



Research Article


# An open-set framework for underwater image classification using autoencoders



Azim Akhtarshenas<sup>1</sup>  · Ramin Toosi<sup>1</sup>

Received: 3 November 2021 / Accepted: 21 June 2022

Published online: 19 July 2022

© The Author(s) 2022  OPEN

## Abstract

In this paper, we mainly intend to address the underwater image classification problem in an open-set scenario. Image classification algorithms have been mostly provided with a small set of species, while there exist lots of species not available to the algorithms or even unknown to ourselves. Thus, we deal with an open-set problem and extremely high false alarm rate in real scenarios, especially in the case of unseen species. Motivated by these challenges, our proposed scheme aims to prevent the unseen species from going to the classifier section. To this end, we introduce a new framework based on convolutional neural networks (CNNs) that automatically identifies various species of fishes and then classifies them into certain classes using a novel technique. In the proposed method, an autoencoder is employed to distinguish between seen and unseen species. To clarify, the autoencoder is trained to reconstruct the available species with high accuracy and filter out species that are not in our training set. In the following, a classifier based on EfficientNet is trained to classify the samples that are accepted by the autoencoder (AE), i.e. the samples that have small reconstruction error. Our proposed method is evaluated in terms of precision, recall, and accuracy and compared to the state-of-the-art methods utilizing WildFish dataset. Simulation results reveal the supremacy of the proposed method.

## Article highlights

- Fish classification could be considered as an open-set problem, since the number of unknown underwater species is extremely high.
- Employing autoencoders, we can make the model to memorize the familiar species and throw away unfamiliar ones.
- The number of species highly affects the performance of the classification, while it roughly does not affect the false alarm rate of non underwater objects.

**Keywords** Fish recognition · Underwater images classification · Convolutional neural network · Autoencoder

## 1 Introduction

In recent years, detection of targets in underwater images and videos has gained much attention in fisheries as the received images and media from such harsh environments are in low resolution and quality. Besides

noise and segmentation problem, diversity posed a crucial problem in accurate classification. There is a great diversity of fishes which requires a lot of exploration and study in order for them to be recognized and also classified into a large number of classes. Based on previous studies, very few works took the open-set problem into

✉ Azim Akhtarshenas, A.akhtarshenas@ut.ac.ir; Ramin Toosi, R.toosi@ut.ac.ir | <sup>1</sup>Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran.



consideration. This in turn degraded the performance of algorithms by lowering accuracy. Due to its tremendous applications, fish species recognition and classification is of great importance in the domains of pattern recognition and image segmentation in oceans and ecosystem monitoring [1].

Detecting a large number of fishes of any kind in a dangerous environment needs a lot of efforts, time, and researches. So, introducing the automatic techniques is a prerequisite to manage and analyze the images captured by many cameras. To this end, Simone et al. [2] adopted a new approach that aims to capture the temporary dynamic of fish amplitude using the genetic algorithm. In order to generalize their method, the authors evaluated their proposed method under various environment conditions. Therefore, their conclusions revealed that their proposed method is very reliable to track fish variations at different time, background confusion, and various illuminance.

Several methods have been devised to help detect the underwater images so far. By increasing the samples and dataset used to train them, designed neural networks have promised to increase the recognition accuracy of the underwater targets. In order to better classify the underwater images and targets, the neural networks is required to be fed a large amount of data [3]. This ensures that the neural network has been generally trained and prepared to detect a large number of targets [4, 5]. So, deep learning and convolutional neural networks are two key solution that significantly help recognize the underwater images and files and also improve the quality of them using the classification techniques. However, obtaining the sufficient data set from some harsh environments sometimes becomes a big challenge of researchers. To address this challenge, some studies and researches aim to better detect the available objects and targets in received files using the combination of autoencoder (AE) and convolutional neural network (CNN). This combination not only reduces the amount of data, but also increases the accuracy of neural network by removing impertinent images and targets using AE [6]. Machine learning algorithms also have a promise to help discriminate data of both high and low quality and restore the lower quality of data.

Under new automated systems and quality controls, data analysis of fish images has become possible with higher performance and precision. This aims to properly classify the vocalization of the marine mammals that will be useful in underwater explorations. Thomas et al. adopted a CNN to classify the vocalizations of three species of whales, noise of non-biological sources, and an ambient noise that is capable to properly differentiate the aforementioned sources and identify the real presence of

whales in a special area. Moreover, the proposed CNN is able to learn high-level representations that ensures the generalization of their proposed network [7].

## 2 Related works

The remarkable growth of underwater data and fish images in recent years has created a critical demand for the fast detection and classification. Deep learning techniques have attracted a lot of attention from the researchers for their astonishing performance in digital image processing. Thanks to the deep neural networks, a huge amount of underwater data and images has been successfully processed and analyzed accurately in an incredibly short time [8].

