Underwater Live Fish Recognition by Deep Learning

Abdelouahid Ben Tamou[1,2], Abdesslam Benzinou[1(✉)], Kamal Nasreddine[1], and Lahoucine Ballihi[2]

[1] Univ Bretagne Loire, ENIB, UMR CNRS 6285 LabSTICC, 29238 Brest, France
{bentamou,benzinou,nasreddine}@enib.fr
[2] LRIT-CNRST URAC 29, Mohammed V University In Rabat, FSR, Rabat, Morocco
ballihi@fsr.ac.ma

Abstract. Recently, underwater videos have gained great interest by marine ecologists for studying fish populations. Actually, this technique produces large amount of visual data and does not affect fish behavior. However, visual processing and analyzing of the recorded data can be subjective, time consuming and costly. We propose in this paper to use the convolutional neural network AlexNet with transfer learning for automatic fish species classification. We extract features from foreground fish images of the available underwater dataset using the pretrained AlexNet network either with or without fine-tunig. For classification, we use a linear SVM classifier. The experiment results demonstrate the effectiveness of the proposed approach on the Fish Recognition Ground-Truth dataset. We achieve an accuracy of 99.45%.

Keywords: Deep learning · Transfer learning
Convolutional neural network · Pretrained model · AlexNet
Fish recognition

1 Introduction

In the last few years, underwater video cameras are extensively used in scientific, industrial and military fields for exploring and studying underwater environments. Marine biologists are interested in using underwater video analysis to study fish populations as species richness and size measurement [1–5], abundance [6] or animal behavior [1]. Automatic processing is an advantage compared to manual processing which is relatively off putting task, subjective, time consuming and costly. Automatic fish classification can be divided into two parts. (1) Fish detection which aims to detect and separate the subject from the background. (2) Fish recognition which aims to identify the species of the detected fish. The underwater environment presents a lot of difficulties and poses great challenges for computer vision. The luminosity changes frequently, the visibility is limited and the background can change rapidly due to moving aquatic plants. There are some attempts to improve image contrast and resolution for underwater images [7,8]. In addition, in fish recognition task, the fish can move in three

A. Mansouri et al. (Eds.): ICISP 2018, LNCS 10884, pp. 275–283, 2018.
https://doi.org/10.1007/978-3-319-94211-7_30

dimensions, it can also hide behind rocks and algae. We also encounter the problems of fish overlapping and of the similarity in shape and patterns among fish of different species. In this paper, we will focus on fish recognition in underwater video images.

Convolutional neural networks (ConvNets) [9] consist of L learned layers. The first layer is the input layer and represents a raw image. The hidden layers typically consist of convolutional layers, pooling layers, normalization layers and fully connected layers. The output layer consists of N-dimensional vector where N is the number of classes, this layer uses Softmax function to predict a single class of N mutually exclusive classes. ConvNets are trained using a standard error-backpropagation algorithm.

Li et al. [2] applied Fast R-CNN (Regions with Convolutional Neural Networks) on underwater images to detect and recognize fish species. They achieved an accuracy of 81.4% on LifeCLEF 2014 dataset that contains 24277 fish images of 12 species. Choi [11] participated in LifeCLEF 2015 task for detecting and identifying fish in underwater videos and achieved the best performance of 81% in this task [12]. He detected fish by using background subtraction and a selective search strategy [13]. Then, he used the GoogleNet [14] based on convolutional neural networks to classify fish species. Qin et al. [3] used convolutional neural networks on the Fish Recognition Ground-Truth dataset consisting of a total of 27370 fish images of 23 species and they reached an accuracy of 98.57%. Qin et al. [4] used also deep architecture to extract the features of fish images. In their architecture, two convolutional layers use Principal Component Analysis (PCA), the non-linear layer uses a binary hashing and the feature pooling layer uses a block-wise histograms. Then, information invariant to large poses are extracted by using spatial pyramid pooling (SPP). Finally, they use a linear SVM classifier for the classification. Despite they introduced hand-crafted layers they have improved marginally the accuracy by 0.07%. Sun et al. [5] extracted features from underwater images by applying two deep learning architectures, PCANet [15] and Network In Network (NIN) [16]. For classification, they used a linear SVM classifier. They tested their model on a database of 15 species and obtained an accuracy of 69.84% with the NIN architecture and 77.27% with the PCANet architecture. Salman et al. [17] created a deep architecture of three convolutional layers to extract features, then they combined the features from multiple layers of the network to feed standard classifiers like SVM and KNN. They achieved an accuracy of 96.75% on test set of 7500 fish images issue from LifeCLEF 2015 Fish dataset.

