3D-Modeling Dataset Augmentation for Underwater AUV Real-time Manipulations*

Chua-Chin Wang[1]
Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan 80424
Email: ccwang@ee.nsysu.edu.tw

Chia-Yi Huang
Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan 80424
Email: mark584967@vlsi.ee.nsysu.edu.tw

Chu-Han Lin
Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan 80424
Email: owen850612@gmail.com

Chia-Hung Yeh[2]
Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan 80424
Email: yeh@mail.ee.nsysu.edu.tw

Guan-Xian Liu
Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan 80424
Email: m063010044@student.nsysu.edu.tw

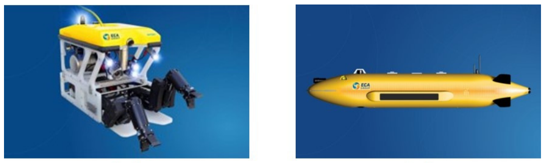
Yu-Cheng Chou
Institute of Undersea Technolody
National Sun Yat-Sen University
Kaohsiung, Taiwan 80424
Email: ycchou@mail.nsysu.edu.tw

***Abstract*—Underwater real-time object recognition is essential to unmanned underwater drones, namely autonomous underwater vehicles (AUV), cruising in the ocean. As the deep learning technology evolves swiftly lately, the attempt for AUVs to fully understand the surrounding environment becomes an emerging demand for marine or military applications. No matter which approach that deep learning manages to adopt, a large dataset with sufficient number of images for each object is required. In this investigation, a dataset augmentation method based on 3D modeling is proposed to resolve the mentioned problem. By rotating and scaling the target object in 3 dimensions with different backgrounds, the number of underwater object images is increased over 1000 times. Through the proposed method, high quality image data are forged to improve the recognition accuracy of those rare underwater objects, which are very hard to collect enough number of images, by 20% based on real-time video clips' experiment.**

***Index Terms*—AUV, real-time recognition, underwater object, 3D modeling, dataset**

I. Introduction

Ocean occupies over 70% of earth surface, where most of the deep sea remains unexplored. Ocean is also considered the ultimate final frontier of natural resources such that the exploration in the deep sea become very competitive worldwide. This is also one of the reason that the market size of the marine industry is over 2/3 of the semiconductor industry globally. Only 5% of the entire ocean is explored by human due to the poor accessibility. The reliable tools for human to look into the deep sea are wither MSV (man submersible vehicle), e.g., submarines, or UUV (unmanned underwater vehicle). The former one might result in the risk of life caused by unexpected pressure or current accidents. Most of the deep sea areas needs the latter to try the pilot runs of exploration [1]. UUVs are classified into two major categories : ROV (remote operating vehicle) [2], and AUV (autonomous underwater vehicle) [3], as shown in Fig. 1. ROVs basically use a cable carrying power and signals from a mother ship so that high bandwidth communication and reliable power supply are feasible to conduct intensive search of seabed. However, the range of ROV exploration apparently is limited by the length of that cable. By contrast, AUVs need no cable, which cruise freely under the water surface. Nevertheless, AUVs' operation time is highly constrained because they are powered by batteries. In other words, the power supply is strictly limited. With the unlimited operation range, AUV is deemed as a better solution to explore unknown underwater world if the power efficiency can be managed. AI (artificial intelligence) is another weapon to be equipped with AUVs so that AUVs will know what the environment is now and what to recognize during the cruise. AI is now known to be realized by deep learning or deep machine learning, which has already achieved tremendous success in image processing for object recognition, e.g., [4].



(a) (b)

Fig. 1. Examples of (a) ROV; (b)AUV

A strong demand in the collection of underwater object images is driven by the exploration and monitoring of underwater ecosystems, including fish, seabed, seagrass meadows,

*This investigation was partially supported by Ministry of Science and Technology, Taiwan, under grant MOST 108-2218-E-110-002 and 109-2218-E-110-007.

[1]Prof. C.-C. Wang is also with Inst. of Undersea Tech., National Sun Yat-Sen Univ, Kaohsiung, Taiwan.

[2]Prof. C.-H. Yeh is also with Dept. of Elec. Eng., National Taiwan Normal Univ., Taipei, Taiwan.

