Enhanced Deep CNN Models for Underwater Fish Classification

Ghazanfar Latif [1*], Kinza Waqar [2], Shifa Khaja [2], Sarah Khan [2], Majid Khan [1], Loay Alzubaidi [2]

[1] Department of Computer Science, Prince Mohamad bin Fahd University, Saudi Arabia.

[2] Department of Computer Engineering, Prince Mohamad bin Fahd University, Saudi Arabia.

[*] Corresponding Author Tel: +966-50-1057-422 E-mail: glatif@pmu.edu.sa

Abstract

Fish recognition has significant importance in the field of biodiversity and study of aquaculture. Conventional classification and recognition methods of underwater fish species poses a great difficulty due to noises in the background, image distortion, other water bodies present in images, occlusion and quality of the image. This paper proposes the idea of using a Convolutional Neural Network to classify and recognize fishes captured in underwater images. Two deep CNN models were proposed and trained with images from the ground truth dataset consisting of 24971 fish images belonging to 5 fish species which were collected from a video. The preprocessed images of the fish were used to adjust the trained deep neural network in order to analyze the classification performance. The experimental results demonstrated that the system performed with high accuracy, reaching 98.12% and the proposed method is able to recognize fish species effectively.

Keywords: Fish Recognition, Convolutional Neural Network, Image Classification, Deep Learning

1. Introduction

Around 70% of the earth's surface is covered by oceans and seas. This vast area makes the study of the underwater species challenging and there is a substantial risk for human to work in deep underwater environment. Although marine ecosystem has been explored by many researchers in various depths, the knowledge of the underwater world is limited. One of the fundamental approaches to study the marine ecological system is to observe the behavior of various fish species. Marine ecologists strive to gain knowledge about the natural underwater environment in order to study their behaviors and aid their preservation. Nowadays, fish recognition is used to monitor marine life in contemplation to study about the quantification of fish species and their attributes. It is also beneficial in commercial and environmental applications as it facilitates ecologists to solve issues regarding food availability. To acquire this biological information, the use of embedded video cameras has become widely common. Manually analyzing large quantities of underwater videos to find useful information proved to be time consuming and onerous. In addition, manual processing is intuitive, off putting and a costly task [1]. With advancement in technology, automatic processing such as computer vision or machine learning techniques can be used to recognize and classify the fish species accurately.

The aim of automatic fish classification is to detect and identify fish species. However, working in underwater environment brings a lot of difficulties and can be challenging for computer vision. For instance, the visibility underwater is limited and due to the movement of the fish species there is a rapid change in background and luminosity. It is a challenging problem due to noise, occlusion and overlap of the digital images. Hence, Convolutional Neural Networks (CNN) can be used for higher accuracy in image classification and recognition [2-3]. These networks contain various layers and are capable of accepting the complete image as an input and extract the prominent features for classification. Additionally, the system can detect and recognize large scale and multi oriented sample of fish species from a dataset belonging to unconstrained environment. Despite the higher accuracy, underwater fish classification based on machine intelligence is challenging as deep learning requires a huge chunk of training data and obtaining this data is complex and costly [4]. Moreover, the results of this method depend on the quality of the data available. It is challenging for any machine learning algorithms to investigate and produce quality results if the input contains noise, irrelevant and unreliable information.

The rest of the article is arranged with the following 2 discusses the recent literature review, section explains the proposed architectures of CNN, section 4 describes the experimental results and finally section 5 concludes the article.

2. Literature Review

Fish recognition is considered an important aspect of environmental studies. In this paper the focus is on the main applications of fish recognition systems such as protection of endangered fish species. Mehdi Chouiten et al. proposed an underwater fish recognition system using classification based on shape, color and texture using the fish analyzer [5]. The dataset consists of 2700 images of fishes. The accuracy achieved by this system is 99%. Pelletier et al. investigates two deep CNN architectures, namely AlexNet and GoogLeNet, in terms of their efficiency in classifying images of marine resources captured in unorganized scenarios [6]. The dataset used for classification was obtained from the Nature Conservancy which consisted 3,777 images. The conclusion showed that the best results were found using trained models and transfer learning, where GoogLeNet and AlexNet gave 94.01% and 96.01% accuracy respectively.

