet the requirements of an embedded system that can run in real-time and with low power consumption. Also these article summaries the results of a designed CNN for object detection in FLS images.

Related work

Classification and detection in FLS images

Over the past few years, underwater vehicles have greatly improved as a tool for undersea exploration. Today with the new technology of high definition FLS, the detection problem can be addressed again.

The new generation of FLS can provide better acoustic imagery with higher frame rates. However, the characteristics of the sonar data introduce different difficulties in the object detection problem. There is a large amount of literature about object classification in the marine domain, some approaches use template matching⁵⁻⁷ while others use engineered feature extractors⁸⁻¹⁰ with a trained classifier. Some researchers suggest that the FLS image features are based on shadows or highlights.³ In order to use the template matching; one needs to define a template for each class, which indicates how a class should look in an FLS image. A maximum likelihood function is used to differentiate between the classes. These methods can perform with close to 95% accuracy, but it is hard to generalize this method for a large number of classes. Between the engineered feature extractors one can find the use of a cascade of features based on Haar features¹¹ and even a use of some feature extractors such as SIFT.¹² These methods are faster but less accurate. These techniques represent the classical image processing methods. It has been shown that in regular camera images CNN concepts can improve the results. Using a CNN instead of the classical image processing methods improved the accuracy and the robustness of the object detection task.¹¹

Object classification and detection in color images using CNN concepts

Since Krizhevsky's¹³ work it has become a known fact that object classification in the image processing world can be done using CNN. While Krizhevsky¹³ was one of the first to demonstrate this in 2012, additional networks were designed for the object detection task.^{14,15} In the 2016 Image Net¹⁶ classification challenge the error was under 3% which indicates the power of the CNN. The Faster R-CNN¹⁷ combined a region proposal network (RPN) with a classification network to address the detection problem. The Faster RCNN can work at 7 Frames per Second (FPS). YOLO¹⁸ tried to tackle the detection problem by increasing the FPS for real-time implementations. This presents a new approach to detection, where a single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. YOLO achieved 45 FPS. The CNN technique showed great results for the object detection problem in regular camera images both in accuracy and runtime. But still, as will mention in the next sections there is a place for improvement in sonar imaging.

CNN concepts for FLS images

Valdenegro-Toro¹⁹⁻²¹ used CNN for detection and localization in FLS images. In his work, an ARIS Explorer 3000 sonar was employed to capture a database of 2500 FLS images. The dataset which contains different objects in a water tank was employed to train a small CNN for detection and classification. In Valdenegro²¹ words: “This key finding signals that deep and convolutional neural networks are a clear direction for future research in sonar image processing”. Juhwan et al.^{22,23} applied CNN to FLS images to localize a remotely operated vehicle (ROV) linked to the AUV through a tether cable. Two different labeled images, the ROV images, and some background images were collected. By using CNN they recognized the ROV and improved the localization of it. Using the same CNN architecture as YOLO they manage to track the ROV in 5FPS. Their conclusion was: “it shows that applying machine learning algorithms on processing sonar image is much more useful”. These two examples show the potential in using CNN concepts for FLS images, but still, both of them didn’t succeed to build a solution for the objects detection task to AUVs.

FLS image dataset

The Hydrocamel II has a Blue View M900x. The sonar has a range of 100 meters and a field of view of 130degrees. The sonar image space is a matrix with a width of 900pixels and a height of 896pixels. This article will only discuss Horizontal FLS images. The pixels can be transformed into a Cartesian coordinate system, an angle and distance from the AUV. The data was taken from a series of experiments in the Red and the Mediterranean Seas. The data was annotated manually. The database is built from 118000 sonar images, in which 1000objects were tagged. 50objects were metallic like objects and the rest were large rocks or unknown objects (Table 1). The small number of objects in an image is a common result, but this can be a problem for training the network. Because of the low ratio between the true labeled images to the false labeled images, the training process will easily converge into the False System, a system that returns false for every image. Controlling the ratio between the true labeled images to the false labeled images will prevent this problem. Another Issue is that in FLS images the raw image pixels to meter factor is depending on the pixel position. The meaning of this is that objects farther than the AUV will look smaller in pixel area than objects that are near the AUV. This can fool the CNN because the objects are labeled with the same notation. A solution can be working in the Cartesian coordination system but a different problem will be the borders of the image. In Cartesian coordination system, the image is a sector of a circle as shown in Figure 2. The last issue that should be mentioned is that today it is impossible to find an open FLS image database that can be used to compare different methods for object detection. An open dataset will increase the development time of perception algorithms in the AUV community.