Two major problems, namely, color degradation, light scattering, and poor light remarkably increased the difficulty of exploration in underwater environments and resulted in the degradation of underwater images. So, the received underwater images and videos are very invaluable in particularly seabed mining and archaeology. However, the appearance of the restoration and quality improvement methods, that help enhance the underwater images, has attracted great attention. Authors in [9] employed a self-adapted method that improves the quality of underwater images. Their method is designed to guide the image to either processing route based on a red deficiency measure. In the color correction phase, the histogram in each RGB channel is conveyed to balancing the image color. An adaptive histogram equalization method is utilized in order to increase the local contrast in the CIE-Lab color space. The dark channel prior haze removal scheme is modified for dehazing in the haze diminution phase. In the last phase, a histogram stretching method is used in the HSI (hue, saturation, intensity) color space to make the image seem more natural. Zheng et al. [10] used a fusion algorithm to restore and enhance underwater images. It consists of a color restoration module, an end-to-end defogging module and a brightness equalization module. In the color module, a color balance algorithm based on CIE Lab color model is applied to alleviate the impact of color deviation. In the end-to-end defogging module, one end is the input image and the other end is the output image. A CNN network is designed to connect these two ends in order to improve the contrast of the underwater images. In the CNN network, a sub-network is utilized to decrease the depth of the network that needs to be designed to obtain the same features. Several depth separable convolutions are considered to reduce the complexity during network training. The basic attention module is also adopted to highlight some important areas in the image. Moreover, in order to improve the defogging

network's ability to extract overall information, a cross-layer connection and pooling pyramid module are added. In the brightness equalization module, the authors used a contrast limited adaptive histogram equalization scheme to coordinate the overall brightness. Higher accuracy is very important in designing CNNs, recognition, and classification of underwater images and targets with higher precision has been one of the most significant area of image and computer vision. For example, the authors in [11] employed the available tools of machine learning to improve and enhance the received underwater images. Their proposed algorithm aimed to increase the accuracy, processing speed, and also the improvement of existing dataset and fish images.

In [12], the authors made use of two methods, namely max-RGB method and shades of gray method in order to improve the underwater images. Combination of two method and CNN have been combined to enhance the received images with about 90 % accuracy. However, the speed of detection is about 50 frame per second which shows the deficiency of their method, specially when a large amount of images and videos are existed.

Krizhevsky et al. faced up with a large amount of dataset and aimed to design a CNN with higher pace of training and also lower level of error [13]. The fully connected layers helped to reduce the overfitting using the dropout method. The simulation results revealed that the achieved error rate is much more less than that of previous works. A large part of ocean ecosystem consists of planktons and classifying of them can be a great help for those who are doing research in ocean ecosystems. The authors in [14] employed a CNN to classify a large number of dataset using a fine-grained classification method. However, majority of datasets belong to a few classes of planktons that causes imbalance class distribution. In order to reduce the negative effects of unbalancing in training of CNN, the authors make use of the transfer learning that reduces the amount of data for big-sized species of planktons. Limited color information of underwater videos and images is a challenging task of their classification that has attracted a lot of attention. In [15], an adaptive Gaussian mixture model was investigated to recognize special targets in poor quality underwater images. To this end, this method employed optimization techniques, namely particle swarm optimization, differential evolution, and genetic algorithm to initialize the variable set that led to the highest accuracy of 99%. Mohammad et al. introduced an image improvement method with two steps that not only helps restore the images, but also increase the accuracy of classifiers. The first step of their method is to correct the color and improve the quality of images. In the second step, a CNN is modeled to enhance the image improvement of previous step. In order to show the efficiency of proposed

algorithm, different images with various conditions were used to fulfill the peak signal to noise ratio (PSNR), mean square error (MSE), and structural similarity (SSIM). The simulation results from different experiments revealed that the super-resolution in the used algorithm yields an accurate results for medium layers of underwater environments using the realtime algorithms [16].

In [17], another effective method based on the CNN was proposed to properly restore the the received underwater images. This model learned to identify the relationship among the underwater stages and global ambient light and finally an optical imaging model was utilized to recover the images that are received from underwater environments. In the following, an image synthetic method was proposed to simulate underwater images received from different underwater depths and environments. Simulation results showed that the proposed CNN based method is effectively help classify and restore the underwater images. Authors in [18] introduce an improved method based on the fusion method that restores accurately underwater images. In this scheme, a single image is considered as the input and a sequence of operations such as white balancing, gamma correction, sharpening, manipulating weight maps are performed on the input image. in the following, multiscale image fusion of the inputs is applied to obtain the resultant output. In the first step, color distorted input image is white balanced to eliminate the color casts maintaining a realistic subsea image. In the second step, the authors employed CLAHE that serves an important role in brilliance enhancement. Concurrently, histogram equalization is carried out on the sharpened image. The weight maps analyze image characteristics that properly recognize the spatial pixel connection and correlation. Finally in the last step, multiscale pyramidal fusion of the inputs and weight maps are applied. The authors in [19] adopted a novel learning classifier system (LCS) which accurately classifies large-size of underwater images using a novel classification convolution autoencoder (CCA). CCAE makes use the combination of the autoencoder and classifier in order to enhance the accuracy and precision of system. CCAE also aims to decompress LCS generated rules to their original input space. To evaluate their proposed solution, the authors conducted experiments on selected underwater synsets of ImageNet dataset. Due to the large amount of images received from underwater environments, another useful processing tool is employing automated methods. As it is mentioned earlier, automation techniques are very helpful in oceanography and have promise to properly identify the marine species. The authors in [20] aimed to analyze a large received images using the automated methods and machine learning algorithms. To this end, a deep technique is employed to automatically annotate and quantify the unlabeled

