



# Crosspooled FishNet: transfer learning based fish species classification model

Monika Mathur<sup>1</sup> · Diksha Vasudev<sup>1</sup> · Sonalika Sahoo<sup>1</sup> · Divanshi Jain<sup>1</sup> · Nidhi Goel<sup>1</sup> 

Received: 19 July 2019 / Revised: 7 July 2020 / Accepted: 16 July 2020 /

Published online: 22 August 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

Fish species classification is an important task for biologists and marine ecologists to frequently estimate the relative abundance of fish species in their natural habitats and monitor changes in their populations. Traditional methods used for fish species classification were laborious, time consuming and expensive. So, there is need for an automated system that can not only detect and track but also categorize fish as well as other aquatic species in underwater imagery, minimizing the manual interference. Absorption and scattering of light in deep sea environment leads to low resolution images making fish species recognition and classification a challenging task. Further, performance of traditional computer vision techniques tends to degrade in underwater conditions due to the presence of high background clutter and highly indistinct features of marine species. For such classification problems, Artificial Neural Networks (ANN) or deep neural network are being increasingly employed for improved performance. But the limited dataset of fish images makes it difficult to train such networks as they require huge datasets for training. Thus to reduce the requirement for a huge amount of training data, an algorithm using cross convolutional layer pooling on a pre-trained Convolutional Neural Networks (CNN) is proposed. The present paper focuses on the development of automatic system for classification, which can detect and classify fish from underwater images captured through videos. Thorough analysis on image dataset of 27,370 fish images gives a validation accuracy of 98.03%. The proposed method will be an efficient replacement to strenuous and time consuming method of manual recognition by marine experts and thus be advantageous for monitoring fish biodiversity in their natural habitats.

**Keywords** Underwater images · Convolution neural network · Fish species · ResNet · Cross convolutional layer pooling

## 1 Introduction

Object recognition and classification is one of the most prominent topic of research for underwater environment [36, 37]. But underwater imagery is a strenuous task, which

---

✉ Nidhi Goel  
nidhigoel@igdtuw.ac.in

requires the high-resolution camera equipment to overcome scattering and absorption of light in ocean water [26]. Marine biologist and ocean researchers are interested in observing the health and status of fish aggregation targeted by fisheries, fish population of the particular species and information about the corresponding abundance of fish and their distribution. Climatic changes due to global warming, weaken the ability of the ocean and coasts to provide the critical ecosystem for marine species thus increasing the pressure on fish and their habitats. Hence, there is a need for dependable as well as cost-effective methodologies for regular surveillance of fishes throughout their habitats.

Earlier analysis and classification of fish species is done manually that involves destructive and time consuming methods like capturing of fishes or visual census by deep sea divers [16]. Also, the other traditional approach for classification makes use of precisely devised low-level features [35]. These features give only satisfactory performance as they lack generalization ability. To overcome these issues there is an increasing requirement for a system which can effectively and quickly recognize and identify fishes without human intervention.

Deep learning networks like Convolutional Neural Networks (CNN) can be the solution for image recognition tasks provided the availability of large datasets [7]. But it is difficult to get sufficient sample images for the training of CNNs for underwater tasks since the acquisition of underwater images is a costly and tedious process. Further, due to the complexity of the marine environment, the images obtained are noisy leading to low contrast background, deteriorative edges and blurred details. This, in turn, renders underwater image recognition to be quite arduous task.

This paper aims at finding the solution to the above mentioned problem of fish classification with high accuracy. In case of fish classification, limited training datasets pose a problem of overfitting which reduces the ability of the model to generalize outside the original dataset. The proposed paper solves the problem of limited training data with the help of transfer learning. Making use of a pre-trained network can reduce the need of training a deep convolutional neural network from scratch. Thus, reducing the requirement for a large amount of training data as well as alleviates the need for high computational power. So, an automatic system which uses crosspooling for fish classification is proposed in this paper.

Rest of the paper is organised as follows: Section 2 gives a brief overview of related work as Literature Survey. Section 3 elaborates proposed work along with cross-convolutional layer pooling. Section 4 discusses the results and comparative analysis. Finally, this paper is concluded in Section 5.

## 2 Literature review

Image classification in underwater environments is one of the most challenging problems in the area of underwater research due to complex environment conditions (e.g. darkness, water pressure, currents, cluttered object [5] etc.). Manual methods of classification requires an enormous amount of time and labour; hence, Artificial Neural Networks are being increasingly employed to increase performance in such classification problems, at the same time decreasing manual effort.

The earlier techniques used for fish classification were generally carried out in constrained environment. Technique proposed by Strachan et al. [31, 35] for fish species classification was only successful on images captured in restrained environments. Such techniques included classification using information about the shape and color of dead fish samples in controlled backgrounds. Daan et al. [30] proposed another technique that utilized

the physical features of certain species like length, height, and thickness of the fish. These features were measured using laser light for 3D modeling of fish species. Larsen et al. [12] also derived shape and texture features for three fish species and classification is done with Linear discriminant analysis (LDA) under constrained condition. Fish species classification in constrained environment requires capturing of fishes from their natural habitats and performing experiments on dead fish samples in the laboratory. These processes are not only time consuming but they also pose threat to the fish population.

To overcome these issues of constrained environment, various work have been done in unconstrained underwater environments. Spampinato et al. [29] and Mori et al. [24] proposed two classical methods for classification of fish based on their textural and physical features in unrestrained environments. While these techniques gave a satisfactory performance on underwater images, they work well only for fish species having rich and clearly distinguishable textural features. Further, Tharwat et al. [33] proposed a robust biometric-based approach for fish species classification in unconstrained environment. This biometric based approach collects images for four fish species from different distances and with various illumination conditions. Finally, Weber's Local Descriptor (WLD) is used for colour and texture analysis and classification is done by Adaboost classifier.

But fish classification in unconstrained underwater environments is a tough and demanding task because images and videos obtained in these unrestrained environments are characterized with excessive turbidity, large variations in luminosity and confusion between target and background due to a complex environment. Another challenge in underwater image classification is inter-class similarity i.e. the resemblance in size, color and textural features of fish species from different classes and intra-class dissimilarity which is dissimilarity among the fish species of the same class. These similarities and dissimilarities are mainly because of the various orientations in which freely moving fishes are captured.

To further improve the classification accuracies in complex unconstrained environment, various machine learning based algorithms [17, 19] are proposed for underwater image recognition and categorization. Rodrigues et al. [23] proposed five different schemes for the species recognition based on combination of three feature extraction methods (SIFT, PCA and PCA+SIFT+VLAD), three data clustering algorithms (k-means, aiNet and ARIA) and three input classifiers (SIFT, k-means and k-NN). Results show that these schemes are less time consuming and cheap [23].

