Underwater Fish Species Recognition using Deep Learning Techniques

B Vikram Deep
Dept. of Computer Science and Engineering
National Institute of Technology, Rourkela
Rourkela, India
Email: vikramdeep.b@gmail.com

Ratnakar Dash
Dept. of Computer Science and Engineering
National Institute of Technology, Rourkela
Rourkela, India
Email: ratnakar.dash@gmail.com

***Abstract*—Underwater fish species recognition has gained importance due to the emerging researches in marine science. Automating the fish species identification using technology would help the marine science to evolve further. Image classification tasks have seen a rise with the introduction of deep learning techniques. In this paper, we have proposed a hybrid Convolutional Neural Network (CNN) framework that uses CNN for feature extraction and Support Vector Machine (SVM) and K-Nearest Neighbour (k-NN) for classification. Both the proposed frameworks are tested on Fish4Knowledge dataset. Our experimental results show that our framework gives better results than most of the traditional as well as existing deep learning techniques.**

***Index Terms*—Image classification, Deep learning, Convolutional Neural Network (CNN), CNN-SVM, CNN-KNN, Fish4Knowledge**

I. Introduction

Underwater object recognition has been an emerging research area because of the increase in the volume of underwater data available by different observatories like CANADA, VENUS etc. Fish recognition is one of the important tasks under underwater object detection due to its prominence in oceanography, marine science. Fish species recognition can help academic researchers, ocean scientists and biologists [1] [2]. Also, it help's in determining the biomass level as well as geological changes in oceans. Due to the importance of fish species recognition, many computer vision methodologies are proposed to classify fish species accurately. The fish species classification can be broadly divided into three application areas on the basis of scope [3]:

- Fish species recognition on dead fish.(eg. conveyor belt classification in industries.)
- Fish species recognition in an artificial habitat. (eg. aquarium, water tanks etc.)
- Fish species recognition in natural habitat. (eg. sea, ocean etc.)

Underwater fish species recognition is a tough task. There were researches on fish recognition on dead fish [4] [5] and fish taken out of water [6] [7] [8] [9] or in artificial tanks. But, there were no significant researches on underwater habitat. This is mainly because of the unrestricted natural habitat of the sea. The underwater videos are of low quality with complex backgrounds and low luminosity. Visualizing the fish species can help us acquire deep knowledge on the movement and activities of the species as a whole. Image classification has been an emerging research area with the introduction of deep learning techniques. This is mainly due to the fact that deep learning methods such as CNN, does not need any feature extraction upfront. Most of the existing frameworks for object recognition are based on the images on the ground.

The remaining paper is organized as follows: Section II provides information on the existing frameworks, Section III presents the methodologies used, Section IV discusses the experimental results, and Section V provides conclusion and future work with references followed.

II. Literature Survey

Fish species recognition in uncontrolled natural habitat is a difficult task because of the complex background and noise in the images. Many researchers have worked on this domain in the last decade and sophisticated methods have been proposed to classify fish species in their natural habitat.

Leilei Jin et al. [10] has proposed a framework that works in small sample size situations for underwater fish species recognition. The sample images are preprocessed with improved median filter. The improved median filter differs from conventional median filter by filtering only impulse noise pixels and unaltering the other pixels. A ConvNet is used that is pre-trained with large ImageNet dataset. It is then fine-tuned and trained with the sample images from Fish4Knowledge dataset. Leilei Jin has achieved an accuracy of 85.05 % on the dataset.

Dhruv Rathi et al. [11] has proposed a framework that is based on deep learning and image processing techniques. The sample images are first pre-processed with Otsus thresholding, erosion and dilation. The resulting image along with original image is passed on to CNN to classify fish species. The dataset used is Fish4Knowledge and has achieved an accuracy of 96.29 %.