Learning deep architectures from scratch necessitates a large dataset because of the huge number of weights to be trained. Available underwater datasets are of small size for learning ConvNets for underwater fish recognition. To overcome this problem, we introduce transfer learning framework [18] to train ConvNets from pretrained networks that could be trained on large datasets. AlexNet [10], GoogleNet [14], VGG [19] and ResNet [20] are some examples of pretrained models that have emerged in this field last few years. In this paper, we propose to transfer the learned weights from AlexNet model to a deep ConvNet for

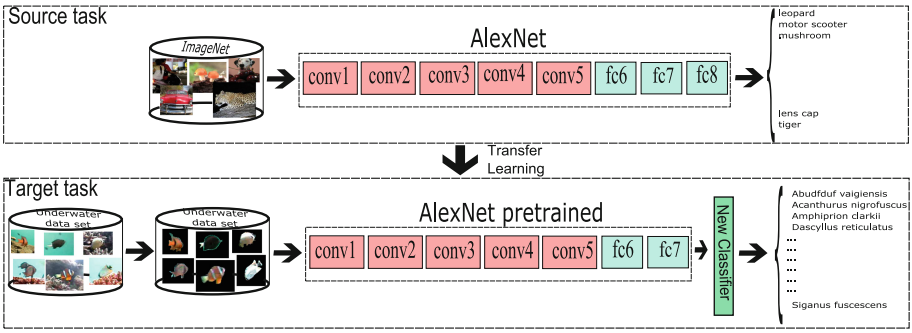

Fig. 1. The proposed approach based on the pretrained AlexNet network with transfer learning technique for fish recognition.

fish recognition in the open sea. We extract the fish features from images of the underwater dataset before and after fine-tuning the model and classify the input images with a linear SVM classifier on the extracted features. We choose AlexNet because this model needs less resources, is faster and has a simple architecture than others networks like GoogleNet (22 layers deep) and VGG (at least 16 convolutional layers) that make fine-tuning the transferred weights difficult especially with limited training data.

The paper is organized as follows: in the next Sect. 2, we describe the proposed approach for live fish recognition based on pretrained model. Then, the Sect. 3 details the experimental scheme and we give a comparative study evaluating the proposed approach on the Fish Recognition Ground-Truth dataset (c.f. Fig. 2). Finally, conclusions and perspectives are given in Sect. 4.

2 Proposed Approach

Available underwater datasets for fish recognition are too small for training deep ConvNets from scratch with random initialization. Moreover, deep learning requires immense resources of memories and processors. To overcome the difficulties imposed by limited training data, we use trained weights of AlexNet to extract fish features from images by removing some layers of the model and then using the rest of the network as a fixed feature extractor for our data. In order to demonstrate the effectiveness of fine-tuning approach, we extract features before and after fine-tuning. Fine-tuning algorithm consists of retraining the classifier on top of the network on the underwater image set and fine-tune the weights of the AlexNet via back-propagation. Finally, we will propose three schemes for classification. These schemes will be detailed below.