coral reefs and their coverage to track any change in coral population. A new classification method of underwater targets was proposed to categorize the received images based on CNN. Markov randomly field-Grabcut algorithm was applied to fragment the images to shadow and seabottom sets and then, CNN was learned to classify three forms group of underwater targets, namely, cylinder, truncated cone and sphere. In order to prove the validity their proposed model, synthetic aperture data sets are fed on and the simulation results showed that the proposed model outperforms SVM and CNN in term of accuracy [21]. Alongside environmental conditions, some other reasons pose a threat to the Bangladesh water resources. Sharmin et al. [1] aimed to identify and classify the freshwater fishes in order to introduce them to the next generations. However, There is not a convincing comparison in their study. Zong-Yao et al. [22] investigated a two-stage scheme to detect and classify fish. They applied SSD to perform video sequences and then utilized it to assess the exactness of SVM classification. The authors in [23] investigated a CNN to recognize fish species using transfer learning and then classify them utilizing SVM classifier. They primarily utilized AlexNet model to distinguish fish features from available dataset and then fine-tuned their proposed model to adequately achieve features. Several misclassified fish images have been represented and left without taking them into consideration.

## 2.1 Motivations and contributions

From the previous section, it can be understood that none of described researches has focused on open-set scenario, where there exist unseen or unfamiliar species that reveal a research gap among previous works. In other words, in spite of the huge underwater exploration, only a small subset of the great creatures' set has been recognized and classified so far. To shed more light on the issue, complete identification of underwater targets seems to be costly and time-consuming process. However, received underwater images and videos contain both recognized and unrecognized species. This combination is an unfortunate for classifiers and gives rise to the higher inaccuracy in the final models in real scenarios. In other words, neural networks are supposed to classify any image and target into available classes while some of those targets do not belong to any class. Therefore, all of the previous studies mostly focused on restoration, image enhancement, and classification of underwater images, regardless of whether that underwater target's image is belonged to their available classes or not. This large diversity among underwater images significantly reduces the accuracy of proposed classifier which has to classify received images under any circumstances and degree of similarity, specially

when the number of their classes are not large enough. Motivated by classification schemes in previous studies, particularly in [7, 21], we proposed a model that appropriately classifies only those underwater images which are belonged to the exiting classes while dropping out the remained anomalous or irrelevant images. This helps substantially increase the accuracy of restoration and classification as the underwater images are not always categorized into their available classes for two important reasons. Firstly, the number of underwater creatures and fishes is big enough where minority of them have been classified so far. Secondly, the propagation problems and fast variation of underwater channels reduce the quality and provide very poor quality of underwater images over a long distance. In order to properly identify and classify the final images, we proposed a CNN that systematically categorizes the familiar images and discard unseen ones. The proposed CNN requires few parameters that results in lower complexity and gives time-saving classification inference. We summarized the main contributions of this paper as follows:

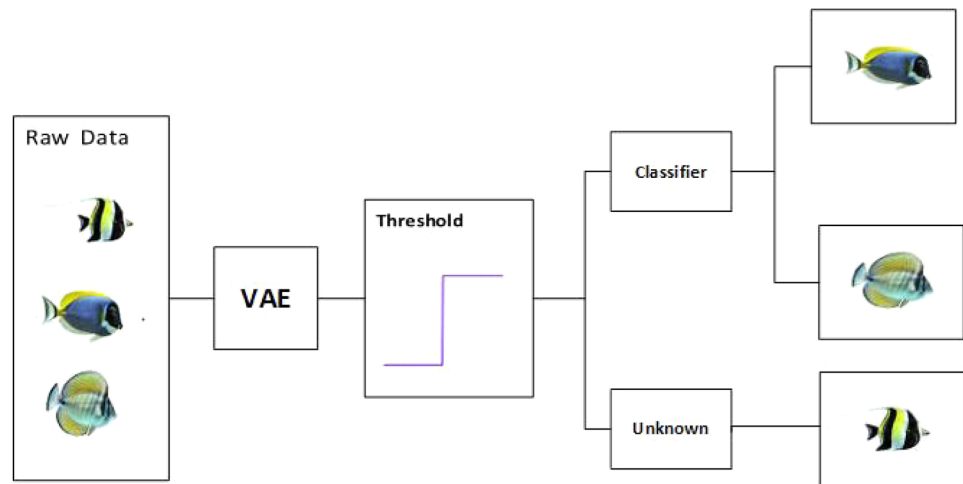
- Proposed AE-CNN throws away underwater images of unknown classes in order to reduce the classification's error and complexity.
- Proposed AE-CNN classifies the underwater images with the higher accuracy than that of previous works.

