2019 19th International Conference on Control, Automation and Systems (ICCAS 2019)
Oct. 15~18, 2019; ICC Jeju, Jeju, Korea

Data Augmentation Method for Object Detection in Underwater Environments

Jung-min Noh[1,2], Ga-Ram Jang[2], Kyoung-Nam Ha[3] and Jae-Han Park[1,2*]

[1]Dept. of Industrial Technology, University of Science and Technology, MS course,
Daejeon, 34113, Korea (yesmin3548@kitech.re.kr)
[2] Dept. of Robot R&D Group, Korea Institute of Industrial Technology,
Ansan, 15521, Korea ({river.j, hans1024} @kitech.re.kr) * Corresponding author
[3] Dept. of Precision Manufacturing &Control R&D Group, Korea Institute of Industrial Technology,
Busan, 46056, Korea(0vincent@kitech.re.kr)

Abstract: This paper proposes a novel data augmentation method for enhancing the performance of deep-learning-based object detection in underwater environments. With deep-learning-based methods, system performance is highly dependent on the learning dataset. In extreme conditions such as underwater environments, however, it is difficult to obtain sufficient image data. To make this easier, we propose an image generation method that augments learning data by applying the special conditions of underwater environments. The proposed method generates virtual underwater images by using the optical properties of water from images taken above ground. The generated data can be used in place of underwater data, thus reducing the training effort. Experimental results show that the proposed method is effective.

Keywords: Data augmentation, Underwater object detection.

1. INTRODUCTION

Object detection is one of the most important functions of computer vision. Before the advent of deep-learning technologies, object detection algorithms were implemented based on visual features extracted with handcrafted algorithms. Many researchers, therefore, sought better visual features for improved object detection.

Deep-learning-based object detection learns feature extraction from data using a back-propagation algorithm [1]. With a sufficient amount of good data, a well-learned neural network can be acquired. However, obtaining enough appropriate data from the application domain is often difficult. In underwater environments, for example, obtaining image data is difficult because it can be accomplished only by trained divers using special equipment. Further, the divers are limited by time and temperature.

Therefore, an easier method is needed to obtain underwater learning data for use in deep-learning algorithms supporting underwater object detection. This paper proposes a method to apply above-ground images to the underwater detection algorithm. We show that the data obtained through our proposed method can replace authentic underwater data.

This paper is organized as follows. Section 2 discusses related studies. Section 3 proposes a method to solve the problem of insufficient data, and Section 4 shows that the proposed method can solve the problem.

2. RELATED WORKS

Object detection algorithms based on convolutional neural networks (CNN) [1] have been widely studied. Unlike traditional handcrafted methods [2], this type of object detection does not require characterization by humans. When data is available, the computer finds features to detect on its own through back-propagation. Typical object detection algorithms include Faster R-CNN, You Only Look Once (YOLO), and Single-Shot MultiBox Detector [3–6]. Particularly, YOLO can support real-time detection because it is very fast and accurate.

These detection algorithms require considerable data. However, collecting sufficient labeled image data is difficult under certain conditions like underwater environments. Many methods have been studied to solve these problems, including transfer learning [7]. According to Yosinski et al. [8], the lower layer of a well-learned CNN model has general parts that can be used for all data. Transfer learning is a method of using these general parts to facilitate learning of a new dataset. Domain adaptation [9], a method that improves the performance of neural networks when feature distribution is changed by factors such as background change, is a type of transfer learning.

Data augmentation [10] increases the amount of data by making changes to the labeled dataset. There are various other ways to augment data, including adding noise or converting the shapes of images. Recent studies have applied data augmentation methods using a generative adversarial neural network (GAN) [11] to domain adaptation [12, 13].

Many of these complementary data strategies can be

used for object detection in underwater environments. For example, Shang et al. used a transfer-learning method with a pre-learned neural network for the ImageNet Large-Scale Visual Recognition Competition [14] to solve the problem of detecting 12 species of fish using approximately 25,000 published data [15, 16].

Applying deep-learning-based object detection algorithms to a robot vision system for object manipulation presents different problems than those when applying it to ultrasound or fish/coral detection. Generally, there are no publicly available image data for underwater objects. Additionally, underwater environments are very difficult to access and have complex properties [17–20]. Therefore, a different approach is needed to collect a sufficient amount of underwater data.