978-1-7281-9396-0/20/$31.00 ©2020 IEEE

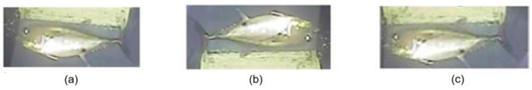


Fig. 2. Combination of basic image processing : (a) rotation; (b) mirroring; (c) partition

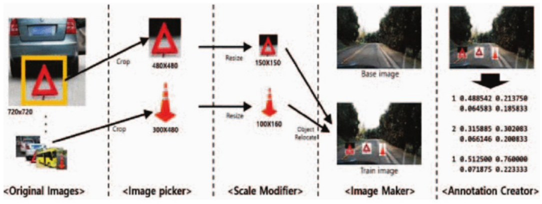


Fig. 3. Example of web crawler [7]

etc. The reason is that without enough images, AI (DNN, CNN, etc.) technologies will not attain accuracy of object recognition, particularly in real-time underwater operations where the light is not good enough and power supply is doubtful. Theoretically, at least 10,000 pictures with various viewpoints, sizes, and positions for the same object is needed to achieve 90% accuracy by commercially available tools, e.g., [4]. This requirement is very hard to meet for rare object underwater, e.g., special plankton, sea spider, etc. To resolve this problem for DNN (deep neural network), CNN (convolutional neural network), or RNN (recurrent neural network) to carry out recognition and classification, a dataset augmentation method using 3D modeling is proposed in this investigation. Without using any new image, the proposed method increases the accuracy over 20%. Experimental simulation also verifies that the dataset generated by the proposed method outperforms many existing benchmark dataset.

II. Underwater Dataset Augmentation by 3D Modeling

Many approaches were reported to resolve the insufficient images for AI learning problem. However, almost all of these reported methods are composed of a few well-known approaches. Two major approaches are analyzed as follows.

• Geometry-based operations : Owing to the scarcity of specific underwater object image, many researchers proposed to apply basic image processing skills, e.g., rotation, partition, mirroring, and scaling [5] [6], to generate more images for AI learning, as shown in Fig 2. Lacking variety of viewpoints is the major problem of this approach.

• Web searching : Apparently, internet is an abundant resource for any image. Lots of researchers developed various web searching technologies to propel the enrichment of underwater images, e.g., searching tags. An example is shown in Fig. 3. Although it is convenient, the problem of this approach is the users of AI learning tools still need time and effort to screen those images which are not they really want and label others.

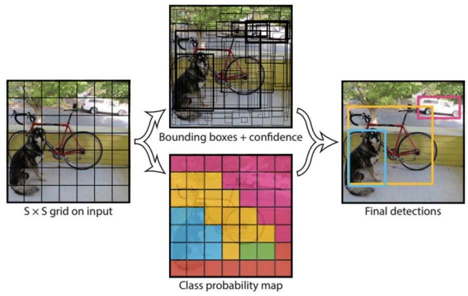


Fig. 4. Example of YOLO training and recognition

A. Real-time recognition fundamentals

Ever since AI algorithms using DNN or CNN were highly promoted in early 2010's, e.g., [8], YOLO has been known one of the best options for real-time objection recognition tools. Basic YOLO operations are similar to other learning networks of which the training can be summarized as the following steps.

(a). Read the entire picture without editing.
(b). The image is divided into many S×S frames. Feature extraction is then executed.
(c). Execute CNN directly on the picture.
(d). By applying Non-max suppression algorithm, the final prediction or recognition is attained based on the confidence factor.
(e). Estimate 5 coordinates in a bounding box, namely (x,y,w,h,f). If the confidence factor is higher than a predefine threshold, the probability of the object is high.

Referring to Fig. 4, it is a typical example of YOLO operations. Nevertheless, any object recognition expected to be over 90% accuracy needs more than 10,000 training images. What even worse is that these training images can not have high similarity, which makes the traditional geometry-based augmentation methods difficult to generate enough number of quality images for training.

B. 3D modeling theory

Apart from the conventional augmentation methods, 3D modeling attains two features : multiple viewpoints and backgrounds. Referring to Fig. 5, assume the target object is placed at origin, (0,0,0). At least 26 viewpoints can be generated, which are from (x, y, z), where $x, y, z \in \{-1, 0, +1,\}$ and $x, y, z \neq 0$ at the same time. The amount of information that a viewer can collect from an object is proportional to the projection area from his/her viewpoint. Therefore, those viewpoints with maximum information is $(x, y, z), \mid x \mid=\mid y \mid=\mid z \mid= 1$. Namely, the 8 corners of the cube in Fig. 5 are the best selections. More viewpoints certainly will introduce more information for learning. As for the added background, basic selections include : clear water, dark water, coral reef, rocks, seagrass.