To benchmarking and evaluating the effectiveness of proposed system, we compare the results of the proposed model with the quality of different approaches [19, 20]. Following tables in simulation section indicates the accuracy of the proposed method and the accuracy of two the state-of-the-art methods. It is clear that the proposed method outperforms two mentioned the state-of-the-art methods. This comparison will reveal that the accuracy of the proposed system is higher than two popular studies. This superiority of the proposed method clearly reveals the significance of using AE before classifiers. The rest of this paper is organized as follows: In Sect. 4, we introduce the system description and proposed CNN. In Sect. 5, experiments, simulation, results, and figures are represented and compared. In Sect. 6, we discuss previous researches and proposed model in this paper. In Sect. 7, we conclude the paper.

## 3 System description

In this section we go through the details of the proposed method which consists of two main parts. As it can be seen in Fig 1, in the first part, an autoencoder is used to learn the characteristics of the known classes by reducing the

**Fig. 1** Blockdiagram of the proposed method



reconstruction error over the available species. Then, in the second part, a CNN based classifier is trained to classify the known species. The proposed classifier only classifies the species that are considered as known by the AE. Each part utilized a loss function which is described in next section with details. This approach not only discards the irrelevant and too-much-degraded images, but also reduces the complexity of identification and classification in classifiers.

### 3.1 Proposed method

CNNs play an important role in image and video classification as a strong tool because they help successfully classify a large number of images into a few certain classes for analytical purposes. To study our problem and address the mentioned concerns with machine learning techniques, we employ a CNN which aims to classify a large number of underwater images.

The proposed CNN consists of multiple layers, and its input layer extracts our image features using a kernel and image matrix. Due to its special operation, current neural networks technically cover the corner pixels once. In order not to lose the information of edges of available images, a padding is used to cover the corner pixels more than once. During the design procedure of the neural network, we also used the process of max pooling to reduce the volume of information and data loading. Some dropout layers also utilized to prevent CNN from overfitting by randomly ignoring or dropping out some nodes in each epoch. In fact, dropout is a regularization topology to estimate learning a great number of CNNs with various structures in parallel. Then fully connected layers are added to the structure before the output layer.

Blockdiagram of the proposed method is represented in Fig. 1 which consists of two main parts. Part one represents AE and part two represents classifier. AE acts as a

filter which receives raw data, calculate the loss value, and remove those images which have the loss value over than threshold. In other words, received images are primarily fed into the AE to calculate the loss value of AE in network. If the calculated loss value of this step for one image is larger than a threshold, the image will be dropped out from image collection and otherwise, the received image is considered to be classified between two groups of subsets.

In stochastic processing, the mean squared error (MSE) of an estimator is used to measure the average of the squares of the errors which is the average squared difference between the estimated values (reconstructed image) and the actual value (original image). Therefore, we employ the MSE as the loss function for detection through the autoencoder:

$$\mathcal{L}(\tilde{y}, y) = MSE(y, \tilde{y}) = E[(y - \tilde{y})^2] \quad (1)$$

Where  $y, \tilde{y}$  show the original and reconstructed images respectively.

The AE architecture is shown in Fig. 2. We train the AE, using the available classes with reconstruction loss under the above equation. Thus, the AE aims to learn the characteristics of the available species to reconstruct them. On the other hand, for unseen species, the reconstruction fails. Thus, by considering a reconstruction loss of the input image to the AE, one could distinguish between seen and unseen classes. In this manner, we could handle the fish classification problem in an open-set manner. To find the appropriate threshold for reconstruction loss, we need to investigate two important errors: i) false rejection or miss which is the error of thrown away a known specie, and ii) false acceptance or false alarm, which is the acceptance of an unknown specie and send it to the classifier. These values are usually having an opposite direction of