3. PROPOSED METHOD

3.1 Underwater Data Augmentation

We propose a method of data augmentation that uses the characteristics of light in underwater environments. Fig. 1 shows the process of the proposed method. First, we modulate the red–green–blue (RGB) rate using absorption data for underwater light wavelengths [17]. Then, we apply Gaussian smoothing to the data expressing water turbidity. There are many opportunities for occlusion, so we apply random erasing [21]. Finally, we fine-tune the pre-trained darknet model [22], which is the backbone network of YOLO v2, via the transfer learning method.

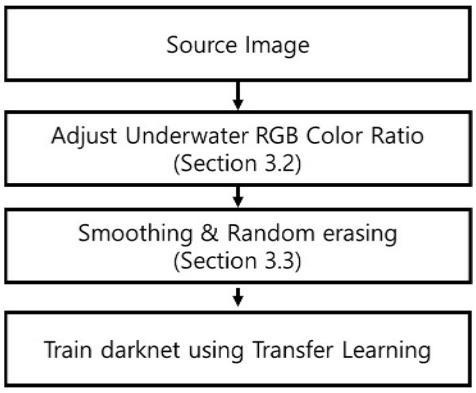

Fig. 1 Overall process of the proposed method.

3.2 Adjust RGB Color Ratio

Objects appear different in underwater environments than they do above water because of light scattering and absorption [17]. Red (600–670 nm) light is more easily absorbed than green (520–560 nm) or blue (440–480 nm). Thus, red colors cannot be seen deeper than a few meters below the surface. Fig. 2 shows the attenuation length of the light wavelengths [20].

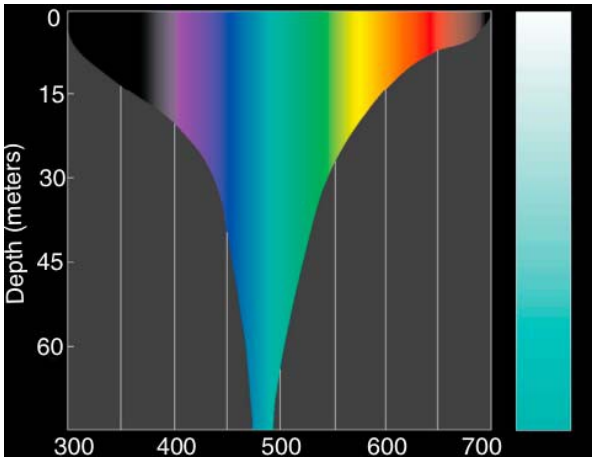

Fig. 2 Attenuation of light with depth in clear oceanic water [20].

The following Beer–Lambert Law helps calculate how much light has been absorbed at a particular wavelength,

$$I = I_0 \cdot e^{-aL}, \tag{1}$$

where a is the absorption coefficient that depends on the wavelength of light, and L is the depth from the water's surface [18].

Applying the coefficient for the RGB wavelength to Eq. (1) provides the following expressions [17]:

$$\begin{aligned} R &= R_s \cdot e^{-0.280 \cdot L}, \\ G &= G_s \cdot e^{-0.042 \cdot L}, \\ B &= B_s \cdot e^{-0.010 \cdot L}. \end{aligned} \tag{2}$$

If the above expressions are applied to the RGB color of the original image, data similar to underwater images are generated. Variable L is derived by adding the random variable, p, to the depth, d, where the operation is expected.

$$L = d + p. \tag{3}$$

Fig. 3 shows the conversion result of Eq. (2). The left figure is the original image taken above ground, and the right figure is the conversion result.

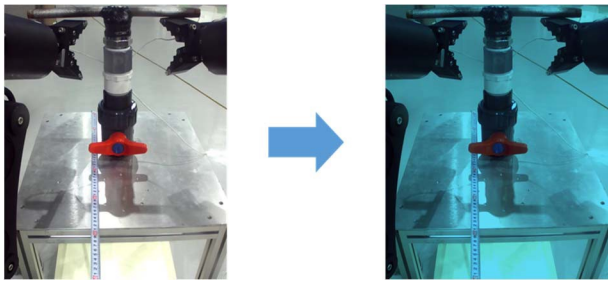

Fig. 3 Above-ground (left) and converted (right).

