MAN-MADE OBJECT RECOGNITION FROM UNDERWATER OPTICAL IMAGES USING DEEP LEARNING AND TRANSFER LEARNING

Xian Yu, Xiangrui Xing, Han Zheng, Xueyang Fu, Yue Huang, Xinghao Ding*

Key Laboratory of Underwater Acoustic Communication
and Marine Information Technology, Ministry of Education, Xiamen University, Xiamen, Fujian, China
School of Information Science and Engineering, Xiamen University, Xiamen, Fujian, China
* yhuang2010@xmu.edu.cn

ABSTRACT

With the development of underwater optical sensors, man-made object recognition from underwater optical images has attracted wide attention. Deep learning methods have demonstrated impressive performance in object recognition tasks from natural images. However, it is difficult to collect large-scale labeled underwater optical images for training such a model. Based on the assumption that it is possible to acquire sufficient labeled in-air images, the proposed work leverages a combination of deep learning and transfer learning to develop a novel recognition system for man-made object from underwater optical images. The extracted features from the proposed network have high representative power, and demonstrate robustness in both in-air and underwater imaging modalities. Therefore, our proposed framework has the ability to recognize underwater man-made objects using only labeled in-air images. The results of experiments on simulated data demonstrate that the proposed method outperforms traditional deep learning methods in the task of underwater man-made object recognition.

***Index Terms*—** underwater optical image, man-made object recognition, deep learning, transfer learning, unsupervised domain adaptation

1. INTRODUCTION

Optical and sonar based systems are the two main imaging modalities used for underwater vision-based navigation [1, 2, 3]. In underwater imaging systems, recognition of man-made objects plays an important role for conducting research in domains such as oceanographic species identification, pipeline overhauling, mine detection, and naval studies, among others [4, 5, 6].

This work was supported in partby the National Natural Science Foundation of China underGrants 61571382,81671766, 61571005, 81671674, U1605252,61671309 in part by the Guangdong Natural Science Foundationunder Grant 2015A030313007, in part by the Fundamental ResearchFunds for the Central Universities under Grant20720160075, 20720150169.

Compared with sonar imaging, optical imaging, due to its ability to capture greater details and color, has found greater applicability in underwater object detection tasks [7]. With the development of underwater optical image sensors, man-made target recognition from underwater optical images has attracted greater attention in both oceanic engineering and image processing [4, 8, 9].

Poor image quality is one of the biggest challenges in underwater optical image analysis (Fig.1). Image quality is often low due to factors such as impurities in the water, and high water density [4]. Besides, limited visibility due to the exponential attenuation of light in deep waters also further degrades image quality [7].

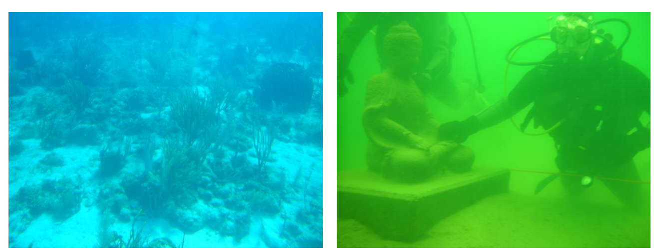

Fig. 1. Examples of underwater optical images with poor image quality.

Very few studies have been conducted in the domain of man-made target recognition from underwater optical images. In both [10] and [11], the authors built systems to identify and recognize underwater man-made objects using color information. Hou et al. [12] proposed a detection method from features based on the color and the shape of underwater man-made objects. Hussian et al. [13] proposed an underwater man-made object recognition framework which integrated a pipeline of different image processing techniques, including equalization for preprocessing, line and edge detection, and Euclidean shape prediction. In [14], the authors reported a system for detecting the presence of man-made objects from unconstrained subsea videos. They extracted object contours as stable features, and then employed a Bayesian classifier to predict the presence of a man-made object in the image.

Recently, deep learning methods, due to their strong representative power, have demonstrated impressive performances in object recognition task from natural images [15, 16]. Therefore, it is natural to consider the use of deep learning to recognize man-made object from underwater optical images. However, there are certain challenges which must be addressed in order to effectively use deep learning techniques for this task. For deep learning, one of the prerequisites is the availability of large-scaled labeled data, needed for the estimation of parameters during the training phase. Also, similar to traditional machine learning methods, deep learning assumes that the training and the testing samples follow a similar distribution [17] - that is, the imaging procedures for capturing the training and the testing samples should be the same or similar. In real-world scenarios, for underwater imaging, it is challenging to collect and label sufficient underwater man-made objects.