Xin Sun et al. [12] has proposed a framework that takes a low resolution image and converts it into high resolution image using single image super-resolution method [13]. Two deep learning techniques PCANet [14] and Network In Network (NIN) [15] are used for feature extraction and a linear SVM is used to classify the species. The dataset used is FishCLEF2015 and the PCANet has achieved a precision of 77.27% and NIN a precision of 69.84%.

Katy Blanc et al. [16] used a processing chain based on background segmentation, selection key points with an adaptive scale, description with Opponent-Sift for feature extraction and a binary linear Support Vector Machine for classification.

Hongwei Qin et al. [17] used a foreground extraction method which uses sparse and low-rank matrix decomposition [18]. to eliminate the complex background of the fish images. Feature extraction is performed to extract the features of the foreground fish images using a deep architecture. The architecture contains Principal Component Analysis (PCA) in two convolutional layers, followed by binary hashing in the non-linear and blockwise histograms in the feature pooling layer. These are followed by a spatial pyramid pooling to extract pose invariant information. Classification is performed using a linear SVM. This framework achieved an accuracy of 98.64% on the Fish4Knowledge dataset.

Salman et al. [19] has proposed a CNN in an hierarchical feature combination framework to learn features that are dependent on species and their variability. The feature extraction is performed using CNN and classification is performed using SVM and KNN. The dataset used is FISHCLEF 2014 and FISHCLEF 2015. The framework achieved a higher accuracy of 97.41%.

Mohammed et al. [20] has proposed a SVM framework along with feature extraction techniques for Nile Tilapia fish classification. SIFT and SURF techniques are used for image feature extraction and produced good results.

These frameworks have used machine learning techniques for the fish species classification task. There are other frameworks that used traditional techniques for the task.

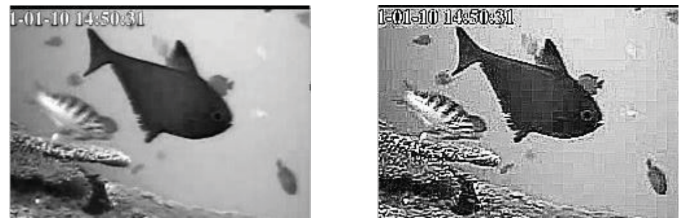
Moniruzzaman et al. [21] provides a survey on different techniques applied on the underwater fish species classification.

III. Methodology

One of the key steps in underwater fish species recognition task is to pre-process the sample images. This is important because the sample images are usually blurry due to the uncontrolled environment. This may lead to the failure of the classifier to learn species dependent features accurately.

A. Image Sharpening

The fish4knowledge dataset is used that contains sample images of 23 species. These images are mostly clear of noise but most of them are blurry. So image sharpening method is used to extract sharp edges of the image so that it can be used to pass on to the framework. A kernel is used for the sharpening process. Usually, a laplacian kernel is used, whose sum of elements is equal to 0. But a slightly modified kernel is used such that the sum of elements is equal to 1. Because the laplacian kernel gives a binary image, the modified kernel gives the color image as output.



(a) Original image (b) Sharpened image

Fig. 1: Sample fish image from Fish4Knowledge dataset where a) original image and b)sharpened image

From Fig. 1, the original image is slightly blurred but the sharpened image shows the little details missed in the original image.

B. Convolutional Neural Network

Image classification has seen an uprise with the introduction of deep learning techniques. These techniques do not need any feature extraction upfront. The network learns the features by itself by constantly updating the weights. Convolutional neural networks can be used as either feature extraction techniques or as a complete classifier.

CNN consists of convolution filters that use a kernel to convolute with the image to extract feature vectors. Pooling layer is responsible for the subsampling process either using maximum pixel value in the kernel (Max pooling) or average pixel values in the kernel (Average pooling). Fully connected layers are used to connect every feature vector from the incoming layers to the next layer.