3.3 Adjust Smoothing and Random Erasing

Underwater environments are not clean, and suspended solids scatter the available light. For this reason, we applied Gaussian smoothing to the image.

Random erasing partially erases images using a random variable. There are many chances for occlusion in the workspace [21], but random erasing prevents failures caused by occlusion.

Fig. 4 shows the process of changing the ground data using this process. Fig. 5 compares the result of applying this method to the above-ground image to a real underwater image obtained at depth of 3 m.

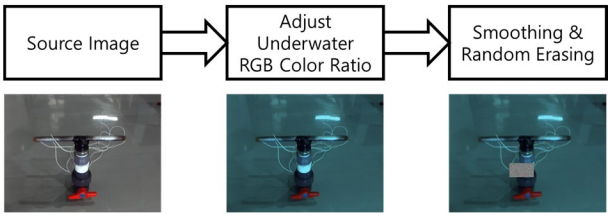

Fig. 4 Conversion process using an above-ground image.

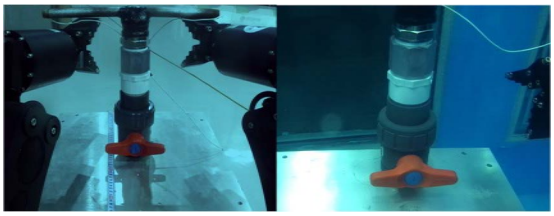

Fig. 5 The converted image and an actual underwater image from a depth of 3 m.

4. EXPERIMENT

4.1 Setup

To evaluate the proposed method, we built three image datasets: underwater images, above-ground images and augmented images. The underwater image dataset comprised images of objects at a 1.2 m depth while the above-ground image dataset comprised images taken above ground. The augmented image dataset comprised augmented above-ground images adjusted by the proposed method.

We fine-tuned the darknet neural network for each dataset. 2,000 underwater and above-ground raw data images were taken of three kinds of objects. Fig. 6 shows the three objects selected for detection testing and compares the conversion results to the real underwater data.

Neural-network learning for each dataset was evaluated using a validation dataset consisting of 600 images. To differentiate between validation and training data, we placed objects, such as profiles and an umbrella, in the background. Fig. 7 shows an example of the validation dataset.

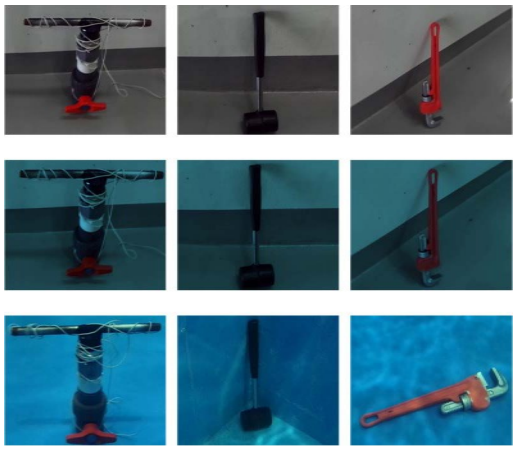

Fig. 6 Objects for detection testing (top), conversion results (middle) and actual data (bottom).

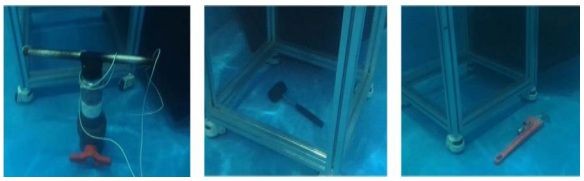

Fig. 7 Examples of validation data.

4.2 Result

Fig. 8 shows the detection results learned from each dataset. To evaluate detector performance, we used a validation image dataset containing three types of target objects and obstacles.

As shown from the detection test results, the detector learned by the above-ground dataset was less effective for object detection. The other two detectors, learned with the underwater and augmented datasets, worked well for validation of images.