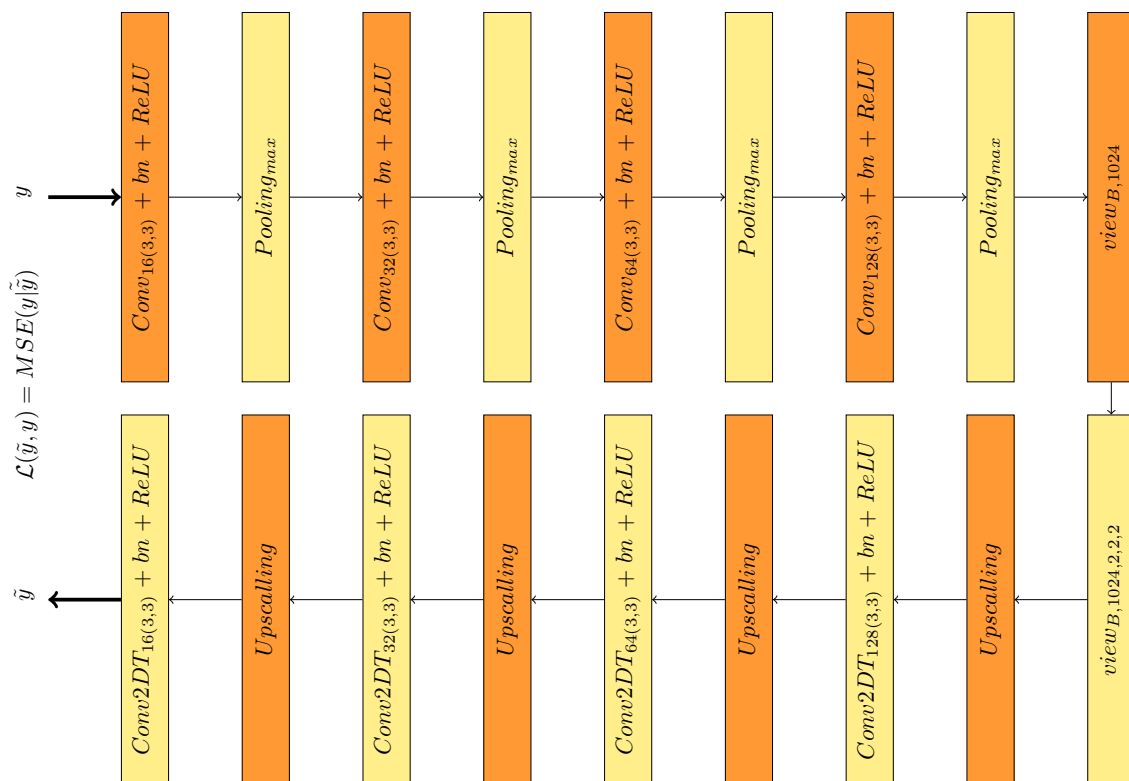


Fig. 2 Unseen detection module architecture (AE)

increasing or decreasing. Thus, a proper threshold must be determined based on the application and each error type importance. Here, we choose the threshold where these two errors become equal which is also called the threshold of equal error rate in the machine learning literature.

In the second part of the algorithm which aims to classify the received images from the AE, we consider a CNN based classifier to classify the underwater images and files into several categorizations. We also employ the following loss function for classification through the neural network:

$$Loss = \sum_i p_i \log(\tilde{p}_i) \tag{2}$$

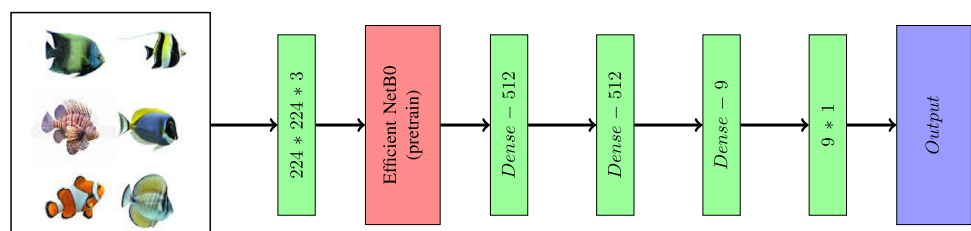
Where  $\tilde{p}_i$  and  $p_i$  show the estimated class index and target class index respectively. The proposed classifier is based on recently published network, EfficientNet which is properly pretrained over Imagenet dataset. We use the pretrained weights as initial weights of the proposed classifier to train

over the known species. As seen in Fig. 3, the Efficient-NetB0 structure is utilized to classify the images. This part technically plays an important role in classification system, since the final decision about classification is made by this classifier and eventually, adding it into the AE completes the structure of the proposed method. AE and CNN jointly cooperate to systematically classify the most appropriate images into corresponded class.

### 4 Experiments and results

Here, we also compare the performance of the proposed method against two state-of-the-art methods. Results are shown in Table 1, where the best method is marked as bold. As can be seen, the proposed method has about %5 higher accuracy in image classification that of two other methods. So, thanks to the AE block, our proposed method

Fig. 3 Image classification module architecture: Underwater targets are fed into classification module as inpts and classified into classes in output block



**Table 1** Accuracy and inference time of the proposed classification against two state of the art methods

Number of classes	300	500	700	Inference (ms)
Proposed method	<b>0.9370</b>	<b>0.8912</b>	<b>0.8159</b>	63
LCS [19]	0.8843	0.8425	0.7592	154
Automated Method [20]	0.8802	0.7975	0.7504	73

Best accuracy is marked as bold

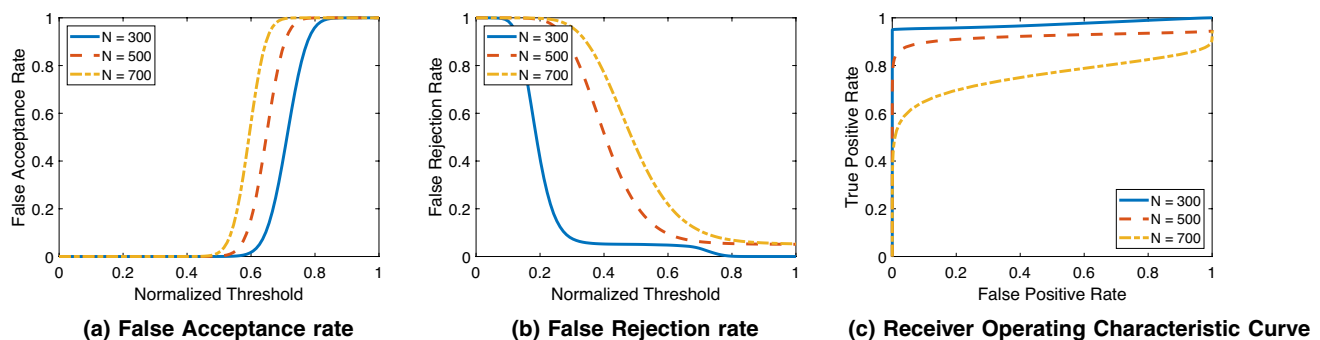
clearly outperforms other methods in terms of accuracy. In the last column the average inference time of the methods are illustrated. As shown, the proposed method has about %15 lower inference time in image classification that of two other methods.