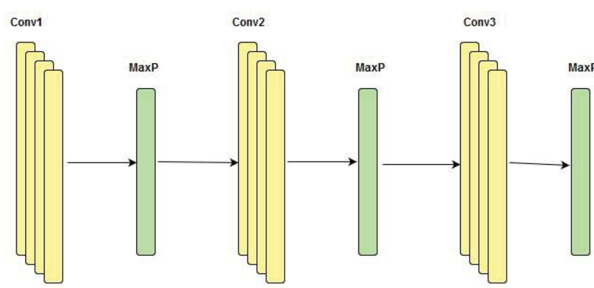


Fig. 2: Architecture of proposed DeepCNN framework

The architecture of the framework used in the fish species recognition is shown in Fig. 2. In the framework, three convolution layers are used that are each followed a max pooling layer. These layers are then connected to two fully connected layers and the last fully connected layer classifies the fish species.

The main reason behind using max pooling is its ability to detect edges. In fish species recognition, detecting the fish border is an important task. Fish images are sharpened using image sharpening. Usage of average pooling may not be suitable as it averages the pixels in the kernel.

The first convolution layer contains 32 filters of 3x3 kernel size, the second convolution layer contains 64 filters of 3x3 kernel and the third convolution layer contains 128 filters of 3x3 kernel size. All the convolution layers are followed by a max pooling layer of 2x2 kernel. A dropout layer is also used before a fully connected layer to ensure that overfitting is avoided. In all the convolution layers Rectified Linear Unit(ReLU) is used as an activation function. In the last fully connected layer Softmax is used as an activation function.

C. Hybrid Convolutional Neural Network

In the conventional method, the convolutional neural network is used for feature extraction as well as a classifier. In the hybrid convolutional neural network, the classification is performed using different classifiers. The following methods are proposed based on the DeepCNN architecture.

1) DeepCNN-SVM: The DeepCNN architecture is used for feature extraction and the features are used to classify the fish species using SVM classifier. The SVM classifier is used with the linear kernel that used one vs. rest for classification of fish species.

2) DeepCNN-KNN: The DeepCNN architecture is used for feature extraction and the features are used to classify the fish species using k-NN classifier. The k-NN classifier is tested with different values of k neighbors and the k value with the better result is proposed.

IV. Experimental Results and Discussion

A. System Overview

The proposed framework is simulated using the Python environment with keras framework and tensorflow backend. The system settings are : Intel(R) Core(TM) i7-3770 CPU: 3.40 GHz. RAM: 4GB.

B. Dataset

The fish4knowledge dataset contains underwater videos of underwater live fishes as well as underwater fish images. This dataset is acquired from the underwater observatories at Nanwan, Lanyu and Houbi Lake of Taiwan. The dataset contains 27370 verified fish samples in image format. These fishes are divided into 23 species. The fishes are divided based on the species dependent features such as the number of fins, shape etc. The dataset is very imbalanced in the sense that the least frequent species is about 1000 times less than the most frequent species. The fish images are obtained using detection and tracking software in [22]. The fish species are manually labeled by following instructions from marine biologists [23]. TABLE I gives the dataset description.

The whole dataset is divided into 90% training and 10% testing as the dataset is highly imbalanced and the least frequent species has only 16 images. In the 90% training set, the species that contains less than 300 images are augmented. Each image is randomly rotated from -10^0 t0 10^0 and added to the dataset. This procedure is repeated 5 times [17]. After the data augmentation, the training dataset is further divided into 90% training set and 10% validation set. The details of the division are shown in TABLE II.