Spampinato et al. performed fish classification and detected fish trajectories to study their anomalous behavior [7]. The idea was initiated to assist marine biologists in conducting their studies regarding fish species. Fish classification was made possible by integrating two kind of features, texture and shape. Information about the texture was extracted using statistical moments of grey level histogram, spatial co-occurrence matrix and Gabor filtering. The shape features were derived using Curvature Scale Space transform and Fourier descriptors boundary histogram. In addition, by applying an affine transformation, the acquired images had the ability to be viewed in 3D. These methods were tested on 360 images of ten various species which resulted in an average accuracy of about 92%.

Peiqin Zhuang et al. proposed the idea of fish recognition using deep learning, open set classification, fine grained recognition and vision-text modeling [8]. The dataset is the WildFish that has 1000 fish species with 54,459 fish images in total. A contemporary open set fish classification along with open set deep learning is used for fish identification in realistic situations. To differentiate between any confused categories in a pair, a new fine-grained recognition technique is utilized. The test accuracy achieved using this method is 74.7%.

Hossain et al. proposed an automatic underwater monitoring system [9]. The objective of this monitoring process is to detect fishes, identify their species and to track the detected fish as well as observe their activities. This is achieved by using Pyramid Histogram of visual Words (PHOW) features along with SVM classifier, GMM based background subtraction method for detection and Kalman Filter to identify fishes. To carry out this experiment, data-set of more than 20,000 sample images belonging to 15 species from the CLEF 2015 is used. The accuracy achieved by the proposed system is 91.7% in still images of high resolution, 48.94 % in low quality videos for detecting, and 40.1% for in low quality videos for identifying fish species.

Hnin and Thidar had proposed a species recognizing system to identify the species based on specimens

[10]. This paper aims to study if the morphometric variation that exists among the different fish species will allow the automated taxonomic recognition of the species. The main approach is to utilize the Feature selection, combination theory, and pattern recognition for implementing the fish recognition system. The dataset had 1516 images that consisted of 20 classes of fishes. The accuracy achieved in this method is 99.1304% using the SMO (SVM) classifier, 79.1304% using the MLP (ANN) classifier and 92.1739% using the J48 classifier.

Jin and Liang investigates a framework for recognizing small sample size situations underwater [11]. The objective of this paper is to eliminate noise from fish images using an improved median filter and to provide an effective CNN based fish recognition framework. The dataset was obtained from ImageNet and the sample size was about 1000 images in each of 1000 categories. The accuracy achieved by the proposed framework was 85.50%. S´ebastien Villon et al. proposed the idea of automatically detecting and recognizing the coral reef fishes in underwater videos using two methods [12]. One, by performing the SVM (support vector machine) classifier and then extracting HOG (histogram of oriented gradients) features. Two, by using deep learning techniques. The dataset consists of 13000 fish images with 8 categories of fishes. The accuracy achieved using the HOG+SVM is 55% and the accuracy achieved using the CNN is 85%.

Siddiqui et al. proposed deep learning techniques for fish species classification of fine – grained images [13]. This paper aims to detect, track and classify fish species without human intervention in underwater videos. In order to avoid a large amount of training data, a cross-layer pooling algorithm that uses pre-trained CNN is proposed. SVM classifier is performed on a dataset containing 2209 images from 16 different species of fishes. The accuracy achieved by the proposed automatic method was 94.3%. Fouad, M. et al. has proposed fish detection and identification using the BAT optimization algorithms (BA) [14]. The dataset had 151 images that was used to identify a specific fish species - Nile Tilapia. The accuracy achieved using the BAT technique is 90% along with a positive result of reduced classification time.