In this work, we assume that it is easier to acquire sufficient training samples of man-made objects from in-air images. For example, it is easy to capture sufficient multi-view images before submerging the man-made objects in water. Based on this assumption, we propose an underwater man-made object recognition framework which uses both deep learning and transfer learning. During the training phase of the proposed framework, we use a large-scale dataset of labeled in-air images of man-made objects and combine this with the unlabeled underwater man-made objects. During the testing phase, we demonstrate that our trained model is able to categorize the underwater man-made object with robustness.

The main contribution of our work is a system which can use in-air images to effectively classify man-made object from underwater optical images. This removes the need to carry out the tedious and difficult task of collecting and annotating large-scale underwater images.

2. METHODS

2.1. Underwater datasets generation

Inspired by He et al. in [18], underwater images are mostly generated based on the depth of field analysis and simulation of underwater environments. Since it can be challenging and expensive to collect depth of field information for ordinary optical acquisition devices, in this paper we introduce a new method to satisfactorily generate underwater images without the need of the extra depth information of field images.

As can be observed in Fig.1, color is the most dominant feature which appears in underwater images. Nguyen et al. [19] proposed a color transfer method based on illumination awareness and 3D gamut to manipulate the color values of source images to generate images with same appearances.

However, only relying on color transfer cannot realistically simulate the underwater environment. Therefore, based on [20], we also apply turbidity simulation on top of color transfer to obtain a better representation. The resultant signal is therefore composed of two components, the first term is direct transmission:

$$D = I_{color} e^{-\eta z}, \tag{1}$$

where I_{color} is the image we obtained through color transfer, η is the the coefficient of diffusion attenuation obtained from a given real underwater patch, and z represents the adjustable distance between I_{color} and the reference underwater image, with a higher value of z representing a higher turbidity.

The second term in the resultant signal is backscattering:

$$B = B_{\infty}(1 - e^{-\eta z}), \tag{2}$$

where B_{∞} is the backscatter in the line of sight (LOS) which extends to infinity in water.

The resultant underwater image is generated by combining the two terms as follows:

$$I_{underwater} = D + B - D \cdot B, \tag{3}$$

and $\cdot$ represents the element-wise multiplication.

2.2. Framework for underwater man-made object recognition

Fig.2 represents the flowchart of our proposed framework. We employ AlexNet, which is a CNN based deep learning implementation, as the base model, in our proposed framework [21]. Our implementation consists of five convolutional layers (conv), and three fully connected layers (fc). A rectified linear unit (ReLU) is applied after the pooling operation on the conv1, conv2 and conv5 layers. The classifiers are implemented by the fully connected layers at the end of the network. The feature vector generated by the last fully connected layer is processed by the soft-max function, while the vector of probabilities represents the final prediction results of the categories.

The maximum mean distance (MMD), a distance metric feature, is applied to both the fc7 and fc8 layers of the neural network as the regularization and the transfer learning element of our proposed framework. This minimizes the distribution of the data from the different imaging procedures - in-air or underwater. According to the theory of transfer learning, the labeled in-air images are assigned as the source domain, while the unlabeled underwater images are assigned as the target domain [22]. Therefore, the MMD can be written in its square form using kernel operations:

$$\begin{aligned} D_k^2(p,q) = & E_{x_p^s x_q^s} k(x_p^s, x_q^s) + E_{x_p^t x_q^t} k(x_p^t, x_q^t) \\ & - 2E_{x_p^s x_p^t} k(x_p^s, x_p^t), \end{aligned} \tag{4}$$

where E denotes the expectation, x_p^s and x_q^s are two samples from the source domain, while x_p^t and x_q^t are two samples from the target domain; and k is the Gaussian kernel