TABLE I: Fish4Knowledge image dataset

S.No	*Species*	*Number of Images*
01	Dascyllus reticulatus	12112
02	Plectroglyphidodon dickii	2683
03	Chromis chrysura	3583
04	Amphiprion clarkii	4049
05	Chaetodon lunulatus	2534
06	Chaetodon trifascialis	190
07	Myripristis kuntee	450
08	Acanthurus nigrofuscus	218
09	Hemigymnus fasciatus	241
10	Neoniphon sammara	299
11	Abudefduf vaigiensis	98
12	Canthigaster valentini	147
13	Pomacentrus moluccensis	181
14	Zebrasoma scopas	90
15	Hemigymnus melapterus	42
16	Lutjanus fulvus	206
17	Scolopsis bilineata	49
18	Scaridae	56
19	Pempheris vanicolensis	29
20	Zanclus cornutus	21
21	Neoglyphidodon nigroris	16
22	Balistapus undulatus	41
23	Siganus fuscescens	25

TABLE II: Fish4Knowledge dataset division

Training set	*Validation set*	*Test set*
30,047	3339	2739

To validate the proposed hybrid frameworks, simulation has been carried out using Fish4Knowledge dataset. For performance comparison accuracy, precision, recall, and f-score [24] are used. These measures are calculated using the following equations.

$$Precision = \frac{t_p}{t_p + f_p} \tag{1}$$

$$Recall = \frac{t_p}{t_p + f_n} \tag{2}$$

$$f - score = 2\frac{Precision X Recall}{Precision + Recall} \tag{3}$$

where t_p and t_n are true positive and true negative, f_p and f_n are false positive and false negative calculated from the confusion table.These performance measures along with accuracy can provide insights into the performance of our proposed frameworks.

TABLE III: Accuracy comparison of proposed frameworks

Framework	*Accuracy(%)*
Random Forest	81.38
k-NN	78.98
SVM	78.78
DeepCNN	98.65
DeepCNN-SVM	98.32
DeepCNN-kNN	**98.79**

TABLE IV: Performance evaluation of proposed frameworks

Framework	*Avg-Precision (%)*	*Avg-Recall (%)*	*Avg-F-score*
Random Forest	62.33	39.94	48.68
k-NN	58.05	40.75	47.88
SVM	70.64	53.29	60.75
DeepCNN	92.11	95.04	93.56
DeepCNN-SVM	96.27	93.79	95.01
DeepCNN-KNN	**98.74**	**96.94**	**97.83**

From TABLE III, proposed deep learning frameworks are performing better than the other classifiers. Also, our frameworks are giving results close to the existing frameworks.

From TABLE IV, performance measures such as precision, recall, and f-score are evaluated for the proposed frameworks. These measures give actual insights into the performance of proposed frameworks. DeepCNN, DeepCNN-SVM, and DeepCNN-KNN produce better results than the other and also the balance between precision and recall is better.

TABLE V: Comparison of proposed frameworks with existing frameworks

Framework	*Accuracy (%)*
Median filtering with CNN [10]	85.08
Ostu's thresholding with CNN [11]	96.29
DeepFish-SVM-aug-scale [17]	**98.64**
DeepCNN	98.65
DeepCNN-SVM	98.32
DeepCNN-KNN	**98.79**

TABLE V compares our proposed frameworks with the existing frameworks that have used Fish4Knowledge dataset with deep learning. All the proposed frameworks perform close to the existing best framework [17] with DeepCNN-KNN giving the best accuracy. One of the main reasons for this is, in our proposed frameworks, we used max-pooling after every convolution layer. Max-pooling is able to detect the presence of edges, which is very important in our fish species recognition. As a result, after every convolution layer, edges present in the image are preserved and passed on to the next layer. This helps in preserving all important features in the feature extraction phase.

Fig. 3 and Fig. 4 shows the accuracy and loss plots of the DeepCNN framework which is the backbone in our hybrid frameworks. In both, the accuracy and loss plots follow constantly increasing and constantly decreasing pattern which is acceptable. The main reason behind this behavior is with every epoch, the same fish image is passed on to the DeepCNN, to adjust the weights accordingly so as to minimize the loss and maximize the accuracy.