Table 1 Sonar images details

Sonar videos	1000
Sonar images	118000
Image size	900 * 896
Object annotated	1000
% of metallic object	5
% of large rocks	95

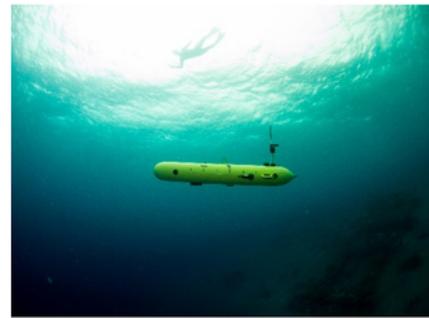


Figure 1 Hydrocamel II.

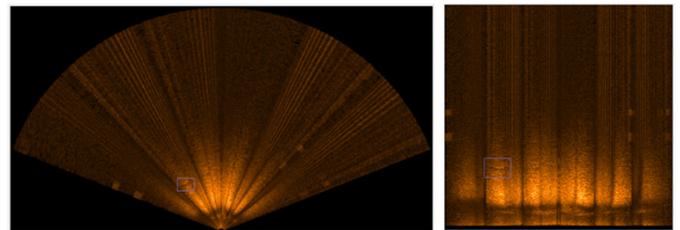


Figure 2 Sonar images examples: Cartesian coordinate system and radial coordinate system.

Basic CNN structure

The basic CNN structure that has been used in this article is composed of three convolutional layers and three fully connected layers, a dropout layer and some pooling layers. The ReLu activation function is being used between the layers. The first convolutional layer is built from three on three kernels with 32filters. The second convolutional layer is built from three on three kernels with 64filters. The last convolutional layer is built from five by five kernels with 128filters. In Figure 3 the design is described. The weights initialized with zero mean normal distribution and variance of 0.001, the biases initialized with the constant 0.01. For training and testing, the images were cropped into N smaller images size MxM. This has been done to address the issue described in the last section. Every slice that has a labeled object on it got the true label and the rest got the false label. The data was in a radial coordination system. Lastly, for every true labeled slice, two false label slices were randomly added. Three datasets were trained and tested on the basic CNN structure. Details about them are summarized in Table 2. Training the basic CNN structure for 5000 iterations with a batch size of 128 gave the results in Figure 4. This basic CNN structure only converges into one of the two possible outcomes. The first one is the false system, for every input the output will be false. The other one is the true system, for every input the output will be true. Trying to change the ratio between the false labeled images and the true labeled images didn’t yield any different results. These results point out that the CNN couldn’t learn the features of the objects. It could be that because of the high SNR the features can’t be observed in a single image.

Table 2 Summary of different Datasets

Dataset Name	M	# of train slices	# of test slices
Database20-r	20	1163052	129228
Database50-r	50	45441	5049
Database100-r	100	10692	1188

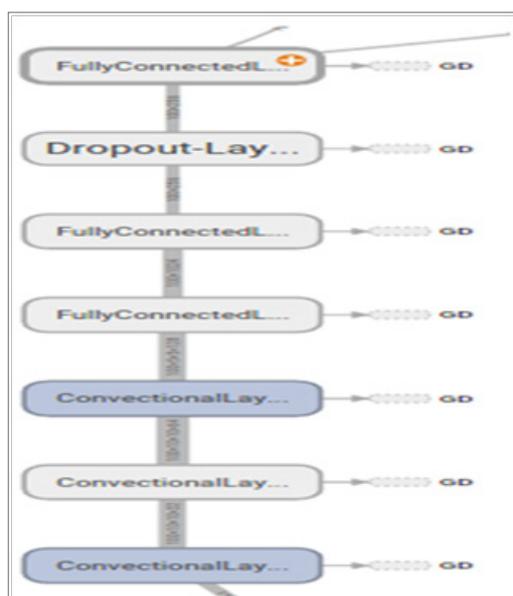


Figure 3 The basic CNN structure used in this article. The Blue layers have pooling layers. The ReLu function was used as an activation function.