Li et al. [13] proposed a fast neural network named as Regions with Convolutional Neural Networks (RCNN) for fish species recognition. RCNN shows accuracy of 81.4% for 24277 fish images of 12 species. Salman et al. [25] adopted a CNN approach to exhibit the efficacy of large datasets for classification of fish species in unrestrained environments and achieves an average classification accuracy to 90%. Qin et al. [18] also proposed deep learning based method for feature extraction of fish images. PCA Technique is used on convolutional layers for feature extraction. Further, binary hashing and block wise histogram are used for nonlinear and the feature pooling layer respectively. Finally, classification is done using linear SVM classifier with an accuracy of 98.64%.

Jin et al. [10] proposed an algorithm that would classify fish images even with small dataset by fine tuning the pre-trained AlexNet model. This framework employed a novel median filter for preprocessing the images and achieves an accuracy of 85.50%. Siddiqui et al. [27] proposed a CNN based method for fish species classification using transfer learning algorithm. ResNet-152 is used as pretrained network and classification is done using SVM classifier with an accuracy of 94.3%. Allken et al. [1] also deployed a deep learning

neural network to automate the fish species classification task. Images in the dataset are captured and simulated by Deep Vision Trawl Camera System. Inception-3 network is fine-grained for the task of training and classification accuracy of 94% is achieved using softmax classification layer.

Further, the hybrid solution for fish species classification in unconstrained underwater environment is proposed by Jalal et al. [9], which combined YOLO deep neural network with Gaussian mixture and optical flow models. This hybrid model achieves the classification accuracies of 91.64 % and 79.8 % on two different datasets. Dassantos et al. [2] proposed a CNN based fish classification model which recognize the fish species along with its family and order to improve the classification of species with similar characteristics in Pantanal region. Fish species classification model proposed by Rauf et al. [21] also uses deep learning framework consisting of 32 deep layers, for extracting discriminating and valuable features from the image. Further, VGGNet architecture is employed to improve the accuracy by adding four convolutional layers to the training of each level in the network. Villon et al. [34] compared the performances of various CNN based model with human ability in terms of speed and accuracy for fish species classification problem. Results indicate that CNN based methods achieved an accuracy of 94.9%, whereas identification by humans was only 89.3% correct.

Most of above methods use large datasets for training the networks and the requirement of large dataset in case of fish species classifications is generally fulfilled by data augmentation techniques as collection of underwater fish images is a tough task due to complex underwater environment. Data augmentation artificially introduces the realistic and unrealistic distortions in the present data, leading to increase in the samples present in the dataset. This increased number of images help in better training of the model, thus, positively impacting the training and test accuracy in most of the cases. On the other hand, data augmentation simultaneously reduces the network robustness and thus decreases its chance for practical implementation.

Thus, the proposed model uses ground-truth database from Fish4Knowledge project for training the model without any data augmentation to achieve robustness and suitability for real time application. Further, the proposed model also reduces the requirement for a large amount of training data by employing transfer learning. This paper proposed the cross convolutional layer pooling algorithm by employing a pre-trained deep CNN for feature extraction. Finally, the classification is done on the basis of high level features extracted by cross convolutional layer pooling, using the softmax activation layer, which is a non-linear version of multinomial logistic regression. The next section presents the proposed algorithm along with detailed description of dataset, implementation environment, feature extraction through cross convolutional layer pooling algorithm and training specifications.

### 3 Proposed work

In fish species classification problem, insufficient training datasets pose overfitting and makes the model incapable of generalizing outside of the original dataset. However, the problem of limited training data can be solved with the help of transfer learning by reducing the requirement for a large number of sample images for training as well as decreasing the computational power. Hence, a fine-grained classification model using cross-convolutional layer pooling on a pre-trained CNN is presented in this paper.

### 3.1 Dataset

The effectiveness of the proposed model is evaluated on the Ground-Truth dataset provided by Fish4Knowledge (F4K) project [3]. This dataset has been obtained from underwater video sequences acquired from the deep sea. It contains 27,370 verified images of 23 fish species. The dataset included manually labeled fish species by following instructions from marine biologists. The fish species along with the number of fishes in each category is depicted in Table 1. It can be observed that dataset is highly imbalanced as the ratio between the availability of most frequent and least frequent species is around 1000 times. There is significant variation in the orientation and position as well as the colors and textures of fishes in the datasets. Figure 1 indicates the fish images for each of the species present in the dataset. For evaluation of proposed framework, all the 23 species of dataset have been considered. The given dataset is split into training dataset of 21,896 images and validation dataset of 5,474 images.

The dataset contains the images of varying sizes and are being resized to 224-by-224, as it is default input image size for ResNet-50 network. The Resizing process does not incur any loss of information present in the images. As the data provided by Fish4Knowledge project is acquired for research purposes only and is distributed without any testing. The image capturing conditions were also not standardized or approved in any way. Varying conditions persisted, captured image quality was subject to the camera quality which is clearly not standardized in the dataset. Since the main objective of this research work is to propose a system which would be robust enough to be implemented in real scenario. So, it is important that images used should be as realistic as possible. But synthetically augmenting data which doesn't conform to any set standards or adhere to any quality criteria, would have further degraded the data quality. The non standardized capturing conditions would have already introduced enough variations in the images of dataset. Thus, the proposed model uses the database without data augmentation to achieve the foremost objective of robustness and suitability for real life application and implementation. Further, the random images from internet (Google Images) has also been used for testing purpose, apart from the validation dataset.

**Table 1** The fish species and its samples from F4K dataset

Species	Samples	Species	Samples
<i>Dascyllus Recticulatus</i>	12112	<i>Plectroglyphidodon Dickii</i>	2683
<i>Chromis Chrysur</i>	3593	<i>Amphiprion Clarkia</i>	4049
<i>Chaetodon Lunulatus</i>	2534	<i>Hemigymnus Melapterus</i>	42
<i>Pomacentrus Moluccensis</i>	181	<i>Zebrosoma Scopas</i>	90
<i>Neoniphon Samara</i>	299	<i>Scolopsis Bilineata</i>	49
<i>Neoglyphidodon Nigroris</i>	16	<i>Lutjanus Fulvus</i>	206
<i>Myripristis Kuntze</i>	450	<i>Scaridae</i>	56
<i>Acanthurus Nigrofuscus</i>	218	<i>Chaetodon Trifascialis</i>	190
<i>Pempheris Vanicolensis</i>	29	<i>Zanclus Cornutus</i>	21
<i>Hemigymnus Fasciatus</i>	241	<i>Canthigaster Valentine</i>	147
<i>Siganus Fuscescens</i>	25	<i>Balistapus Undulates</i>	41
<i>Abudefduf Vaigiensis</i>	98		

							
Siganus Fuscescens	Pempheris Vanicolensis	Chromis Chrysur	Amphiprion Clarkia	Chaetodon Lunulatus	Chaetodon Trifascialis	Myripristis Kuntze	Acanthurus Nigrofuscus
							
Scolopsis Bilineata	Zebrasoma Scopas	Abudefduf Vaigiensis	Canthigaster Valentine	Pomacentrus Moluccensis	Neoniphon Samara	Hemigymnus Melapterus	Lutjanus Fulvus
							