In convolutional neural networks, there is no fixed rule to find the number of filters and kernels in the network. These are found out by using trial and error depending on the task. In the proposed framework, 32, 64, 128 filters are found to give good results for the setup. Also, max pooling has given good results than average pooling, but it cannot be generalized as it is dependent on the task. In the fish species

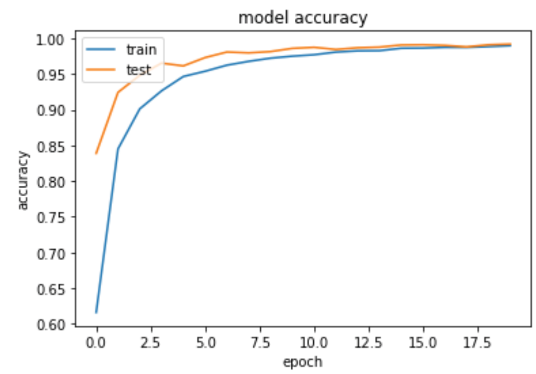


Fig. 3: Accuracy plot of DeepCNN on Fish4Knowledge dataset

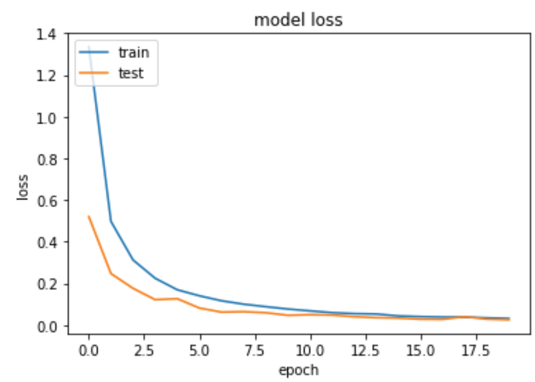


Fig. 4: Loss plot of DeepCNN on Fish4Knowledge dataset

recognition, highlighted features are important in the images, so max pooling is a good choice. Also, adding one more layer of convolution is not a good choice, because, the input image size is 48x48, so hardly any pixels are left if a fourth layer is added which may affect the performance of the framework.

V. Conclusion and Future Work

Underwater fish species recognition is a complex task due to itś uncontrolled habitat. We have proposed DeepCNN, DeepCNN-SVM, and DeepCNN-KNN frameworks, that are performing close and better than existing frameworks for underwater fish species recognition in terms of accuracy, precision, recall, and f-score. Our proposed framework DeepCNN-KNN gives an accuracy of 98.79%. There is still scope to improve the performance by preprocessing the image using image processing as well as other deep learning techniques such as such as transfer learning to enhance the performance. of the framework.

References

[1] Heithaus, M. R. and Dill, L. M., "Food availability and tiger shark predation risk influence bottle nose dolphin habitat use," in *Ecology*, vol. 83(2), 2002, pp. 480–491.

[2] Rova, A., Mori, G., and Dill, L., "One fish, two fish, butterfish, trumpeter:recognizing fish in underwater video," in *IAPR Conference on Machine Vision Applications*, 2007, pp. 404–407.

[3] Shafait, F., Mian, A., Shortis, M., Ghanem, B., Phil Culverhouse, F., Edgington, D., Cline, D., Ravanbakhsh, M., Seager, J., and Harvey, E., "Fish identification from videos captured in uncontrolled underwater environments," *ICES Journal of Marine Science*, vol. 73(10), pp. 2737–2746, 2016.

[4] Strachan, N. J. C., "Recognition of fish species by colour and shape," *Image and Vision Computing*, vol. 11, pp. 2–10, 1993.

[5] Strachan, N. J. C. and Kell, L., "A potential method for the differentiation between haddock fish stocks by computer vision using canonical discriminant analysis," *ICES Journal of Marine Science*, vol. 52, pp. 145–149, 1995.

[6] White, C. and Strachan, N. J. C., "Automated measurement of species and length of fish by computer vision," *Fisheries Research*, vol. 80, pp. 203–210, 2006.