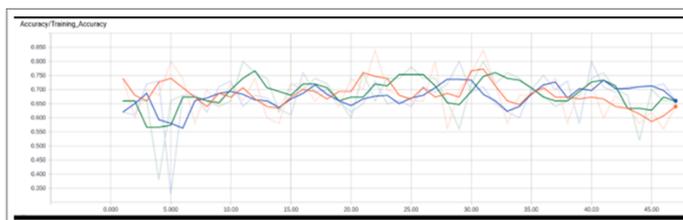


Figure 4 the training accuracy of the different databases. The blue line is for the Database20-r, the orange line is for the Database50-r and the green line is for the Database100-r.

Conclusion

This article discusses the object detection problem in FLS images. Some findings of the features structure were described in detail and also a basic CNN structure was introduced and tested. The results show that much more work is needed to be done. A deeper network should be considered but the power consumption could be a problem for AUVs. In contrast to regular camera images the features in FLS images are harder to learn. Trying to use a basic CNN structured did not yield any results. A recurrent network can be considered for this type of problem.

Acknowledgements

None.

Conflict of interest

The author declares there is no conflict of interest.

References

1. Welcome to LAR.
2. Braginsky B, Guterman H. Obstacle avoidance approaches for autonomous underwater vehicle: simulation and experimental results. *IEEE Journal of Oceanic Engineering*. 2016;41(4):882–892.
3. Hurtós N, Ribas D, Cufí X, et al. Fourier-based registration for

robust forward-looking sonar mosaicing in low-visibility underwater environments. *Journal of field robotics*. 2014;32(1):123–151.

4. Baruch A, Kamber E, Arbel I, et al. Navigation approaches for hovering autonomous underwater vehicles. Science of Electrical Engineering (ICSEE), IEEE International Conference; 2016 Nov 16-18; Eilat, Israel: IEEE; 2016.
5. Myers V, Fawcett J. A template matching procedure for automatic target recognition in synthetic aperture sonar imagery. *IEEE Signal Processing Letters*. 2010;17(7):683–686.
6. Hurtós N, Palomeras N, Nagappa S, et al. Automatic detection of underwater chain links using a forward-looking sonar. OCEANS-Bergen, 2013 MTS/IEEE; 2013.
7. Midelfart H, Groen J, Midtgaard Ø. Template matching methods for object classification in synthetic aperture sonar images. 2009.
8. Fandos R, Zoubir A, Siantidis K. Unified design of a feature-based ADAC system for mine hunting using synthetic aperture sonar. *IEEE Transactions on Geoscience and Remote Sensing*. 2014;52(5):2413–2426.
9. Sawas J, Petillot Y, Pailhas Y. Cascade of boosted classifiers for rapid detection of underwater objects. ECUA 2010 Istanbul Conference; 2010.
10. Sawas J, Petillot Y, Pailhas Y, et al. Target recognition in synthetic aperture and high resolution side-scan sonar. 2010;10:1–8.
11. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition*; 2001:1–9.
12. Lowe D. Object recognition from local scale-invariant features. Proceedings of the International Conference on Computer Vision; 1999 Sept 20-27; Kerkyra, Greece: IEEE; 2002.
13. Krizhevsky A, Ilya S, Geoffrey H. Image net classification with deep convolutional. 2012.
14. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014.
15. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. Springer; 2014:818–833.
16. Fei-Fei L, Russakovsky O, Deng J, et al, Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*. 2015;115(3):211–252.
17. Ren S, He K, Girshick R, Sunet al. Faster r-cnn: Towards real-time object detection with region proposal networks. 2015.
18. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. 2016.
19. Braginsky B, Baruch A, Guterman H. Tracking of autonomous underwater vehicles using an autonomous surface vehicle with ranger interrogator system. OCEANS 2016 MTS/IEEE Monterey; 2016.
20. Valdenegro-Toro M. Objectness scoring and detection proposals in forward-looking sonar images with convolutional neural networks. 2016.
21. Valdenegro-Toro M. End-to-end object detection and recognition in forward-looking sonar images with convolutional neural networks. Autonomous Underwater Vehicles (AUV), 2016 IEEE/OES; 2016.
22. Valdenegro-Toro M. Object recognition in forward-looking sonar images with convolutional neural networks. OCEANS 2016 MTS/IEEE Monterey; 2016.
23. Juhwan K, Son-Cheol Y. Convolutional neural network-based real-time rovdetection using forward-looking sonar image. Autonomous Underwater Vehicles (AUV), 2016 IEEE/OES; 2016.
24. Juhwan K, Hyeonwoo C, Juhyun P, et al. The convolution neural network based agent vehicle detection using forward-looking sonar image. OCEANS 2016 MTS/IEEE Monterey. 2016.