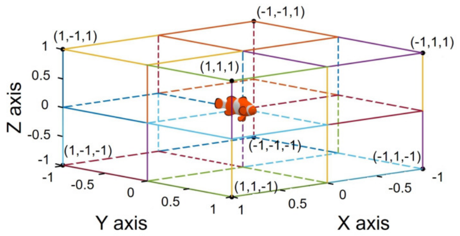


Fig. 5. Multiple 3D modeling viewpoints

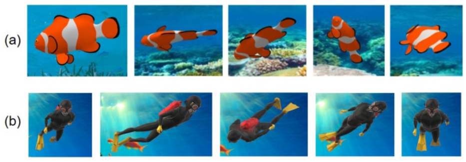


Fig. 6. Examples of 3D modeling : (a) clown fish; (b) diver

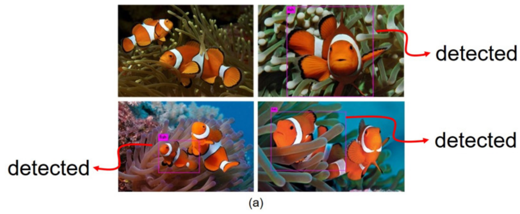


Fig. 8. clown fish recognition without 3D modeling dataset augmentation

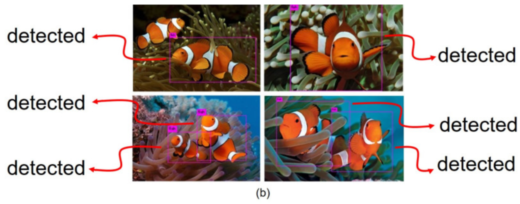


Fig. 9. clown fish recognition with 3D modeling dataset augmentation

C. 3D modeling dataset augmentation procedure

Most the available DNN or CNN tools for AI learning should be able to read the entire image without any processing besides labeling. The recognition is then considered as a regression problem. Thus, to speed up the learning and elevate the recognition efficiency, we propose the following dataset augmentation methods.

(1). Generating the 3D model of a target object
(2). Change the viewpoint to this target object
(3). Add underwater background for this target object
(4). Change the position and size of this target object in the image with a background

The best feature of this method is that the location of the target object is known before labeling preprocess such that the effort of identification is drastically reduced. With reference to Fig. 6, the clown fish and the diver images can be easily generated.

Moreover, after the addition of underwater background, the images are very close to real ones as shown in Fig. 7. Labeling preprocess can be easily carried out as well.

The advantage of 3D modeling method compared with prior augmentation methods can be demonstrated by the following equations.

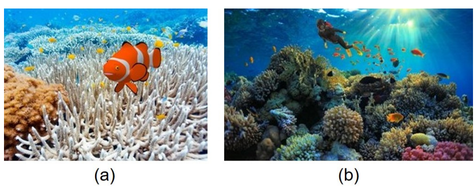


Fig. 7. 3D modeling with underwater background : (a) clown fish; (b) diver

$$\text{Ratio}_{\text{geom}} = \#\text{rotation} \times \#\text{scale} \times \#\text{mirroring} \quad (1)$$

$$\text{Ratio}_{\text{3D}} = \#\text{rotation} \times \#\text{scale} \times \#\text{mirroring} \times \#\text{viewpoint} \times \#\text{background} \quad (2)$$

where $\text{Ratio}_{\text{geom}}$ and Ratio_{3D} denote the ratios of dataset augmentation using conventional geometry-based methods and our proposed method, respectively. The typical numbers are 40 ($4\times5\times2$) vs. 1600 ($4\times5\times2\times8\times5$) for the ratios of these two methods.

D. Primitive trials of 3D modeling

To make a fair comparison with and without 3D modeling, we firstly use clown fish as the object to be recognized. Fig. 8 shows that 3 out of 7 clown fish are found in the 4 pictures. By contrast, 6/7 are recognized in Fig. 9. A brief summary of the clown fish trial is given in Table I, where a simple application upon clown fish recognition shows at least 14% increase of accuracy by the 3D modeling approach.

TABLE I
CLOWN FISH RECOGNITION

	w/o 3D modeling	with 3D modeling
# images	200	1200
# test	90	90
# recognized	51	63
accuracy	56.7%	70.0%

III. EXPERIMETAL SIMULATIONS ON UNDRWATER BENCHMARK DATASET

The proposed dataset augmentation method based on 3D modeling is physically realized to justify the expected performance. Several benchmark datasets are used to make a fair comparison with the proposed method.

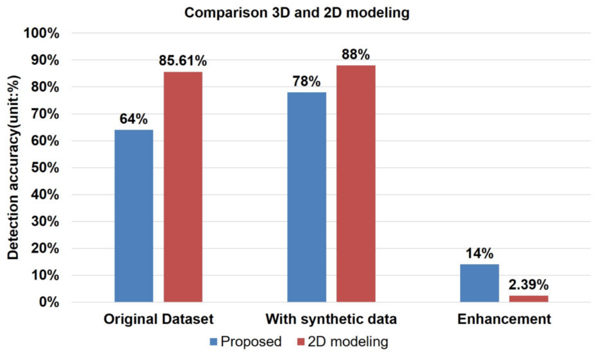


Fig. 10. Accuracy enhancement of the proposed method vs. the conventional approach [10]