2.1 Architecture of AlexNet

As shown in the Fig. 1, AlexNet [10] has five convolutional layers. The number of filters and their size in these layers are 96 filters of size $11 \times 11 \times 3$, 256 filters of

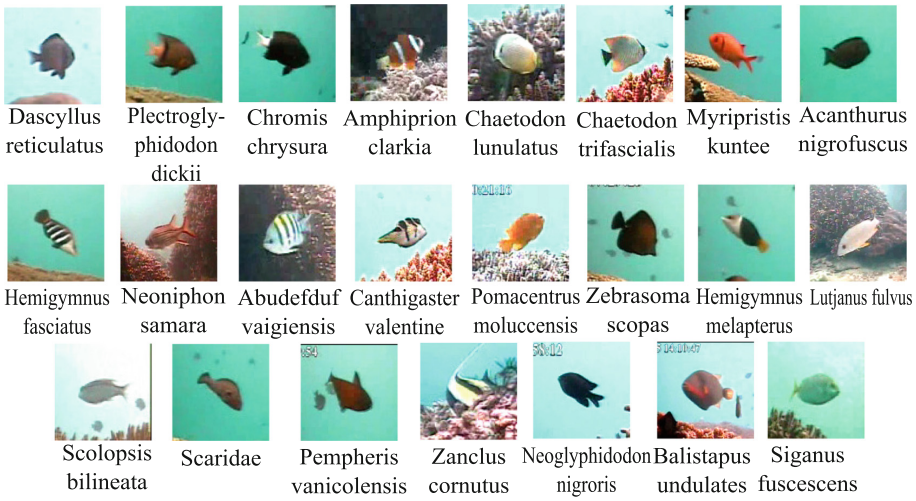

Fig. 2. Sample images of 23 fish species in Fish Recognition Ground-Truth dataset.

size $5 \times 5 \times 48$, 384 filters of size $3 \times 3 \times 256$, 384 filters of size $3 \times 3 \times 192$ and 256 filters of size $3 \times 3 \times 192$ respectively. It has also three fully-connected layers with 4096 neurons in the two first layers and the last one has 1000 neurons. AlexNet has been trained over 1.2 million images from the ImageNet dataset [21]. It can classify images into 1000 categories of objects (such as keyboard, mouse, coffee cup, pen and many animals).

2.2 Input Images

First, we eliminate the background of images using the fish masks given in the dataset (c.f. Fig. 3). These masks are generated by Qin et al. [22] who proposed a foreground extraction method for underwater videos based on sparse and low-rank matrix decomposition. Then, we use foreground fish images as input training images after a resizing to the same size $227 \times 227 \times 3$.

2.3 Feature Extraction and Classification

We first employ the AlexNet model without any fine-tuning to extract learned features by removing the output layer fc8 and using the value outputs of fc7's layer as feature descriptor for each fish image, we denote this scheme by *Alex-SVM*.

It is recommended to fine-tune the model, especially, when data similarity is very low between the original data and the new data. As the dataset used in this work is totally different from ImageNet, we will retrain the AlexNet model on the underwater dataset. We initialize a new fully-connected layer to replace the old one fc8 with a random values of 23 outputs corresponding to the 23 species in

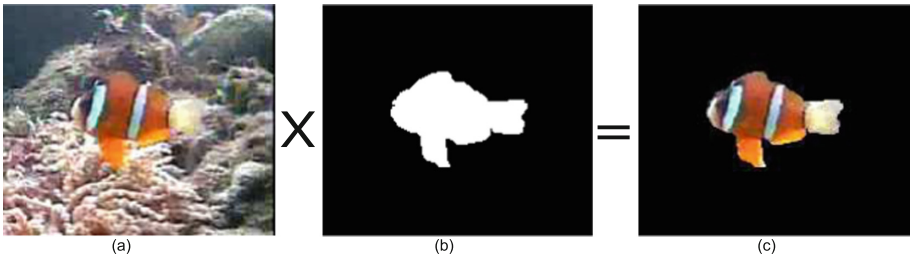

Fig. 3. Example of foreground extraction. (a): original image. (b): fish mask. (c): fish foreground.

the considered dataset. Then, we retrain only this layer and we keep the weights of the lower layers. This is because the features captured in the lower layers are universal like edges and curves that are also pertinent to our task. We get a new AlexNet model retrained, we denote this scheme by *Alex-FT-Soft*.

Finally, we use the retrained AlexNet model to re-extract fish features from images as in the first one. We denote this last scheme by *Alex-FT-SVM*.