Hemigymnus Fasciatus	Balistapus Undulate	Plectroglyphidodon Dickii	Zanclus Cornutus	Neoglyphidodon Nigroris	Scaridae	Dascyllus Reticulatus	

**Fig. 1** Examples of 23 fish species from F4K dataset

## 3.2 Implementation

Absence of adequate amount of labeled data for classification leads to overfitting of the model on the training images with good training accuracy and poor classification accuracy on validation images. To overcome this problem of limited data, the method of transfer learning is advantageous. In order to accomplish transfer learning in the proposed method, ResNet 50 model [8] is used for feature extraction prior to classification. It is a 50 layered deep residual network that has been trained on huge datasets containing millions of images of hundreds of objects provided by the huge benchmark ImageNet [4] dataset. This enhances the ability of the network to learn specific features from a wide range of images. The ResNet-50 model accepts images of size 224-by-224 as input and uses skip connections to alleviate the vanishing gradient problem. Features extracted from ResNet-50 model acts as input to the model and finally classification is done using Softmax activation layer.

### 3.2.1 Local feature extraction

Pre-trained CNN (i.e. AlexNet [11], VGG [28], Resnet [8] etc.) for feature extraction meets the prerequisites for the task of image classification. Instead of training and implementing CNN from scratch, proposed model uses Resnet-50 for feature extraction. It extracts features by considering the available activations just before the fully connected layer of the pre-trained model. Feature vector in machine learning model is carried by these activations which can further classify it by learning. Initial stages of implementation uses VGG16 model as feature extractor for fish images in the dataset. Owing to lower classification accuracy by VGG16 model, later ResNet 50 has been used for local features extraction through cross-pooling to improve the accuracy.

### 3.2.2 Cross convolutional layer pooling algorithm

The cross convolutional layer pooling method utilizes the convolutional layer activations from higher layers of a pre-trained network as parts annotations [14, 15] and activations from lower convolutional layer as local features. The annotations produced by the convolutional layer activations are bound to be unsatisfactory as compared to the ones provided by experts. But the annotations provided by a large number of feature maps of higher layers are comparable and many a time even better than manually prepared parts annotation. In Cross-layer pooling algorithm, activations are required from two successive convolutional layers lying in same spatial dimensions for point-wise multiplication operation. Activations from lower convolutional layer provide local features whereas activations from higher convolutional layer provide parts annotation [14]. Therefore, weights from higher level activations are used to pool features from the lower layers. Figure 2 represents the cross convolutional layer pooling using two consecutive convolutional layers.

So, Cross-Convolutional Layer Pooling is a method for extracting image representations from successive convolutional layers. It can extract significant details from convolutional layers and comprises of the following two steps:

1. Extracting local features from lower convolutional layer.
2. Pooling these local features extracted in step1 by employing activations from the higher convolutional layer as guidance.

In this manner, single pooling channel is created for each detected region. This special type of pooling performed in two steps wherein first the activations from the two successive layers undergo point-wise multiplication followed by summing these pooling to obtain a feature vector for each feature map. These feature vectors from different pooling channels are then concatenated to obtain the final image representation. Finally, the image representation extracted from cross-layer pooling is shown by (1).

$$P^n = [P_1^n, P_2^n, \dots, P_K^n, P_{M(n+1)}^n] \tag{1}$$

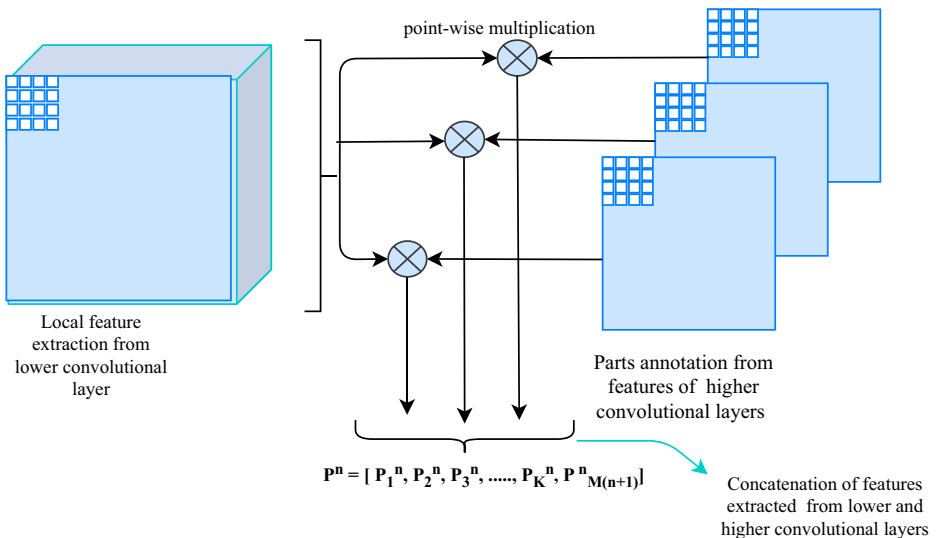


Fig. 2 Cross convolutional layer pooling using two consecutive convolutional layers

here,

$$P_K^n = \sum_{i=1}^{N_i} x_i^n y_{i,K}^{n+1} \quad (2)$$

In (1) and (2),  $P^n$  represents the cross-pooled feature of the  $n^{\text{th}}$  convolutional layer. The cross-pooled feature is obtained by combining the pooled feature of each pooling channel, i.e.  $P_K^n$ ,  $K = 1, \dots, M_{n+1}$ .  $x_i^n$  represents the  $i^{\text{th}}$  local feature in the  $n^{\text{th}}$  convolutional layer. The feature maps of the  $(n + 1)^{\text{th}}$  convolutional layer are obtained by convolving the feature maps of the  $n^{\text{th}}$  convolutional layer with a  $m \times n$  sized kernel.  $y_i^{n+1}$  denotes the feature vector of  $(n + 1)^{\text{th}}$  layer in the  $K^{\text{th}}$  dimension which is used to weight local features  $x_i^n$  in the  $K^{\text{th}}$  pooling channel. The proposed model implies various combinations of different layers of ResNet 50 network for feature extraction based on Hit and Trial method. Detailed description of different layer combinations and their associated accuracies has been shown in Table 3 in result section. Finally, Feature extraction in the proposed model has been done by cross pooling layer 154 and layer 157 of ResNet50 pre-trained model, which are then used for further training and classification. The classification is done by applying the softmax activation in the final layer of the model.

### 3.2.3 Model training

Once the crosspooled features are extracted using the ResNet50 pre-trained model, the classifier model is then trained on these extracted features of fishes in the training dataset. The training is conducted using Keras, which is a high-level API for building neural networks and runs on top of Tensorflow or Theano. Before training the images are resized to 224-by-224 as it is default input image size for ResNet-50. The parameter specifications for the training experiments using NVIDIA Quadro K2200 GPU are as follows:-

- 21,896 training images and 5,474 validation images.
- Each run cycle comprises of 300 epochs.
- The Batch size is fixed to 32.
- Learning rate is set to 0.001.
- Tanh is used as activation function.
- Adamax is used as optimizer.