In order to evaluate the performance of the AE and the total proposed system in term of accuracy, we represent simulation figures and tables in this part. To this end, we use WildFish [24] dataset to evaluate the proposed method. This dataset consists of 1000 image categories and 54459 images. Through this section the proposed method is referenced to as AE-CNN. We also portray the curve of FA and FR for three values of  $N=300$ , 500, and 700, where  $N$  is the number of training categories. First, we discuss the simulations of the autoencoder part of the proposed method, which aims to distinguish between the seen and unseen categories. In the following, the seen samples are delivered to the classifier block while unseen ones are already dropped out. Then we evaluate the classification part of the proposed method. Finally, the numerical results of the AE-CNN are compared to the state-of-the-art methods.

In the first experiment, the performance of autoencoder part of the AE-CNN is investigated and evaluated. It is worthy to mention that false acceptance (FA), false rejection (FR), equal error rate (EER), and area under the ROC curve (AUC) are four metrics employed to evaluate the proposed method. EER is the point where FA and FR becomes equal.

The lower the EER value, the higher the accuracy of the proposed system. The number of categories used to train the network, affects the performance, thus, all aforementioned metrics are calculated for different values for the number categories in train set. Fig. 4 indicates FA, FR, and ROC for different values of the number of training categories, respectively. As it can be seen, Fig. 4 show the FA, FR versus normalized threshold and also Fig. 4 represents ROC versus false positive rate. As expected, the performance value decreases as the number of categories increases. This is because of the fact that as the number of training categories increases, the task of the AE module becomes harder and it leads to performance reduction accordingly. AE, in other words, needs to model larger number of fish categories, which is harder due to the diversity of received images. The EER and AUC, calculated for the plots in Fig. 4, are shown in Table 2. As expected, the performance in terms of EER and AUC decreases as the number of training categories increases. In the second experiment, we investigate the performance of the proposed method against object categories other than fish. In other words, what happens if we feed a person to the AE. For this sake, we used the CIFAR-100 [25] dataset. It has 100 categories with 600 images for each category. However, we removed the aquatic mammals and fish super-classes to only investigate the performance of the proposed method against objects outside the sea. Fig. 5 represents the FA of the proposed AE. As can be seen, the performance of the proposed AE slightly decreases as the number of training categories increases. These two experiments obviously show that as the diversity of training samples increases, the generalization ability of the network decreases. However, for categories other than fish, this decrement is negligible.

In the third experiment, we investigate the classification part of the proposed method. As what we did before, again the number of classes could affect the performance of the classifier. Thus, the accuracy of the classifier is investigated as the number of classes increases. First, we evaluate the

**Fig. 4** Performance evaluation of the proposed autoencoder model

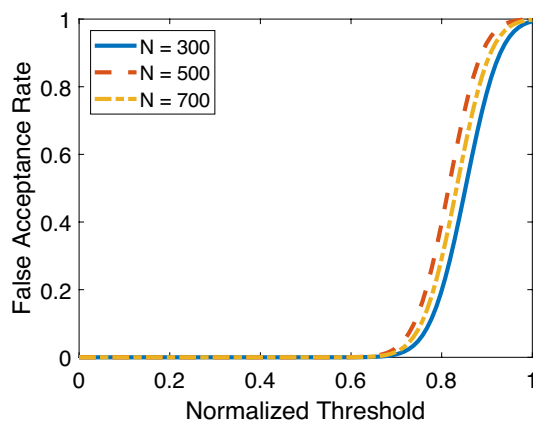
**Table 2** Equal error rate and area under curve of the proposed method against increasing in the number of training classes

Number of classes	300	500	700
AUC	0.9734	0.9204	0.7565
EER	0.0483	0.1003	0.2712

effect of structure on the performance of the proposed method. As stated in Fig. 3, EfficientNet-B0 is used as the feature extractor. We also evaluated other structures including, Resnet18 [26], VGG16 [27], DenseNet-121 [28], Xception [29], and Inception-v3 [30]. Results are shown in Table 3, where the number of parameter for each structure is also stated. As can be seen, EfficientNet-B0 has the best accuracy with the lowest number of parameter in various number of training examples. Xception is the closest structure in terms of accuracy while the number of parameters is much higher than EfficientNet-B0. Thus, we chose EfficientNet as the feature extractor in our proposed method.