3 Experimental Results

The performances evaluation of the proposed system is carried out on the Fish Recognition Ground-Truth dataset[1]. We implement the algorithms in Matlab and use MatConvNet[2] library for training the deep ConvNet.

Table 1. The fish species distribution in the Fish Recognition Ground-Truth dataset.

Species	Samples	Species	Samples
Dascyllus reticulatus	12112	Pomacentrus moluccensis	181
Plectroglyphidodon dickii	2683	Zebrasoma scopas	90
Chromis chrysura	3593	Hemigymnus melapterus	42
Amphiprion clarkia	4049	Lutjanus fulvus	206
Chaetodon lunulatus	2534	Scolopsis bilineata	49
Chaetodon trifascialis	190	Scaridae	56
Myripristis kuntee	450	Pempheris vanicolensis	29
Acanthurus nigrofuscus	218	Zanclus cornutus	21
Hemigymnus fasciatus	241	Neoglyphidodon nigroris	16
Neoniphon samara	299	Balistapus undulates	41
Abudefduf vaigiensis	98	Siganus fuscescens	25
Canthigaster valentine	147		

[1] http://groups.inf.ed.ac.uk/f4k/GROUNDTRUTH/RECOG/.
[2] http://www.vlfeat.org/matconvnet/.

The Fish Recognition Ground-Truth dataset is an underwater live fish image dataset acquired from a live video dataset made by the European project Fish4-Knowledge[3] (F4K) [23]. The dataset contains 27370 fish images and their fish masks of 23 different species. The fish species are manually labeled by following instructions from marine biologists. The dataset is imbalanced in the number of different fish species where the number of the most frequent species is about 1000 times more than the least one. The Fig. 2 shows examples of the 23 fish species and Table 1 shows the distribution of the fish species in the dataset.

We use 7-Fold Cross-Validation in order to estimate the performance of the proposed approach [4]. The results of classification using the three proposed schemes are given in the Table 2 in terms of accuracy and precision.

Table 2. Comparison of fish recognition performances of various methods on the Fish Recognition Ground-Truth dataset.

Species	*Alex-SVM*	Alex-FT-Soft	Alex-FT-SVM	DeepFish-SVM-aug [4]	DeepFish-SVM-aug-scale [4]	Deep-CNN [3]
Dascyllus reticulatus	99.01	98.54	99.73	99.31	99.25	-
Plectroglyphidodon dickii	98.02	96.12	99.37	97.13	97.39	-
Chromis chrysura	97.80	95.24	99.47	98.64	98.24	-
Amphiprion clarkia	99.75	99.39	99.90	100	100	-
Chaetodon lunulatus	99.80	98.97	99.84	100	100	-
Chaetodon trifascialis	93.69	75.79	98.41	92.59	96.30	-
Myripristis kuntee	97.99	92.43	99.11	98.44	100	-
Acanthurus nigrofuscus	77.98	60.04	86.23	64.52	67.74	-
Hemigymnus fasciatus	98.74	90.88	99.59	100	100	-
Neoniphon samara	99.67	99.00	99.34	100	100	-
Abudefduf vaigiensis	97.96	84.69	97.96	92.86	92.86	-
Canthigaster valentine	96.60	85.71	95.92	95.24	95.24	-
Pomacentrus moluccensis	99.45	92.22	100	100	100	-
Zebrasoma scopas	85.62	48.90	88.92	84.62	84.62	-
Hemigymnus melapterus	88.10	35.71	92.86	66.67	66.67	-
Lutjanus fulvus	96.60	89.31	99.51	96.55	96.55	-
Scolopsis bilineata	97.96	79.59	100	85.71	85.71	-
Scaridae	98.21	76.79	96.43	100	100	-
Pempheris vanicolensis	96.43	82.86	100	100	100	-
Zanclus cornutus	80.95	19.05	90.48	66.67	100	-
Neoglyphidodon nigroris	59.52	33.33	57.14	50	50	-
Balistapus undulates	95.24	46.19	97.62	83.33	83.33	-
Siganus fuscescens	91.67	67.86	95.24	100	100	-
Average Precision	93.34	76.03	**95.35**	90.10	91.91	-
Accuracy	98.57	96.61	**99.45**	98.59	98.64	98.57

[3] www.fish4knowledge.eu.