Finally, the classification is done on the basis of high level features extracted by cross convolutional layer pooling, using the softmax activation layer. Due to the use of cross convolutional layer pooling for feature extraction, the proposed model is named as Crosspooled FishNet. The proposed Crosspooled FishNet made predictions on validation data from training dataset and achieves the accuracy of 98.03%. The next section presents the results that are achieved by training and testing of the proposed model. The obtained results have been compared with other state of art methods and are also discussed in next section.

## 4 Results and comparative analysis

Crosspooled FishNet is evaluated using different parameters as well as comparison with various state of art models. VGG-16 and ResNet-50 pre-trained networks are experimented for the purpose of cross-pooling. Both models are trained for 100 epochs each for feature extraction. The accuracies are calculated using different combinations of activation functions, optimizers and classifiers for VGG-16 (for layers 16 and 17) and ResNet-50 (for layers 154



and 157) networks. The results have been depicted in Table 2 and it can be observed that maximum accuracy of VGG-16 network is 83.34% for tanh activation function, adamax optimizer and softmax classifier whereas Resnet-50 gives the accuracy of 98.03% for same combination of activation function, optimizer and classifier. This is due to fact that ResNet-50 has more number of hidden layers, thus creating deep representations at every layer and learning more abstract features of the input. So, ResNet-50 performs better than VGG-16 and thus used in proposed model for feature extraction.

The proposed framework implies various combinations of different layers of ResNet-50 and VGG-16 network for feature extraction based on Hit and Trial method. Table 3 shows the accuracy achieved using various combinations of layers for cross pooling of pre-trained VGG-16 and ResNet50 networks. It can be observed that layer 15 and 17 of VGG-16 network gives an accuracy of 85.95% which is better than other combinations of layers like layer 15-16 and layer 16-17. Similarly various combinations of layers for Resnet-50 have also been used to improve the accuracies. Layer 154 and 157 of ResNet-50 gives an accuracy of 98.03% for validation dataset which is best among all other combinations. Thus it can be inferred from the Table 3 that layer 154 and layer 157 of ResNet-50 is the best combination for cross-pooling along with tanh as activation function, softmax as classifier and adamax as optimizer.

The results of prediction for four fish species samples are shown in the Fig. 3 with their actual name and the predicted names by Crosspooled FishNet. To check the efficiency of proposed Crosspooled FishNet model, random fish images from World Wide Web were downloaded. These images were given to the proposed model and Fig. 3 shows that proposed model performs effectively on validation images as well as the unseen random images (Google Images) from internet. Taking random images from the internet was only for application purpose and evaluation of proposed model on unseen images. These images were not used to calculate the test/validation accuracy.

**Table 2** Accuracies calculated using different parameters for VGG-16 and ResNet-50 models

Models	Pre-Trained Model	Activation	Classifier	Optimizer	Accuracy
Model a	VGG16	tanh	sigmoid	rmsprop	Does not converge
Model b	VGG16	tanh	softmax	sgd	52.07 %
Model c	VGG16	relu	softmax	rmsprop	55.82 %
Model d	VGG16	tanh	softmax	nadam	60.67 %
Model e	VGG16	tanh	softmax	adam	68.93 %
Model f	VGG16	tanh	softmax	rmsprop	80.79 %
Model g	VGG16	tanh	softmax	adadelta	82.71 %
Model h	VGG16	tanh	softmax	adamax	83.34 %
Model i	ResNet50	relu	sigmoid	rmsprop	Does not converge
Model j	ResNet50	tanh	sigmoid	rmsprop	Does not converge
Model k	ResNet50	tanh	softmax	sgd	86.56 %
Model l	ResNet50	relu	softmax	rmsprop	9.8 1%
Model m	ResNet50	tanh	softmax	adam	66.82 %
Model n	ResNet50	tanh	softmax	adam	88.25 %
Model o	ResNet50	tanh	softmax	rmsprop	92.58 %
Model p	ResNet50	tanh	softmax	adadelta	94.49 %
Model q	ResNet50	tanh	softmax	adamax	98.03%

**Table 3** Accuracies calculated by cross pooling different layers of VGG16 and ResNet 50 models

Model	Cross-pooled layers	Accuracy
VGG16	Layer 15 and layer 16	78 %
VGG16	Layer 16 and layer 17	83.34 %
VGG16	Layer 15 and layer 17	85.95 %
ResNet50	Layer 154 and layer 155	92.05 %
ResNet50	Layer 155 and layer 156	93.03 %
ResNet50	Layer 156 and layer 157	94.40 %
ResNet50	Layer 155 and layer 157	97.69 %
ResNet50	Layer 154 and layer 157	98.03 %

Performance evaluation of Crosspooled FishNet classification model on a given validation dataset is done using confusion matrix. A confusion matrix is a table, which uses a set of test data for which the true values describe the performance of a classification model. Each row of the matrix represents the number of species in an actual class while each column represents the number of species in a predicted class. Confusion Matrix is a performance measurement for machine learning classification and it depicts if the system is confusing between the classes.

The confusion matrix also mentions the true positive (TP) cases, true negative (TN) cases, false positive cases (FP) and false negative cases (FN) of the participating categories of classes, which are further used to calculate important parameters such as precision and recall. Diagonal elements in the confusion matrix of Fig. 4 represents the true positives (model correctly predicts the positive class) and other elements except diagonal elements represents the false positives (model incorrectly predicts the positive class). Each of the 23 classes included in the analysis of confusion matrix has 15 samples. A scale on the right hand side of the graph shows the number of samples for each class and intensity of blue colour increases as number of samples increases.

It can be observed from the confusion matrix of Fig. 4 that performance of Crosspooled FishNet is quite good as maximum true positives are in dark shades of blue which inferred that there are more samples in true positive rather than false positives. For example, it can be observed from Fig. 4 that from 15 samples of Scaridae fish, 12 are true positive (correctly predicted) and 3 are false positive (incorrectly predicted as *Dascyllus Reticulatus*). Similarly, for *Neoniphon Samara* all 15 samples are true positives and 0 samples are false positive. Further, confusion matrix has been used for calculation of Kappa score, which is explained in detail in the later part of the results section.