### 5 Discussion

Underwater object recognition is an important issue in exploring the underwater environments and oceanography. So, there is a demand for automatic systems in order to precisely recognize and classify marine species in underwater files without human interference. Fish recognition is one of the most applicable and changeable task in marine system understanding. The number of fish species is huge, so we only could access a small set of species to train our algorithm, while there exist lots of species not available for the algorithms to be trained or even unknown for ourselves. Thus, while there exist various classification



**Fig. 5** False acceptance rate for objects out of the sea

**Table 3** The effect of various feature extraction structures on the classification performance of the proposed method

Method	Number of parameters	ACC@300	ACC@500	ACC@700
EfficientNet-B0	5.3M	0.9370	0.8912	0.8159
ResNet18	11.4M	0.8269	0.7947	0.7192
VGG16	134.7M	0.7831	0.7594	0.6508
DenseNet-121	7.2	0.8846	0.8301	0.7436
Xception	23M	0.9182	0.8769	0.8036
Inception-v3	24M	0.9047	0.8620	0.7861

based studies, we must put forward some proposals to deal with an open-set problem in real scenarios.

Leilei et al. [31], proposed a new improved median filter that eliminates noise in fish images. To this end, they designed a CNN and trained it with images from ImageNet. However, there exist a lack of comparison with state-of-the-art methods. We tried to evaluate various aspects of our method and compared it with other state-of-the-art method to evaluate the performance of the proposed method. Results show that the proposed method outperforms other state-of-the-art methods. Imbalanced data is another challenge in underwater classification problems. Vikram et al. [32], classified various fish images captured by cameras utilizing neural network. They aimed to improve the accuracy utilizing different structures of networks. However, dataset size and imbalance classes may result in high inaccuracy in open-set scenarios. In our evaluation, we always balanced the training dataset to prevent the bias of the output class in favor of a specific class or classes.

Low number of samples or categories in the dataset is another challenge for fish classification. FANGFANG et al. [33], employed a simple CNN in order to perform identification and classification of the fish behaviors using a group behavior discrimination approach. However, Small sample size in both studies may lead to overfitting and less generalization ability in CNNs in spite of the high precision for the training set. In this study, we used a large dataset of fish images with 1000 categories. A large number of examples and categories make the evaluation more reliable and gives the model more generalization ability.

Allken et al. [34], employed a deep learning neural network to automatically classify the species in *synthetic* underwater images. However, the performance in real scenarios maybe different from synthetic inputs. Here, we use a dataset of real images to ensure that the performance is valid for real scenarios. Unfortunately, majority of previous works ignore unseen species while quite a lot of species received from marine life are unseen or degraded due to environmental conditions and noise. These unfamiliar



species and low quality images will pose an extraordinary problem for the neural networks.

## 6 Conclusion

In this paper, we introduced a new framework based on autoencoder and CNN to address the fish recognition problem in an open-set manner. However, due to a very large number of available images, complexity increases remarkably and gives rise to a huge cost and time waste. So, removing irrelevant images or unseen targets can be a significant strategy to help reduce the volume of the input images to the classification module. One of the most applicable techniques to properly tackle this critical challenge is to employ an autoencoder that technically removes the unseen or inappropriate images before classification step. For example, some images with insufficient quality and inappropriate resolution are eliminated by AE. In the proposed method, an AE was employed to properly train and carefully distinguish between seen and unseen species. In other words, the AE filtered out species that are not in our training set. To do so, the AE was trained to reconstruct the available species with high accuracy. Thus, in the case of unseen species, the reconstruction error was too high. Therefore, we could prevent the unseen species from going to the classifier section with 95.2% accuracy. After that, a classifier based on EfficientNet was trained to precisely classify the samples that are accepted by the AE, i.e. the samples that had small reconstruction error. Moreover, the proposed CNN in this paper is small and has fast inference time. It is worth mentioning that the size of dataset helps avoid any overfitting. Our proposed method was evaluated utilizing WildFish dataset and compared to the state-of-the-art methods. The proposed method achieved 93.7% accuracy in open-set scenario.

**Funding** The authors have not disclosed any funding.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Ethical approval** The manuscript is not submitted elsewhere and is an original research. There are no animal and human studies involved in this research. The authors have no conflicts of interest to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate