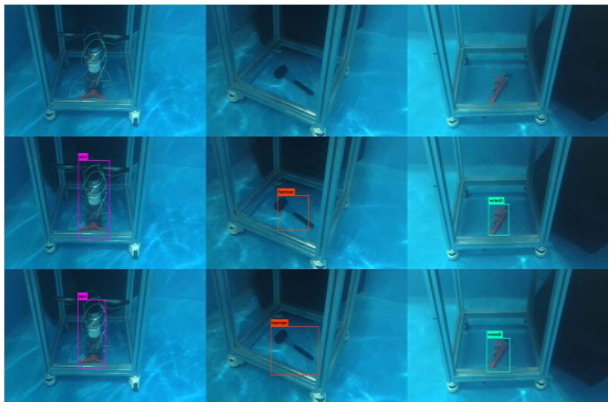

Fig. 8 Detection results of the detector learned from each dataset: the above-ground dataset (top), augmented dataset (middle) and real underwater dataset (bottom).

To analyze the experimental results, we quantitatively evaluated the detection results. We used average precision (*AP*) and perception rate (*PR*) as the

performance index of detection algorithm. Precision and recall are required to calculate the average precision (*AP*) of the detection algorithm. These are calculated as follows:

$$Precision = \frac{TP}{TP + FP}, \tag{4}$$

$$Recell = \frac{TP}{TP + FN}, \tag{5}$$

where *TP* and *FP* are true positive and false positive, respectively, and *FN* is false negative [23].

AP is the average of precisions corresponding to the recall value, and *mAP* is the mean of all *AP*. The criterion for determining whether the answer is correct is *IoU*. The *IoU* is defined as

$$IoU = \frac{area\left(B_P \cap B_{GT}\right)}{area\left(B_P \cup B_{GT}\right)}, \tag{6}$$

where B_P is the predicted bounding box of the detector and B_{GT} is the bounding box of ground truth, as shown in Fig. 9. If the *IoU* exceeds the 0.5 threshold, the detection can be considered correct.

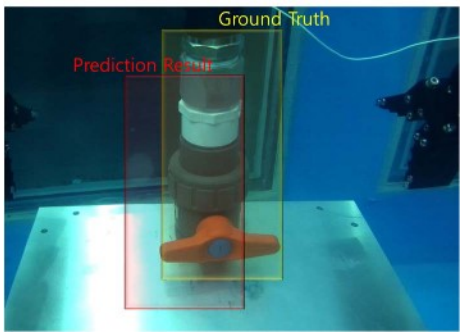

Fig. 9 An example of the bounding box for ground truth and prediction results.

Table 1 shows the evaluated performance of *AP*. Overall, the detector learned from the real underwater dataset showed the highest *AP*, while the detector using only the above-ground dataset showed the lowest *AP*.

The *mAP* of the detector learned from the augmented data was approximately 84%. It was approximately 87.4% of the *mAP* for the real underwater dataset and 158% for the above-ground only dataset. Moreover, the *AP* for the wrench was approximately 97.25%, indicating a similar performance to that on the underwater dataset. Fig. 10 shows the AP for each class as a bar graph.

Table 1 Comparison of AP among different datasets

	Above-ground	Augmented	Underwater
mAP (%)	53.10	84.00	96.14
AP_{pipe} (%)	49.07	75.04	94.22
AP_{hammer} (%)	48.73	79.70	95.68
AP_{wrench} (%)	61.51	97.25	98.53

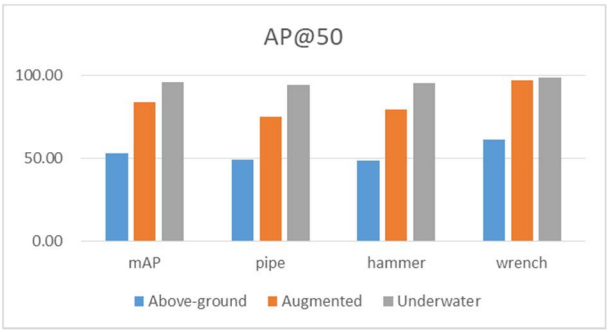

Fig. 10 Graph of *AP* for test objects.

We also used the perception rate (*PR*) as a metric for evaluating the detector recognition performance. *PR* is calculated as

$$PR = \frac{TP}{TP + FN + FP}. \tag{7}$$

Table 2 shows the evaluated detection performance for *PR*. For pipes and hammers, the underwater dataset performed higher than other datasets, but for wrenches, the augmented dataset was the highest at 90.6%. Fig. 11 shows the *PR* for each class as a bar graph.