Rodrigues et al. addresses the issue of automatic fish species classification [15]. The proposed automatic framework uses artificial immune systems and the experiment is carried out in two main categories with complex datasets. The first set consists of dataset of 162 images belonging to 6 species whereas the second set consists of dataset of 48 images belonging to 3 species. The accuracy achieved by the proposed framework was 92%. Miyazono et al proposes fish recognition using CNN and image processing techniques [16]. The dataset contains 1000 fish images from 50 various species of fishes. The accuracy achieved in fish recognition using this method is 91.4%.

In study [17], Ding et al. investigates several CNN architectures to identify which parameters give the desired accuracy and robustness. There are three CNN models who vary in terms of their layer type, pooling size, feature maps and filter size. The three models were trained using a sample of 22,437 images of 4 different species from the Ground – Truth dataset. The results demonstrated that as the number of iterations increased, CNN Model 2's accuracy increased as well, and reached about 96.52%. The parameters that brought about highest performance from Model 2 were smaller filter size and higher number of feature maps on the P type layer 7. Kratzert and Mader proposed the fish recognition in underwater video by using CNN where the image is used as an input [18]. The dataset used is 8099 fish species with 10 distinctive species and has acquired an accuracy of 93%. Additional information such as the length of the fish and the migration date have also been added to improve its accuracy.

3. Proposed Method

In this section, miscellaneous steps of the proposed system are discussed for recognizing fish species from underwater images. A convolutional neural network was used and trained with images from the Ground - Truth dataset. The first step was to regularize the training samples to a uniform size. Secondly, the training process involves selection of randomly selected training samples, to be inputted into the deep CNN models. Therefore, batches of fixed number of training samples were passed onto the convolutional, pooling and dense layers as small datasets each time. As the training sample is central to making the CNN model learn, detrimental effect caused by poor data needs to be prevented. This follows to the third step where weights were updated using back propagation for each small dataset batch. It halts the training upon approaching a precise number of iterations or when the error attains the threshold. Finally, the test data was inputted into the trained CNN models to provide an unbiased evaluation of the trained network and gain results of the performance of the recognition system.

3.1. Underwater Fish Dataset

The underwater ROV for fish recognition categorizes 5 fish species by collecting information from a video dataset of 24971 fish images. Although this dataset is uneven as information is spread such that the most common species are detected 1000 times higher than the least ones, its ability to identify fish species is still relatively accurate. The fish classification methodology used video processing software to obtain the dataset. Each species from the dataset was then labelled manually, as instructed by marine biologists. The design methodology used around 10 cameras that recorded 12 hours a day during daylight. The fish detection component discovers the fish in the image and follows the fish's trajectory as time progresses. The three important aspects that were evaluated in regards to the fish detection were bounding box, contour and trajectory of the fish. A tool was developed that made use of automatic algorithms to run on consecutive frames to give suggestions about the bounding box, contour and tracking. It also gives user the flexibility to change these suggestions. Specifically, for this dataset the bounding box and initial contour were found by background subtraction methods. This made it possible for user to verify the trajectory of fish. The dataset of the five clusters can be found in the table 1 below.

Table 2. Description of Different Underwater Fish Species with total number of Samples

#	Species	No. of Detections	No. of Trajectories
1	Dascyllus Reticulatus	12112	4240
2	Plectroglyphidodon Dickii	2683	1225
3	Chromis Chrysura	3593	1175
4	Amphiprion Clarkii	4049	1021
5	Chaetodon Lunulatus	2534	536

3.2 Convolutional Neural Network (CNN)

Convolutional Neural Network parameters play an integral role in optimizing the quality of the deep network [19]. They help in determining the over fitting or under fitting of the model on a specific dataset. In this section, various CNN parameters are discussed which were used to design the proposed CNN models. Firstly, as the output size and shape is impacted when the input image is passed through many convolutional layers, a kernel size more than 5x5 should preferably not be used.