Results of the Crosspooled FishNet are also verified using precision and recall for different species of fish demonstrated through the graphs. The precision in case of classifiers is defined as a ratio of correctly classified images to the total number of predicted images which lies in a particular class and is given by equation 3.

$$Precision = \frac{TruePositive}{TruePositives + FalsePositives} \quad (3)$$

On the other hand, Recall in case of classifiers is defined as a ratio of correctly classified images to total number of images which belongs to a particular class and is given by equation 4. The precision and recall of the participating classes have been plotted in Fig. 5. It can be observed that Precision is higher than recall in most of the cases, indicating that false

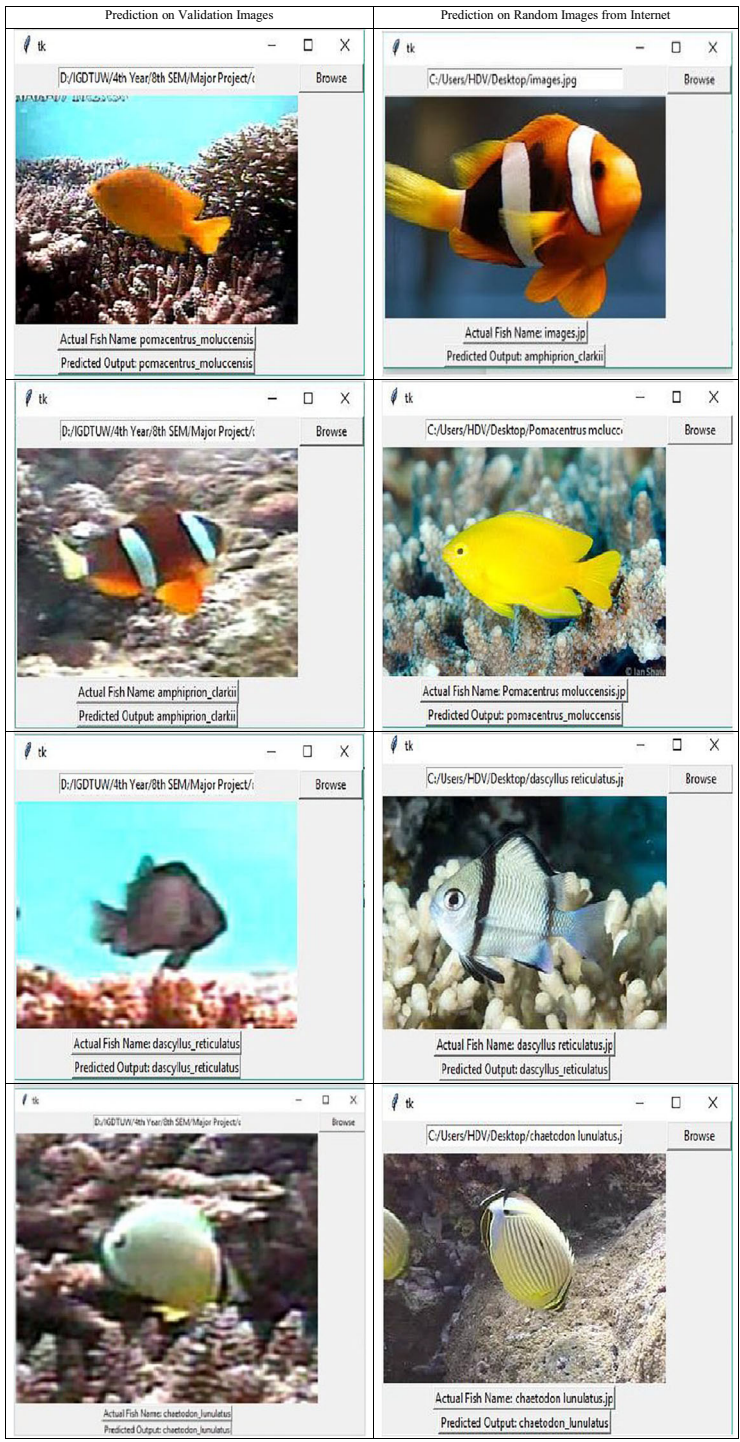


Fig. 3 Prediction accuracy of various fish samples both from validation and internet images

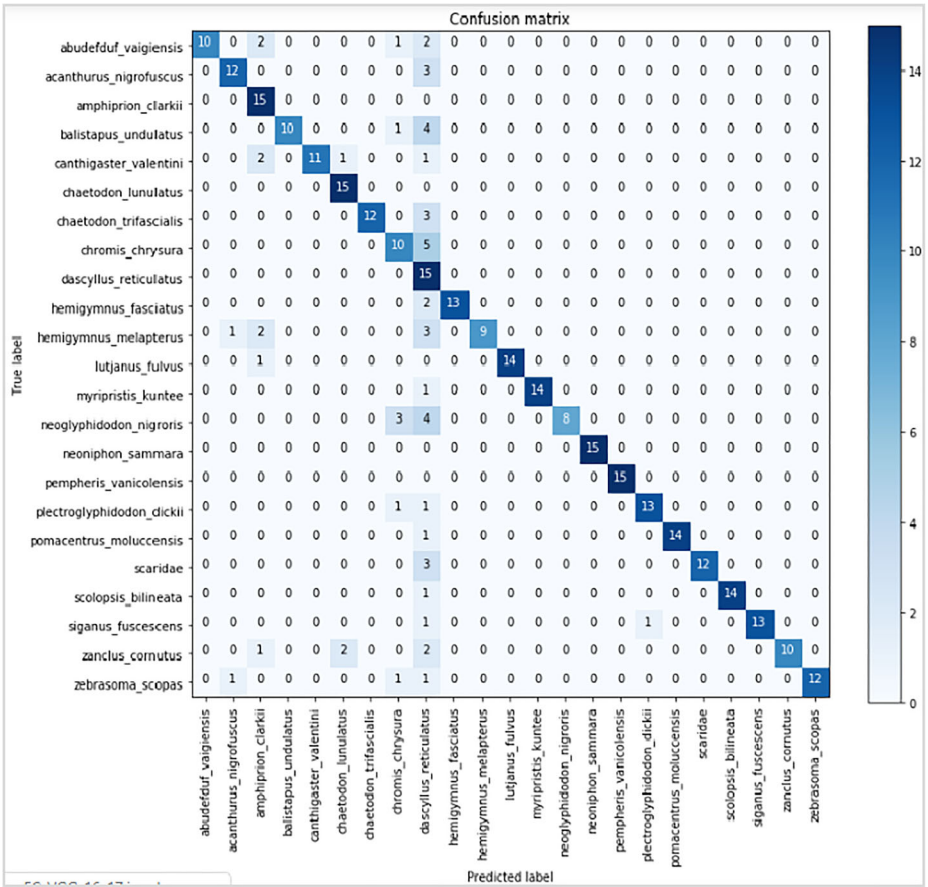


Fig. 4 Confusion Matrix for Crosspooled FishNet

negatives are more common than false positives in the proposed model. High value of precision graph shows that high proportion of predicted positives is truly positives indicating that the number of false positives is very low and Cross-pooled FishNet model is correctly classifying the species. However, in some case like Neoniphon Samara and Pempheris Vanicolensis, both the precision and recall plots have high values indicating that the classifier is better for these two classes and it gets almost all of the images of “Neoniphon Samara and Pempheris Vanicolensis” which are “Neoniphon Samara and Pempheris Vanicolensis” correctly.