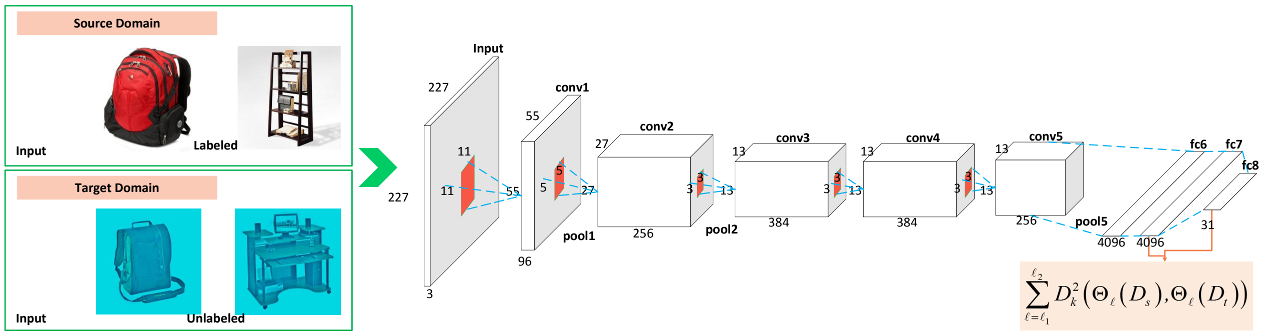

Fig. 2. The training procedure of the proposed framework, where both labeled in-air images and unlabeled underwater images are employed to train the network. The MMD feature metric is added in the last two layers for regularization. $Conv$ denotes the convolutional layer, and fc denotes the fully connected layer.

function defined by $k(x_i, x_j) = e^{-\|x_i - x_j\|^2/\gamma}$. We denote $D_s = \{x_i^s, y_i^s\}_{i=1}^{N_s}$ as the set of N_s labeled samples from the source domain, and $D_t = \{x_j^t\}_{j=1}^{N_t}$ as the set of N_t unlabeled samples from the target domain; x_i^s represents the i_{th} sample with y_i^s as the associated label in the source domain; and x_j^t represents the j_{th} sample in the target domain. Then the objective function can be defined as:

$$min \frac{1}{N_s} \sum_{i=1}^{N_s} J(\Theta(x_i^s), y_i^s) + \lambda \sum_{\ell=\ell_1}^{\ell_2} D_k^2(\Theta_\ell(D_s), \Theta_\ell(D_t)), \tag{5}$$

where the first term J is a common cross-entropy loss function, which is consistent with the corresponding part in AlexNet [21]; Θ represents all parameters in CNN model, and $\Theta(x_i^s)$ denotes the conditional probability of assigning sample x_i^s to label y_i^s. Since, we do not have any information regarding the labels in the target domain, in the function J, both x_i^s and y_i^s are obtained from the source domain. Further, $\Theta_\ell(D_s)$ and $\Theta_\ell(D_t)$ denote outputs of the ℓ_{th} layer of the source and the target domains respectively. The ℓ_1 and ℓ_2 terms refer to the fc7 and fc8 respectively in our setting. We set $\lambda(\lambda > 0)$ as the hyper parameter used to provide a trade-off for the loss function. Therefore, in our setting, the objective function can take advantage of both deep learning and transfer learning methods.

During the testing phase, the underwater man-made object images are directly predicted by the trained network.

3. EXPERIMENTS

3.1. Datasets descriptions

The Amazon dataset is used as the original in-air man-made object dataset. The dataset consists of 2817 images of man-made objects downloaded from *amazon.com*. There are 31 categories, with each category containing between 36 to 100 images. While previous works of research have mainly used objects with regular shapes and sizes, the objects in the Amazon dataset are of irregular shapes captured from different views [23].

The proposed work includes three experiments that demonstrate our contributions. The images in each category in the Amazon dataset and simulated underwater datasets are equally divided into two parts: part 1 and part 2 with no overlap between them. In the first experiment, both the training and the testing data are taken from the underwater imaging system. Thus, the training and the testing data are simulated underwater images have the same turbidity values, which are generated from images in part 1 and part 2 respectively. The experiment is set up to validate the performance of AlexNet, when the training and the testing data are generated using the same imaging system.

The second experiment is designed to evaluate the performance of AlexNet when the training data contains both labeled in-air images from the source domain and unlabeled simulated underwater images from the target domain from part 1, and the testing data only contains unlabeled simulated underwater images from part 2. Experiment 3 is set up with similar data and objectives as experiment 2. It is set up to validate the performance of the proposed framework while using transfer learning along with the traditional CNN model.

The simulated underwater images used in the three experiments are generated from the in-air images through a series of steps as follows. The top-left red-box in Fig.3 indicates a sample from the original Amazon dataset, and I and II denote two real underwater optical images used as reference images. As shown in columns A and B of Fig.3, based on the works described in [19, 20], we generate three simulated underwater datasets with three different values of turbidity for each reference image by adjusting turbidity factor z in Eq.3. The value of turbidity is increased from the top to the bottom of Fig.3, with a larger value of z denoting a higher turbidity. We denote the simulated underwater datasets with different turbidity and reference images as A_1, A_2, A_3, B_1, B_2 and B_3 respectively.