Due to the iterative nature of gradient descent, updating the weights through a single epoch is not enough. A single epoch would give an under fitted graph, increasing the number would lead to optimal and further increasing would lead to overfitting curve. The right number of epochs chosen for a neural network increases with diversity of a dataset, however it should be noted that it cannot be increased to a level that will cause overfitting of the cost/weight graph. Large batch size increases the performance in terms of accuracy and regularization but should not cause overfitting of gradient descent curve as it is data dependent. Larger batch sizes are not able to converge as fast as the smaller batch sizes but have a faster progress in training. The number of hidden layers should be kept minimum as more number of layers would lead to an over fit and increase computation. Dropout layer, either hidden or visible results in the network learning independent representation thus it becomes capable of more generalizations and less sensitive to fitting its result to the training data provided. After training, the dropped neurons are replaced with their original weights.

The dense layer is affected by the activation functions in determining how fast the network can learn and gives accurate results much faster as well as prevents overfitting of the neural network. A nonlinear and differentiable activation function is able to back propagate in the network to compute gradients of error with respect to weights. With a finite range like in rectified linear unit (ReLU) activation function, gradient-based training methods tend to be stable but may result in problem of dead neurons. The leaky ReLU solves this issue by increasing the range of ReLU from negative infinity to infinity. The softmax function is used for the added advantage of the output probabilities range. The sum of all the probabilities equals to one as its range is from zero to one. When this function is applied for a multi-classification model, the target class has the high probability and it gives the probabilities of each class.

Furthermore, the optimization algorithm Adam is used for large data or parameters. Adam is a combination of the best properties from the AdaGrad and RMSProp algorithms that provides an optimization algorithm which is capable of handling sparse gradients on noisy problems. Using an appropriate combination of the above parameters, the goal of the research is to minimize cross-entropy loss function as much as possible for better performance.

3.3 Proposed CNN Models

The presented CNN architecture of Model#1 starts with input image of size 64x64 as shown in Figure 1 and then the first convolution layer uses 5x5 kernel sized 64 filters. A 3x3 sized max pooling layer is applied to reduce the features size to 3 times and then the ReLU activation function based layer is applied. The second convolution layer consisted of 3x3 kernel sized 48 filters and ReLU activation function. Following, 10% dropout is used for regularization. The third convolution is similar but reduces the filter size to 32 with a kernel size of 3 x 3. After which, another dropout of 10% is applied. Next, the fourth convolution layer applied is exactly the same as the third convolution layer. The next few steps consist of a MAX Pooling layer of size 2 x 2, a convolution of 32 filters, 3 x 3 kernel size and ReLU function. The images are then flattened to a one-dimensional vector so that they can be classified. Finally the three dense layers are used with one dimensional vector features reducing to 512, 256 and 128 respectively to reach to the out layer of 5 classes.