if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Sharmin I, Islam NF, Jahan I, Joye TA, Rahman MR, Habib MT (2019) Machine vision based local fish recognition. *SN Appl Sci* 1(12):1–12
2. Marini S, Fanelli E, Sbragaglia V, Azzurro E, Fernandez JDR, Aguzzi J (2018) Tracking fish abundance by underwater image recognition. *Sci Rep* 8(1):1–12
3. Roy SK, Krishna G, Dubey SR, Chaudhuri BB (2019) Hybridsn: exploring 3-d-2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 17(2):277–281
4. Ahmad M, Shabbir S, Roy SK, Hong D, Wu X, Yao J, Khan AM, Mazzara M, Distefano S, Chanussot J (2021) Hyperspectral image classification-traditional to deep models: a survey for future prospects. *IEEE J Sel Top Appl Earth Obs Remote Sens* 15:968–999
5. Dubey SR, Chakraborty S, Roy SK, Mukherjee S, Singh SK, Chaudhuri BB (2019) diffgrad: an optimization method for convolutional neural networks. *IEEE Trans Neural Netw Learn Syst* 31(11):4500–4511
6. Teng B, Zhao H (2020) Underwater target recognition methods based on the framework of deep learning: a survey. *Int J Adv Robot Syst* 17(6):1729881420976307
7. Thomas M, Martin B, Kowarski K, Gaudet B, Matwin S (2019) Marine mammal species classification using convolutional neural networks and a novel acoustic representation. In: *Joint European conference on machine learning and knowledge discovery in databases*. Springer, Berlin, pp 290–305
8. Moniruzzaman M, Islam SMS, Bennamoun M, Lavery P (2017) Deep learning on underwater marine object detection: a survey. In: *International conference on advanced concepts for intelligent vision systems*. Springer, Berlin, pp 150–160
9. Chang H-H, Chen P-F, Guo J-K, Sung C-C (2020) A self-adaptive single underwater image restoration algorithm for improving graphic quality. *EURASIP J Image Video Process* 2020(1):1–21
10. Zheng M, Luo W (2022) Underwater image enhancement using improved CNN based defogging. *Electronics* 11(1):150
11. Seese N, Myers A, Smith K, Smith AO (2016) Adaptive foreground extraction for deep fish classification. In: *2016 ICPR 2nd workshop on computer vision for analysis of underwater imagery (CVAUI)*. IEEE, pp 19–24
12. Han F, Yao J, Zhu H, Wang C (2020) Underwater image processing and object detection based on deep CNN method. *J Sens*. <https://doi.org/10.1155/2020/6707328>
13. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105
14. Lee H, Park M, Kim J (2016) Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning. In: *IEEE international conference on image processing (ICIP)*. IEEE, pp 3713–3717
15. Kannan S (2020) Intelligent object recognition in underwater images using evolutionary-based Gaussian mixture model and shape matching. *Signal Image Video Process* 14:877–885

16. Moghimi MK, Mohanna F (2020) Real-time underwater image resolution enhancement using super-resolution with deep convolutional neural networks. *J Real-Time Image Process* 18:1653–1667
17. Hu Y, Wang K, Zhao X, Wang H, Li Y (2018) Underwater image restoration based on convolutional neural network. In: Asian conference on machine learning, pp 296–311
18. Mohan S, Simon P (2020) Underwater image enhancement based on histogram manipulation and multiscale fusion. *Procedia Comput Sci* 171:941–950
19. Irfan M, Jiangbin Z, Iqbal M, Arif MH (2021) Enhancing learning classifier systems through convolutional autoencoder to classify underwater images. *Soft Comput* 25:10423–10440
20. Mahmood A, Bennamoun M, An S, Sohel FA, Boussaid F, Hovey R, Kendrick GA, Fisher RB (2018) Deep image representations for coral image classification. *IEEE J Ocean Eng* 44(1):121–131
21. Zhu K, Tian J, Huang H (2018) Underwater object images classification based on convolutional neural network. In: 2018 IEEE 3rd international conference on signal and image processing (ICSIP). IEEE, pp 301–305
22. Wu Z-Y, Tseng S-L, Lin, H-Y, Chen H-Y, Van Luan T (2019) Incorporating stereo with convolutional neural networks for real-time fish detection and classification. In: 2019 IEEE international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM). IEEE, pp 83–88
23. Tamou AB, Benzinou A, Nasreddine K, Ballihi L (2018) Transfer learning with deep convolutional neural network for underwater live fish recognition. In: 2018 IEEE international conference on image processing, applications and systems (IPAS). IEEE, pp 204–209
24. Zhuang P, Wang Y, Qiao Y (2018) Wildfish: a large benchmark for fish recognition in the wild. In: 2018 ACM multimedia conference on multimedia conference. ACM, pp 1301–1309
25. Krizhevsky A, Hinton G et al (2009) Learning multiple layers of features from tiny images
26. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. In: European conference on computer vision. Springer, Berlin, pp 630–645
27. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition, arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
28. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
29. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
30. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2015) Rethinking the inception architecture for computer vision. *CoRR*, vol. abs/1512.00567, 2015. [arXiv: org/abs/1512.00567](https://arxiv.org/abs/1512.00567)
31. Jin L, Liang H (2017) Deep learning for underwater image recognition in small sample size situations. In: OCEANS 2017-Aberdeen. IEEE, pp 1–4
32. Deep BV, Dash R (2019) Underwater fish species recognition using deep learning techniques. In: 2019 6th international conference on signal processing and integrated networks (SPIN). IEEE, pp 665–669
33. Han F, Zhu J, Liu B, Zhang B, Xie F (2020) Fish shoals behavior detection based on convolutional neural network and spatiotemporal information. *IEEE Access* 8:126907–126926
34. Allken V, Handegard NO, Rosen S, Schreyeck T, Mahiout T, Malde K (2019) Fish species identification using a convolutional neural network trained on synthetic data. *ICES J Mar Sci* 76(1):342–349

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.