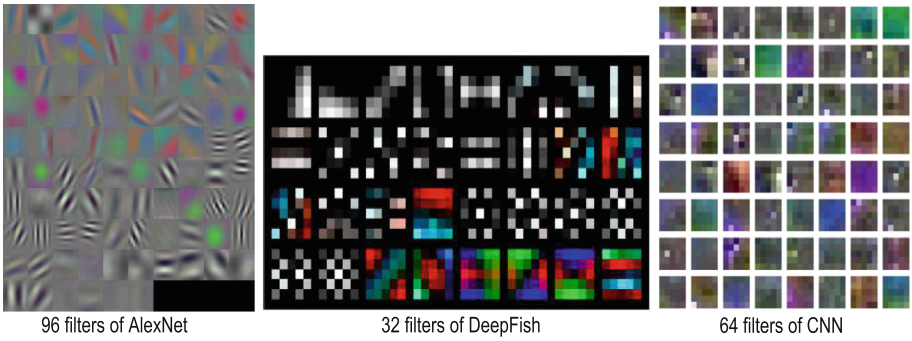

Fig. 4. Visualization of the first convolutional layer filters of AlexNet [10] (96 filters of size $11 \times 11 \times 3$), DeepFish [4] (32 filters of size $5 \times 5 \times 3$) and DeepCNN [3] (64 out of 72 filters of size $5 \times 5 \times 3$) (Color figure online)

As shown in Table 2, proposed schemes are efficient and give promising results. *Alex-FT-SVM* performs better than *Alex-SVM*, this is because in *Alex-SVM* the higher layers of the network are more precise to the details of the objects contained in ImageNet dataset. However, after fine-tuning, these layers become more precise to the details of the fish species contained in our dataset, therefore, we achieve the best accuracy of **99.45%**. We conclude that the fine-tuning improves the performance of the system. We can also see that the Softmax classifier is less robust than SVM classifier, especially, for species with fewer samples like '*Zebrasoma scopas*', '*Hemigymnus melapterus*', '*Scolopsis bilineata*', '*Scaridae*', '*Pempheris vanicolensis*', '*Zanclus cornutus*', '*Balistapus undulates*', '*Neoglyphidodon nigroris*' and '*Siganus fuscescens*'.

Table 2 shows also the comparison of our approach with state-of-the-art methods on the Fish Recognition Ground-Truth dataset. In this work, we want to test a purely deep learning-based system without any layers that contain hand-crafted methods like PCA, block-wise histograms, spatial pyramid pooling (SSP) [4], nor any method to improve the performance like data augmentation or scale images as in DeepFish-SVM-aug and DeepFish-SVM-aug-scale [4]. We can observe that *Alex-FT-SVM* outperforms the state-of-the-art methods even those with hand-crafted layers. In Deep-CNN [3], the authors created a ConvNet with three convolutional layers and trained the network from scratch. As we can see the network trained with transfer learning gives better results than networks trained from scratch.

The Fig. 4 visualizes weights of the 96 filters, 32 filters and 64 filters in the first convolutional layer of the adopted AlexNet, DeepFish and DeppCNN respectively. The color filters extract low-frequency features and the grayscale filters extract high-frequency features. As we can see, the AlexNet has more filters than DeepFish and DeepCNN which means more variety of selective filters extracting more features at different scales and different orientations. We note that additional filters are all very nice, smooth, well-formed and without noisy patterns that make AlexNet richer by feature representations.

4 Conclusion and Future Work

Underwater live fish recognition will become a necessary tool to assist marine ecologists in studying the biodiversity in underwater areas because traditional techniques are destructive, affect fish behavior, demand time and labor costs. Proposed convolutional neural networks for fish identification require large datasets due to the huge number of parameters to be trained, especially in deeper networks. Transfer learning is a solution for training this kind of networks by using pretrained models which have been trained on a large dataset.