$$Recall = \frac{TruePositive}{TruePositives + FalseNegatives} \tag{4}$$

Further, the performance of proposed model is also tested on the basis of F1 score and kappa score, which are generally used in classification problems with imbalanced datasets. F1 score is more realistic measure of a test’s performance by providing a balance between precision and recall. F1 score is calculated as harmonic mean of Precision and Recall values. Formula for F1-score is depicted in Equation 5. Precision, Recall and F1 scores for 23 species used in classifications are shown in Table 4. It can be observed from Table 4 that macro-averaged Precision, Recall and F1-score for proposed model are 0.92, 0.83 and

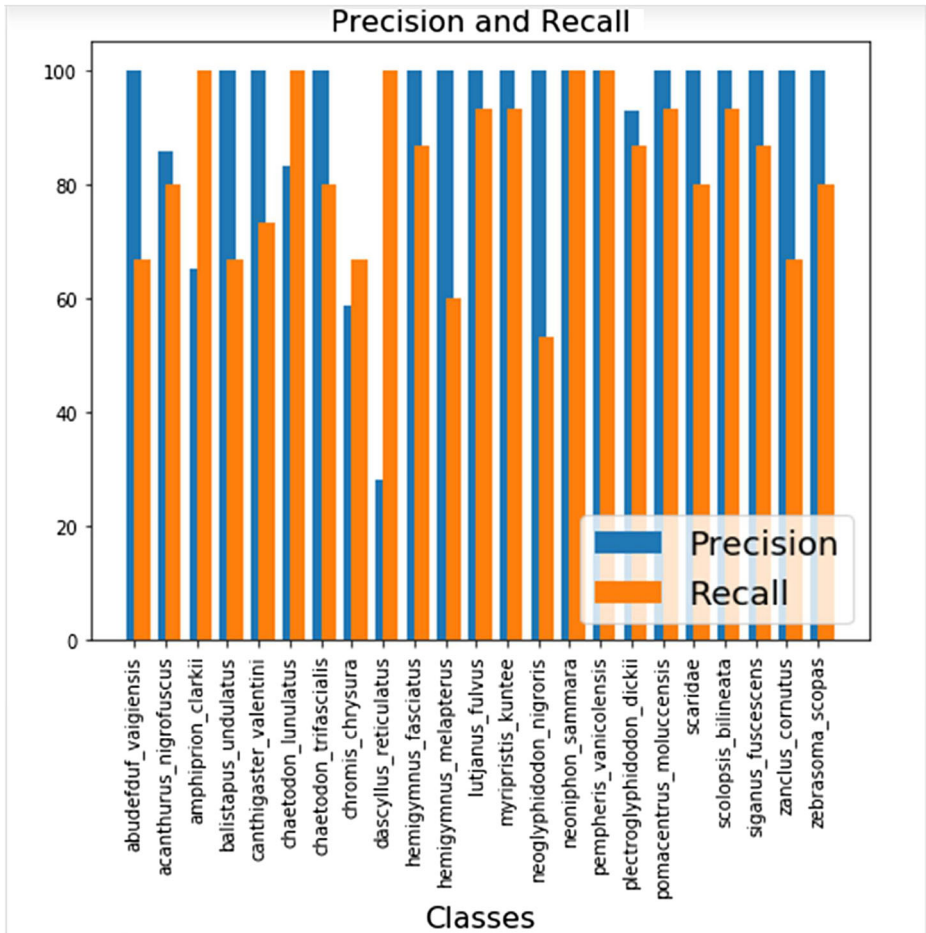


Fig. 5 Precision and Recall of 23 species on validation dataset

0.85 respectively, which indicates that performance of model is quite good as scores are approaching 1 i.e. maximum value for these measures.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

For further evaluation, Kappa score is also calculated for proposed Crosspooled FishNet model. Kappa score is a metric that compares an observed accuracy with an expected accuracy (random chances). Use of random chances with a random classifiers make kappa score more appropriate than using the accuracy for network’s evaluation. It can be calculated with the help of confusion matrix by using the formula shown in Equation 6. Here, observed accuracy is ratio of correctly predicted fishes in the entire confusion matrix to the total fish samples. Expected accuracy is the accuracy which would be achieved by any random classifier based on the confusion matrix. The expected accuracy is directly related to the number of instances of each 23 fish species, along with the number of instances where the predicted labels agreed with true labels. Observed and Expected accuracies for Crosspooled FishNet

**Table 4** Precision, Recall and F1 score for 23 fish species

Species	Precision	Recall	F1 Score	Species	Precision	Recall	F1 Score
<i>Abudefduf Vaigiensis</i>	1	0.66	0.80	<i>Myripristis Kunttee</i>	1	0.93	0.96
<i>Acanthurus Nigrofuscus</i>	0.86	0.8	0.82	<i>Neoglyphidodon Nigroris</i>	1	0.53	0.69
<i>Amphiprion Clarkia</i>	0.65	1	0.78	<i>Neoniphon Samara</i>	1	1	1
<i>Balistapus Undulates</i>	1	0.66	0.80	<i>Pempheris Vanicolensis</i>	1	1	1
<i>Canthigaster Valentine</i>	1	0.73	0.84	<i>Plectroglyphidodon Diekii</i>	0.92	0.86	0.89
<i>Chaetodon Lunulatus</i>	0.83	1	0.90	<i>Pomacentrus Moltuccensis</i>	1	0.93	0.96
<i>Chaetodon Trifascialis</i>	1	0.8	0.88	<i>Scaridae</i>	1	0.8	0.88
<i>Chromis Chrysur</i>	0.58	0.66	0.62	<i>Scolopsis Bilineata</i>	1	0.93	0.96
<i>Dasyllus Recticulatus</i>	0.28	1	0.44	<i>Siganus Fuscescens</i>	1	0.86	0.92
<i>Hemigymnus Fasciatus</i>	1	0.86	0.92	<i>Zanclus Cornutus</i>	1	0.66	0.80
<i>Hemigymnus Melapterus</i>	1	0.6	0.75	<i>Zebrasoma Scopas</i>	1	0.8	0.88
<i>Lutjanus Fulvus</i>	1	0.93	0.96	Macro-averaged Scores for 23 fish species	0.92	0.83	0.85