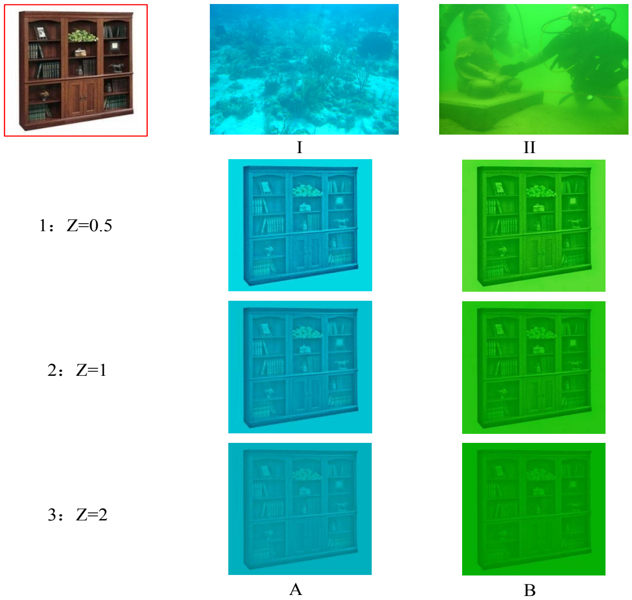

Fig. 3. Examples of underwater optical datasets.

3.2. Implementation details

In our proposed implementation, the basic server settings are: a 56 Intel(R)Xeon(R) CPU E5-2683 V3@ 2.00GHz, with 64G RAM and a NVIDIA GeForce 1080 GPU. All the images fed into the neural network are resized to the same size of 227×227 pixels. The proposed network is pre-trained on ImageNet [21, 24], and then fine-tuned with our own data.

3.3. Experimental results

As shown in Fig.4, first, we compare the accuracies of the three experiments for datasets with different turbidities and reference settings. The blue, yellow and green bars denote the recognition results of the first, the second, and the third experiments respectively. With an average value of 55.70%, the AlexNet in the first experiment achieves the highest recognition accuracy among all the three experiments. This is because in the first experiment, both the training and the testing data are from the same domain of underwater images. However, since in the second and the third experiments, the training data and the testing data are generated using different imaging systems, the performance of AlexNet in these experiments decreases dramatically. For the second experiment, AlexNet has an average accuracy of 17.33%. However, from Fig.4, we can observe that our proposed framework significantly outperforms AlexNet. The average value of accuracy for the third experiment is 38.50%. This can be explained that the proposed framework has the ability to transfer the knowledge learned from the source domain to the target domain, that is, from the in-air images to underwater images.

For a more specific comparison, we also calculate the ac-

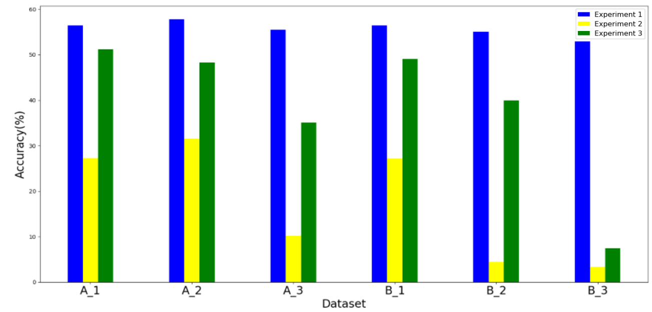

Fig. 4. Comparison of the average accuracy of each simulated underwater dataset.

curacies from the 31 categories of the dataset A_1 for the three experiments. As shown in Fig.5, this dataset has the best performance across all the categories in experiment 1 and the worst performance in experiment 2. The accuracy of the dataset in the third experiment is slightly worse than that in experiment 1. The red curve indicates the number of training data per category. We observe that the accuracies of all three experiments decrease for smaller sizes of training data, for example, for categories such as bottle, trash can, etc.

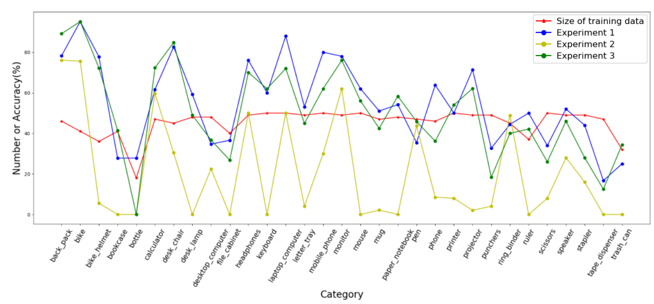

Fig. 5. The accuracies of three experiments on 31 categories with dataset A_1 . The size of the training data (red) for each category is also plotted with the accuracies in the figure.