In this paper, we proposed to transfer learned weights of the pretrained network AlexNet which has been trained on ImageNet dataset to recognize fish species in underwater images. We have extracted fish features from images by AlexNet to feed a linear SVM before and after fine-tuning.

Experiments on the Fish Recognition Ground-Truth dataset demonstrate that the proposed approach outperforms various other approaches employed for fish species identification.

In future work, we plan to extend the method proposed here for larger underwater video datasets with more classes.

Acknowledgments. The authors would like to thank the Région Bretagne for financial support.

References

1. Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y.H.J., Fisher, R.B., Nadarajan, G.: Automatic fish classification for underwater species behavior understanding. In: Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams, pp. 45–50. ACM, October 2010
2. Li, X., Shang, M., Qin, H., Chen, L.: Fast accurate fish detection and recognition of underwater images with fast R-CNN. In: OCEANS 2015 MTS/IEEE Washington, pp. 1–5. IEEE, October 2015
3. Qin, H., Li, X., Yang, Z., Shang, M.: When underwater imagery analysis meets deep learning: a solution at the age of big visual data. In: OCEANS 2015 MTS/IEEE Washington, pp. 1–5. IEEE, October 2015
4. Qin, H., Li, X., Liang, J., Peng, Y., Zhang, C.: Deepfish: accurate underwater live fish recognition with a deep architecture. Neurocomputing **187**, 49–58 (2015)
5. Sun, X., Shi, J., Dong, J., Wang, X.: Fish recognition from low-resolution underwater images. In: International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 471–476. IEEE, October 2016
6. Chuang, M.C., Hwang, J.N., Williams, K.: Automatic fish segmentation and recognition for trawl-based cameras. In: Computer Vision and Pattern Recognition in Environmental Informatics, pp. 79–106. IGI Global (2016)
7. Schettini, R., Corchs, S.: Underwater image processing: state of the art of restoration and image enhancement methods. EURASIP J. Adv. Signal Process. **2010**(1), 746052 (2010)

8. Chambah, M., Semani, D., Renouf, A., Courtellemont, P., Rizzi, A.: Underwater color constancy: enhancement of automatic live fish recognition. In: Color Imaging IX: Processing, Hardcopy, and Applications. International Society for Optics and Photonics, Vol. 5293, pp. 157–169, December 2003
9. LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., Jackel, L.D.: Handwritten digit recognition with a back-propagation network. In: Advances in Neural Information Processing Systems, pp. 396–404 (1990)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
11. Choi, S: Fish identification in underwater video with deep convolutional neural network: SNUMedinfo at LifeCLEF fish task 2015. In: CLEF (2015). Working Notes
12. Joly, A., et al.: LifeCLEF 2015: multimedia life species identification challenges. In: Mothe, J., et al. (eds.) CLEF 2015. LNCS, vol. 9283, pp. 462–483. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24027-5_46
13. Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. Int. J. Comput. Vis. **104**(2), 154–171 (2013)
14. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
15. Chan, T.H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y.: PCANet: a simple deep learning baseline for image classification? IEEE Trans. Image Process. **24**(12), 5017–5032 (2015)
16. Lin, M., Chen, Q., Yan, S.: Network in network (2013). arXiv preprint arXiv:1312.4400
17. Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., Harvey, E.: Fish species classification in unconstrained underwater environments based on deep learning. Limnol. Oceanogr. Methods **14**(9), 570–585 (2016)
18. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). arXiv preprint arXiv:1409.1556
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
21. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition CVPR 2009, pp. 248–255. IEEE, June 2009
22. Qin, H., Peng, Y., Li, X.: Foreground extraction of underwater videos via sparse and low-rank matrix decomposition. In: 2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI), pp. 65–72. IEEE, August 2014
23. Boom, B.J., Huang, P.X., He, J., Fisher, R.B.: Supporting ground-truth annotation of image datasets using clustering. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 1542–1545. IEEE, November 2012