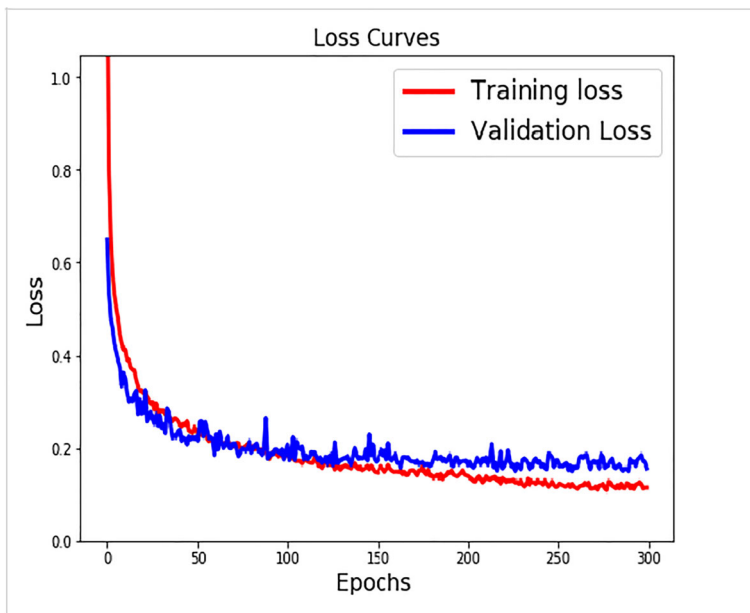
model are 0.8289 and 0.0434 respectively, which gives the Kappa score as 0.8211. This value of kappa metric is excellent score according to Landis and Koch [22] and Fleiss [6] studies.

$$KappaScore = \frac{ObservedAccuracy - ExpectedAccuracy}{1 - ExpectedAccuracy} \quad (6)$$

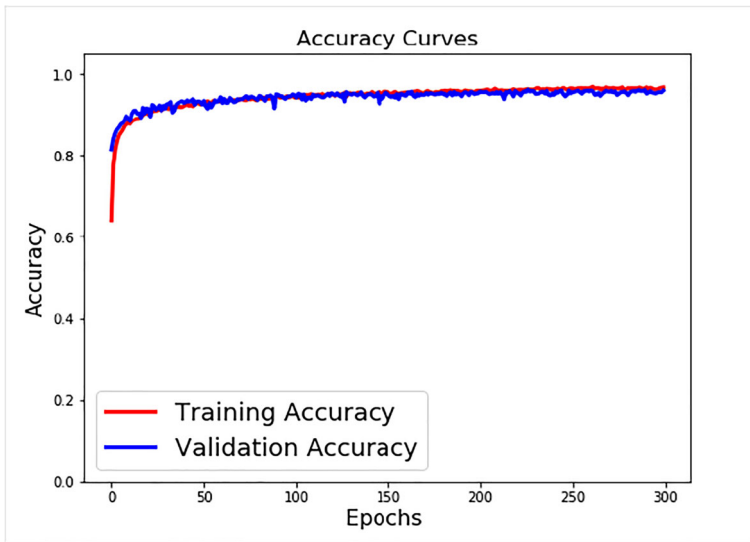
Figures 6 and 7 shows loss and accuracy curves respectively for both training as well as validation datasets for the Cross-pooled FishNet model. In Fig. 6, red line shows the loss curve for training dataset and blue line shows the loss curve for validation/test dataset. Both the curves are exponentially decreasing which shows the loss is decreasing as epochs are increasing from 0 to 300. Loss curves in Fig. 6 shows training and validation loss are decreasing to a point of stability with a minimal gap between the two final loss values, which is characteristic of a good fit loss curve. This depicts that proposed model is performing well.

Similarly in Fig. 7, red line shows the accuracy curve for training dataset and blue line shows the accuracy curve for validation/test dataset. Both the curves are exponentially increasing, which shows that the accuracy is increasing as epochs are increasing from 0 to 300. Validation accuracy in Fig. 7 is slightly lower than Training accuracy, which indicates that model is performing effectively without under or over fitting.

To further validate the potential of Crosspooled FishNet network, it is compared with other existing networks for fish classifications on the same dataset obtained from Fish4knowledge in terms of accuracy. F1 and kappa scores are calculated for the evaluation of the proposed Crosspooled FishNet but due to unavailability of these scores for the state-of-the-art methods, F1 and kappa scores are not included in the comparative analysis. The results of comparison are shown in Table 5. It can be observed that accuracy of 98.03% is achieved by the Crosspooled FishNet, which is good as compared to existing methods.



**Fig. 6** Loss curves for training and validation datasets



**Fig. 7** Accuracy curves for training and validation datasets

To make comparisons, state-of-the-art machine learning algorithms have been used as reference. Jin et al. [10] preprocess the underwater images using Median filter followed by classification through pre-trained network and achieves an accuracy of 85.08%.

An accuracy of 87.56% is obtained by training a softmax classifier with raw pixels [18]. DeepFish-Softmax-Augmented [18] also uses softmax classifier on augmented data by extracting only the foreground information from the underwater images. Further, principal component analysis (PCA) technique is used between two convolutional layers, which is followed by block-wise histograms in the feature pooling layer and binary hashing in the non-linear layer to achieve an accuracy of 92.55%. Rathi et al. [20] pre-processed the fish images using Gaussian Blurring, Morphological Operations, Otsu's Thresholding and Pyramid Mean Shifting to further feed them to a Convolutional Neural Network for classification. An accuracy of 96.29% is achieved. Alex-FT-Soft [32] uses AlexNet for transfer learning and softmax for fish species classification achieving an accuracy of 96.61%.

Crosspooled FishNet is also compared with some of the latest fish species classification networks [2, 9, 17, 19, 21, 34] and it has been observed that the proposed Crosspooled FishNet framework performs better in terms of accuracy. However, the use of different datasets by these latest fish species classification networks make this comparison invalid to incorporate in the present work.

**Table 5** Comparative analysis of Crosspooled FishNet with other state-of-the-art methods

Model	Accuracy (%)
Jin et al. [10]	85.05
Raw-Pixel Softmax [18]	87.56
DeepFish-Softmax-Augmentation [18]	92.55
Rathi et. al. [20]	96.29
Alex-FT-Soft [32]	96.61
Crosspooled FishNet (Proposed)	98.03



An accuracy of 98.03% achieved from Crosspooled FishNet outperform the state-of-the-art methods. It also demonstrates that fine-tuning an existing neural network is computationally less complex as there is no requirement of training the network from scratch, which is tedious and time consuming task. Such networks are also easier to train using relatively little labelled data i.e. only 27,370 images for Crosspooled FishNet instead of millions of images required for training a new model from scratch. So, transfer learned Crosspooled FishNet network is effective and accurate solution to fish classification problem in real time.

## 5 Conclusion

Due to global environmental changes, there is rapid pace in underwater explorations including the fish species classification and their census and surveys. To achieve this goal, automated fish species classification models are required. Proposed Crosspooled FishNet framework is a step forward for this problem with practical approach. The proposed model is automated as it uses deep neural networks and moreover it works on the limited data available for fish without compromising the test accuracy. The technique of cross-convolutional layer pooling derives discriminative features from layer154 and layer157 of ResNet-50 improves the performance of Crosspooled FishNet model. Result analysis through various evaluation parameters on image dataset of 27,370 fish images indicates that proposed model achieves an accuracy of 98.03% with high precision value as 0.92, F1 score as 0.85 and Kappa score as 0.82. Further, comparison with the state-of-the-art methods indicate that Crosspooled FishNet model is more accurate for fish species classification in complex underwater environment. The proposed framework can be used for monitoring fish biodiversity as well as for exploratory applications such as fisheries, for assessing the stock and species richness, examining the population distribution and status of the aquatic ecosystem. Proposed model can also be easily extended to other detection and classification applications.