4. CONCLUSIONS

This work presents a framework for recognizing underwater man-made objects from optical images. The work is based on the assumption that labeled in-air images of man-made objects are easy to acquire. By introducing transfer learning to a CNN model, the proposed method can simultaneously extract features that are representative as well as robust across different imaging systems. This allows us to avoid having to explicitly collect and annotate underwater images for training the model. The recognition performances of our proposed algorithm denote that the framework can be considered as an effective basic deep learning tool for optical image analysis in underwater vision-based systems.

5. REFERENCES

[1] Y. Lee, J. Choi, N. Y. Ko, and H. T. Choi, "Probability-based recognition framework for underwater landmarks using sonar images.," *Sensors*, vol. 17, no. 9, 2017.

[2] Natlia Hurts Vilarnau, "Forward-looking sonar mosaicing for underwater environments," *Universitat De Girona*, 2014.

[3] Xiaoou Tang and W. Kenneth Stewart, "Optical and sonar image classification: Wavelet packet transform vs fourier transform," *Computer Vision & Image Understanding*, vol. 79, no. 1, pp. 25–46, 2000.

[4] Jules S. Jaffe, "Underwater optical imaging: The past, the present, and the prospects," *IEEE Journal of Oceanic Engineering*, vol. 40, no. 3, pp. 683–700, 2015.

[5] Yoav Y Schechner and Nir Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 3, pp. 570–587, 2006.

[6] Yujie Li, Huimin Lu, Jianru Li, Xin Li, Yun Li, and Seiichi Serikawa, "Underwater image de-scattering and classification by deep neural network," *Computers & Electrical Engineering*, vol. 54, pp. 68–77, 2016.

[7] K. Srividhya and M. M. Ramya, "Accurate object recognition in the underwater images using learning algorithms and texture features," *Multimedia Tools & Applications*, pp. 1–17, 2017.

[8] Huimin Lu, Yujie Li, Yudong Zhang, Min Chen, Seiichi Serikawa, and Hyoungseop Kim, "Underwater optical image processing: a comprehensive review," *Mobile Networks & Applications*, pp. 1–8, 2017.

[9] Pooria Pakrooh, Louis L. Scharf, and Mahmood R. Azimi-Sadjadi, "Underwater target classification using a pose-invariant matched manifold classifier," in *IEEE International Workshop on Machine Learning for Signal Processing*, 2016, pp. 1–5.

[10] Isabelle Quidu and Luc Jaulin, "Color-based underwater object recognition using water light attenuation," *Intelligent Service Robotics*, vol. 5, no. 2, pp. 109–118, 2012.

[11] Stphane Bazeille, Isabelle Quidu, and Luc Jaulin, "Identification of underwater man-made object using a colour criterion," 2007.

[12] Hou, "Underwater man-made object recognition on the basis of color and shape features," *Journal of Coastal Research*, vol. 32, no. 5, pp. 1135–1141, 2016.

[13] Syed Safdar Hussain and Syed Sajjad Haider Zaidi, "Underwater man-made object prediction using line detection technique," in *International Conference on Electronics, Computers and Artificial Intelligence*, 2015, pp. 1–6.

[14] Adriana Olmos and Emanuele Trucco, "Detecting man-made objects in unconstrained subsea videos," in *British Machine Vision Conference*, 2002, pp. 517–526.

[15] G. E. Hinton, S Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527, 2006.

[16] P Sermanet, S Chintala, and Y Lecun, "Convolutional neural networks applied to house numbers digit classification," in *International Conference on Pattern Recognition*, 2012, pp. 3288–3291.

[17] Zhiyuan Chen and Bing Liu, "Lifelong machine learning," vol. 10, no. 3, pp. 1–145, 2016.

[18] Kaiming He, Jian Sun, and Xiaoou Tang, "Single image haze removal using dark channel prior," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 1956–1963.

[19] R. M. H. Nguyen, S. J. Kim, and M. S. Brown, *Illuminant Aware Gamut-Based Color Transfer*, The Eurographs Association & John Wiley & Sons, Ltd., 2014.

[20] Yoav Y. Schechner and Nir Karpel, "Clear underwater vision," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, pp. I–536–I–543 Vol.1.

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[22] Sinno Jialin Pan and Qiang Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge & Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[23] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell, "Adapting visual category models to new domains," in *European Conference on Computer Vision*, 2010, pp. 213–226.

[24] Zhen Dong, Yuwei Wu, Mingtao Pei, and Yunde Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247–2256, 2015.