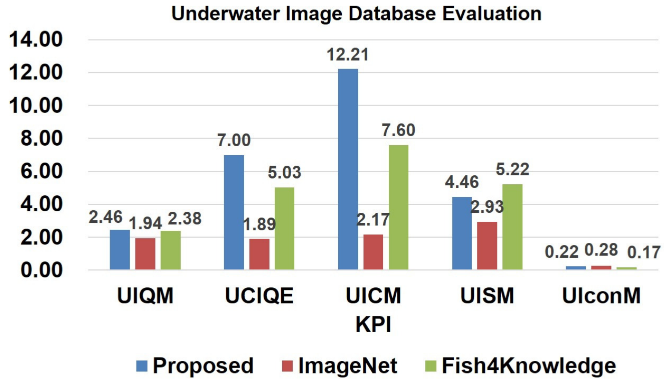


Fig. 11. Performance Comparison

A. Comparison with geometry-based approach

The comparison of the proposed 3D modeling augmentation method with conventional geometry-based approaches can be told by Fig. 10. Our method increases the accuracy around 12% far higher than the prior work [10].

B. Comparison of benchmark datasets

Many widely recognized marine or ocean related datasets are spread over the internet, e.g., Singapore Maritime Dataset (SMD) [9], ImageNet [11], Fish4Knowledge [12]. SMD is for the vessels over the water so that it is apparently not for underwater objects. We have conducted the same learning and testing procedures to the latter 2 datasets and ours to make a fair comparison using the following key performance indices (KPI) [13] [14].

- UIQM (underwater image quality measure)
- UCIQE (underwater color image quality evaluation)
- UICM (underwater image colorfulness measure)
- UISM (underwater image sharpness measure)
- UIconM (underwater image contrast measure)

The above 5 KPIs of three underwater datasets are tabulated in Table II. The proposed 3D modeling augmentation methods are the best of 3 KPIs and the second best of the other 2. Fig. 11 is graphical demonstration of the these KPIs.

TABLE II
PERFORMANCE COMPARISON OF 3 UNDERWATER DATASETS

	ImageNet [11]	Fish4Knowledge [12]	3D modeling ours
UIQM	1.94	2.38	**2.46**
UCIQE	1.89	5.03	**7.00**
UICM	2.17	7.60	**12.21**
UISM	2.93	**5.22**	4.46
UIconM	**0.28**	0.17	0.22

IV. CONCLUSION

This investigation demonstrates a cost-effective dataset augmentation method using 3D modeling. Not only theoretical analysis of viewpoints is presented, the detailed procedure to use the proposed method is fully described. The comparison with existing underwater datasets fully shows that the proposed method outperforms the prior works in most of the well-defined KPIs. All of these facts justify the superiority of the proposed method.

ACKNOWLEDGMENT

The authors would like to express our deepest appreciation to TSRI (Taiwan Semiconductor Research Institue) in NARL (Nation Applied Research Laboratories), Taiwan, for the assistance of EDA tool support.

REFERENCES

[1] https://scitechvista.nat.gov.tw/c/hK4J.htm.
[2] https://www.ecagroup.com/en/solutions/h300-or-rov-remotely-operated-vehicle.
[3] https://www.ecagroup.com/en/solutions/a27-m-auv-autonomous-underwater-vehicle.
[4] https://www.itread01.com/content/1548042482.html
[5] T.-L. Wey, C.-H. Wu, C.-H. Lin, and H.-H. Lin, "Fish recognition based on deep learning," in *Proc. 2016 Conference on Information Technology and Applications in Outlying Islands,* pp. 31-38, May 2016.
[6] https://tw.openrobot.club/article/index?sn=10999
[7] H.-J. Jeong, K.-S. Park, and Y.-G. Ha, Image preprocessing for efficient training of YOLO deep learning networks, in *Proc. 2018 IEEE Conference on Big Data and Smart Computing (BigComp),* pp. 635637, May 2018.
[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You Only Look Once: unified, real-time, object detection, in *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 779788, Dec. 2016.
[9] Singapore Maritime Dataset (SMD), https://sites.google.com/site /dilipprasad/home/singapore-maritime-dataset
[10] H.-C. Shin, K.-I. Lee, and C.-E. Lee, "Data augmentation method of object detection for deep learning in maritime image," in *Proc. 2020 IEEE International Conference on Big Data and Smart Computing (BigComp),* pp. 463-466, Feb. 2020.
[11] http://www.image-net.org/
[12] http://groups.inf.ed.ac.uk/f4k/
[13] https://ieeexplore.ieee.org/abstract/document/7300447
[14] https://ieeexplore.ieee.org/abstract/document/7305804