## Compliance with Ethical Standards

**Conflict of interests** The authors declare that they have no conflict of interest.

## References

1. 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
2. Anderson Aparecido dos Santos AA, Gonçalves WN (2019) Improving pantanal fish species recognition through taxonomic ranks in convolutional neural networks. *Ecological Informatics* 100977:53
3. Boom BJ, Huang PX, He J, Fisher RB (2012) Supporting ground-truth annotation of image datasets using clustering. In: *Proceedings of the 21st international conference on pattern recognition (ICPR)*, pp 1542–1545
4. Deng J, Dong W, Socher R, Li L, Li K, Li F-F (2009) Imagenet: a large-scale hierarchical image database. In: *2009 IEEE computer society conference on computer vision and pattern recognition workshops (CVPR workshops)*, pp 248–255
5. Fan D-P, Cheng M-M, Liu J-J, Gao S-H, Hou Q, Borji A (2018) Salient objects in clutter: bringing salient object detection to the foreground. In: *The european conference on computer vision (ECCV)*, pp 186–202
6. Fleiss JL (1981) *Statistical methods for rates and proportions*, 2nd edn. Wiley, New York

7. Gandarias JM, García-Cerezo AJ, Gómez-de-Gabriel JM (2019) Cnn-based methods for object recognition with high-resolution tactile sensors. *IEEE Sensors J* 19(16):6872–6882
8. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
9. Jalal A, Salman A, Mian A, Shortis M, Shafait F (2020) Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics* 101088:57
10. Jin L, Liang H (2017) Deep learning for underwater image recognition in small sample size situations. In: *OCEANS 2017 - Aberdeen*, pp 1–4
11. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
12. Larsen R, Olafsdottir H, Ersbøll BK (2009) Shape and texture based classification of fish species. In: *Image analysis*. Springer, Berlin, pp 745–749
13. Li X, Shang M, Qin H, Chen L (2015) Fast accurate fish detection and recognition of underwater images with fast r-cnn. In: *OCEANS*, pp 1–5
14. Liu L, Shen C, van den Hengel A (2015) The treasure beneath convolutional layers: cross-convolutional-layer pooling for image classification. In: *IEEE conference on computer vision and pattern recognition (CVPR)*
15. Liu L, Shen C, van den Hengel A (2017) Cross-convolutional-layer pooling for image recognition. *IEEE Trans Pattern Anal Mach Intell* 39(11):2305–2313
16. Mathur NGM (2019) Automatic fish species classification in underwater images using transfer learning. In: *International conference on signal processing, VLSI and communication engineering*
17. Nagaoka Y, Miyazaki T, Sugaya Y, Omachi S (2019) Automatic mackerel sorting machine using global and local features. *IEEE Access* 7:63767–63777
18. Qin JLH, Li X, Zhang C (2016) Deepfish: accurate underwater live fish recognition with a deep architecture. *Neurocomputing* 187:49–58
19. Qiu C, Zhang S, Wang C, Yu Z, Zheng H, Zheng B (2018) Improving transfer learning and squeeze- and-excitation networks for small-scale fine-grained fish image classification. *IEEE Access* 6:78503–78512
20. Rathi D, Jain S, Indu S (2017) Underwater fish species classification using convolutional neural network and deep learning. In: *Ninth international conference on advances in pattern recognition (ICAPR)*
21. Rauf HT, Ikram Ullah Lali M, Zahoor S, Shah SZH, Rehman AU, Bukhari SAC (2019) Visual features based automated identification of fish species using deep convolutional neural networks. *Comput Electron Agric* 167:105075
22. Richard Landis J, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
23. Rodrigues MTA, Freitas MHG, Pádua F, Gomes RM, Carrano EG (2015) Evaluating cluster detection algorithms and feature extraction techniques in automatic classification of fish species. *Pattern Anal Applic* 18(4):783–797
24. Rova A, Mori G, Dill LM (2007) One fish, two fish, butterflyfish, trumpeter: recognizing fish in underwater video. In: *IAPR conference on machine vision applications*
25. Salman A, Jalal A, Shafait F, Mian A, Shortis M, Seager J, Harvey E (2016) Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods* 14(9):570–585
26. Schettini R, Corchs S (2010) Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing* 2010(1):746052
27. Siddiqui S, Salman A, Malik I, Shafait F, Mian A, Shortis M, Harvey E (2018) Automatic fish species classification in underwater videos: exploiting pretrained deep neural network models to compensate for limited labelled data. *ICES J Mar Sci* 75:1–16
28. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556:1–14](https://arxiv.org/abs/1409.1556)
29. Spampinato C, Giordano D, Salvo RD, Chen-Burger Y-HJ, Fisher RB, Nadarajan G (2010) 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
30. Storbeck F, Daan B (2011) Fish species recognition using computer vision and a neural network. *Fish Res* 51:11–15
31. Strachan NJC, Kell L (1995) A potential method for the differentiation between haddock fish stocks by computer vision using canonical discriminant analysis. *ICES J Mar Sci* 52(1):145–149
32. Tamou B, Ben A (2018) Nasreddine underwater live fish recognition by deep learning. In: *Image and signal processing*, pp 275–283

33. Tharwat A, Hemedan AA, Hassanien AE, Gabel T (2018) A biometric-based model for fish species classification. *Fish Res* 204:324–336
34. Villon S, Mouillot D, Chaumont M, Darling ES, Subsol G, Claverie T, Villéger S (2018) A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological Informatics* 48:238–244
35. White C, Svelling DJ, Strachan NJC (2006) Automated measurement of species and length of fish by computer vision. *Fish Res* 80:203–310
36. Zhao J, Cao Y, Fan D, Cheng M, Li X, Zhang L (2019) Contrast prior and fluid pyramid integration for rgb-d salient object detection. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 3922–3931
37. Zhao J-X, Liu J-J, Fan D-P, Cao Y, Yang J, Cheng M-M (2019) Egnnet: edge guidance network for salient object detection. In: The IEEE international conference on computer vision (ICCV), pp 8779–8788

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

## Affiliations

Monika Mathur<sup>1</sup> · Diksha Vasudev<sup>1</sup> · Sonalika Sahoo<sup>1</sup> · Divanshi Jain<sup>1</sup> · Nidhi Goel<sup>1</sup> 

Monika Mathur  
monika009phd0215@igdtuw.ac.in

Diksha Vasudev  
diksha024btece15@igdtuw.ac.in

Sonalika Sahoo  
sonalika004btece15@igdtuw.ac.in

Divanshi Jain  
divanshi026btece15@igdtuw.ac.in

<sup>1</sup> Department of ECE, IGDTUW, Kashmere Gate, Delhi-06, Delhi, India