Table 2 Comparison of PR among different datasets

	Above-ground	Augmented	Underwater
PR_{mean} (%)	35.83	69.63	80.91
PR_{pipe} (%)	32.49	54.27	79.19
PR_{hammer} (%)	35.00	63.98	87.93
PR_{wrench} (%)	40.00	90.63	75.61

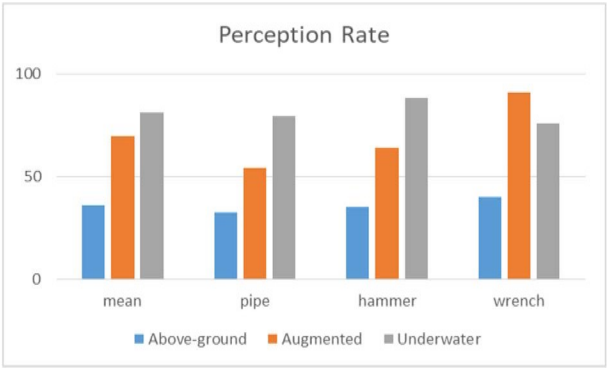

Fig. 11 Graph of PR for each class.

5. CONCLUSION

In this paper, we present a method of data augmentation specific to underwater environments, in which above-ground data are augmented using features of underwater light. To demonstrate that an actual underwater dataset can be replaced with a dataset using our data, we conducted experiments using an actual underwater dataset, an above-ground dataset, and a dataset obtained by our method. We used YOLO detectors learned from each data and compared the performances. The experiments confirmed that augmented data obtained from our method can be used instead of actual underwater data. The proposed method may therefore be useful for underwater object detection because it can be utilized without the difficulty of acquiring underwater data directly.

Optical properties were used in this study, and this method is based on human knowledge. In future research, we will study methods that apply generally to all underwater environments without the use of human knowledge.

ACKNOWLEDGEMENT

This research was financially supported by the Korea Institute of Industrial Technology (KITECH) through the In-House Research Program (JA190004).

REFERENCES

[1] Y. Lecun and Y. Bengio, *The handbook of brain theory and neural networks*, MIT Press Cambridge, MA, USA, 1995.

[2] K. A. Nugroho, "A comparison of handcrafted and deep neural network feature extraction for classifying optical coherence tomography (oct) images," *arXiv:1809.03306*.

[3] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv:1804.02767*.

[4] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," *arXiv:1612.08242*.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *arXiv:1506.01497*, 2015.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," *Proc. European Conf. Comput. Vis.*, pp. 21–37, 2016.

[7] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey on transfer learning," *J. Big Data*, pp. 3-9, 2016.

[8] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *arXiv:1411.1792v1*.

[9] G. Csurka., "Domain adaptation for visual applications: A comprehensive survey," *arXiv:1702.05374v2*.

[10] L. Perez, and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv:1712.04621*.

[11] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *arXiv:1406.2661*.

[12] S. W. Huang, C. T. Lin., S. P. Chen., Y. Y. Wu., P. H. Hsu., and S. H. Lai, "AugGAN: Cross domain adaptation with gan-based data augmentation," *Proc. of ECCV,* 2018

[13] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial Discriminative Domain Adaptation," *Proc. of CVPR*, pp. 7,167–7,176, 2017.

[14] Image-net.org/challenges/LSVRC.

[15] X. Li, M. Shang, H. Qin, and L. Chen, "Fast accurate fish detection and recognition of underwater images with Fast R-CNN," *Proc. of OCEANS,* pp. 1–5,Washington, 2015.

[16] www.imageclef.org/2014/lifeclef/fish.

[17] R. M. Pope, and E. S. Fry, "Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements," *Appl. Opt.,* Vol. 36, pp. 8,710–8,723, 1997.

[18] D. F. Swinehart, "The Beer–Lambert law," *J. Chem. Educ*, Vol. 39, No. 7, pp. 333, 1962.

[19] J. Y. Chiang., and Y. C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Proc.,* Vol, 21, No. 4, April, 2012.

[20] J. Marshall., "Vision and lack of vision in the ocean," *Current Biol.,* Vol. 27, pp. 431–510, 2017.

[21] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," *arXiv.1708.04896*.

[22] https://pjreddie.com/darknet/yolo.

[23] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) challenge," *I.J. Comput. Vis.*, vol. 88, pp. 303, 2010.