[7] Storbeck, F. and Daan, B., "Fish species recognition using computer vision and a neural network," *Fisheries Research*, vol. 51, pp. 11–15, 2001.

[8] Nery, M., Machado, M., Padua, F., Carceroni, R., and Queiroz-Neto, J., "Determining the appropriate feature set for fish classification tasks," in *Proceeding of the XVIII Brazilian Symposium on Computer Graphics and Image Processing*, 2005.

[9] Chuang, M., Hwang, J. N., and Rose, C., "Aggregated segmentation of fish from conveyor belt videos," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 1807–1811.

[10] Jin, L. and Liang, H., "Deep learning for underwater image recognition in small sample size situations," in *OCEANS 2017-Aberdeen*, 2017, pp. 1–4.

[11] Rathi, D., Jain, S., and Dr. Indu, S., "Underwater fish species classification using convolutional neural network and deep learning," in *International Conference of Advances in Pattern Recognition*, 2017.

[12] Sun, X., Shi, J., Dong, J., and Wang, X., "Fish recognition from low-resolution underwater images," in 9^{th} *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2016.

[13] Yang, C. and Yang, M., "Fast direct super-resolution by simple functions," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 561–568.

[14] Chan, T., Jia, K., Gao, S., Lu, J., Zeng, Z., and Ma, Y., "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[15] Lin, M., Chen, Q., and Yan, S., "Network in network," in *International Conference on Learning Representations*, 2014.

[16] Blanc, K., Lingrand, D., and Precioso, F., "Fish species recognition from video using svm classifier," in *MAED'14 Proceedings of the* 3^{rd} *ACM International Workshop on Multimedia Analysis for Ecological Data*, 2014, pp. 1–6.

[17] Qin, H., Li, X., Liang, J., Peng, Y., and Zhang, C., "Deepfish: Accurate underwater live fish recognition with a deep architecture," *Neurocomputing*, vol. 187, pp. 49–58, 2016.

[18] Qin, H., Peng, Y., and Li, X., "Foreground extraction of underwater videos via sparse and low-rank matrix decomposition," in *ICPR Workshop on Computer Vision for Analysis of Underwater Imagery(CVAUI)*, 2014, pp. 65–72.

[19] Salman, A., Harvey, E., Jalal, A., Shafait, F., Mian, A., Shortis, M., and Seager, J., "Fish species classification in unconstrained underwater environments based on deep learning," *Limnology and oceanography, methods*, vol. 14, pp. 570–585, 2016.

[20] Fouad, M. M. M., Zawbaa, H. M., El-Bendary, N., and Hassanien, A., "Automatic nile tilapia fish classification approach using machine learning techniques," in 13^{th} *International Conference on Hybrid Intelligent Systems (HIS)*, 2013, pp. 173–178.

[21] Moniruzzaman, M., Islam, S., Bennamoun, M., and Lavery, P., "Deep learning on underwater marine object detection: A survey," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, 2017, pp. 150–160.

[22] Boom, J., Huang, P. X., Spampinato, C., Palazzo, S., He, J., Beyan, C., Beauxis-Aussalet, E., van Ossenbruggen, J., Nadarajan, G., Chen-Burger, J. Y., Giordano, D., Hardman, L., Lin, F., and Fisher, R. B., "Long-term underwater camera surveillance for monitoring and analysis of fish populations," in *Proceedings of International Workshop on Visual observation and Analysis of Animal and Insect Behavior (VAIB), in conjunction with ICPR*, 2012.

[23] Boom, J., Huang, P. X., He, J., and Fisher, R. B., "Supporting ground-truth annotation of image datasets using clustering," in 21^{st} *International Conference on Pattern Recognition (ICPR)*, 2012.

[24] Nowak, S., Lukashevich, H., Dunker, P., and Rüger, S., "Performance measures for multilabel evaluation: A case study in the area of image classification," in *Proceedings of the international conference on Multimedia information retrieval*, 2010, pp. 35–44.