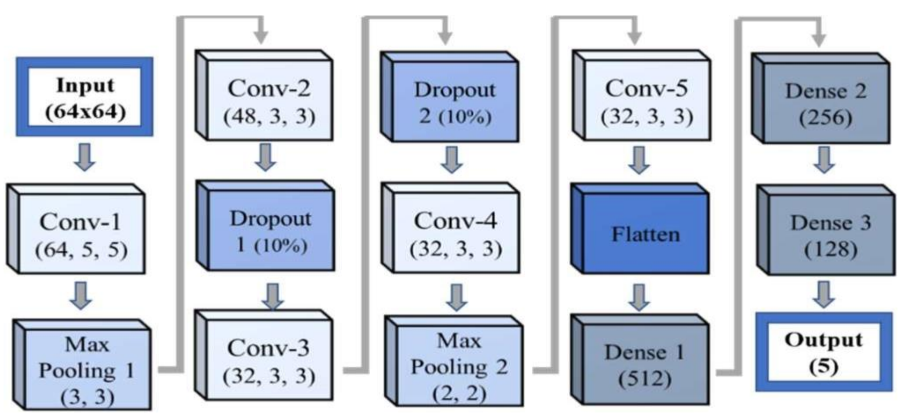

Figure 1. First proposed CNN Model for Fish Classification

In proposed Model 2, some changes were made as shown in Figure 2. The first convolutional layer of 5x5 sized kernel with 64 filters and ReLU layer is added, which is followed by a MAX Pooling layer of 2x2 size. Following, another convolution layer of 48 filters, 3x3 kernel size and ReLU activation is applied. After a 10% dropout and 2x2 MAX Pooling, the images are passed into the third convolution layer whose 32 filters, 3x3 kernel size and ReLU layer enable extraction of features. This is followed by another 10% regularization through the second dropout. The next steps are a series of convolutions followed by flatten layer. The fourth convolution layer consisted of 32 filters, 3x3 kernel size with ReLU and is followed by the fifth convolution layer of decreased number of filters i.e. 16 is applied. It has 3 x 3 kernel size and ReLU layer which is also followed by a flatten layer. They are then passed through three consecutive dense layers with 512, 256, and 64 features where the last dense layer applies a softmax function, making it possible for the image to be classified based on its probability distribution.

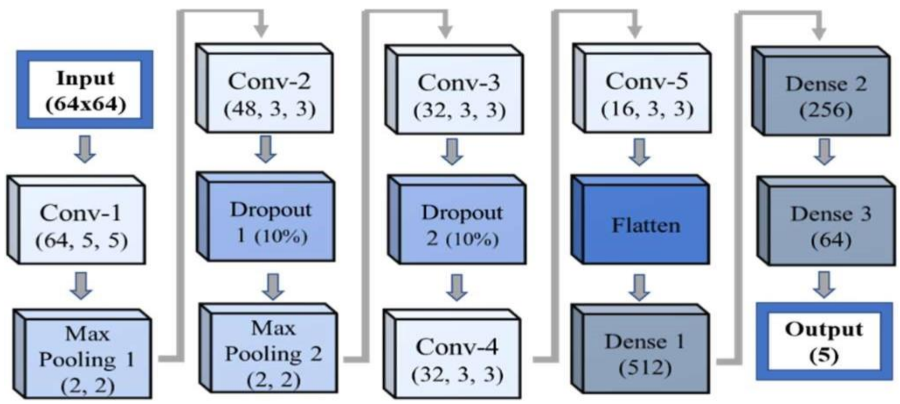

Figure 2. Second proposed CNN Model for Fish Classification

4. Experimental Results

The experimental results displaying the performance of the two proposed CNN architectures can be found in Table 2. Epoch sizes of 10, 25, 50 and 100 were used for different experiments with batch size of 50, 100 and 250. The experimental results table is divided into three columns with Validation Accuracy (%), Test Accuracy (%) and Computation Time (sec). Various design parameters were changed and their impact on the level of accuracy was evaluated. A total of 23 experiments were carried out on model 1 and model 2 as both implemented with the mutual objective of identifying an optimized architecture.

The experimental results demonstrated in Table 2 and the test accuracies accomplished are influenced by the number of epochs and number of batch size. The number of epochs directly affects the test accuracy; as the epoch size increases, the test accuracy increases as well. On the contrary, as the batch size increases, the test accuracy decreases. Primarily, it can be observed that the best test accuracy, although with a relatively higher computation time, is achieved when the epoch size is equal to the batch size. This is the same in the case of model 1 (98.122%) and model 2 (97.753%). The average test accuracies approximately achieved by model 1 is 95.64% and model 2 is 95.27%. The outcome resulted in an improved accuracy level than those found in recent literature. Furthermore, it provided a basis for utilizing CNN architecture in the use of underwater image classification.

Table 2. Summary of results of presented CNN Models for Fish Recognition

Model	Batch Size	Epochs	Validation Accuracy	Test Accuracy	Computation Time (Sec.)
1	100	10	98.848%	96.320%	39.954
1	200	10	98.878%	93.116%	34.951
1	50	25	99.585%	92.257%	122.076
1	100	25	99.609%	96.333%	94.494
1	200	25	99.550%	96.609%	81.450
1	50	50	**99.876%**	**98.122%**	**241.455**
1	100	50	99.733%	95.452%	187.926
1	200	50	99.802%	95.087%	160.581
1	50	100	99.545%	96.498%	482.914
1	100	100	99.703%	95.644%	376.348
1	200	100	99.956%	96.547%	319.560
2	50	10	99.125%	93.405%	54.621
2	100	10	99.125%	95.879%	44.610
2	200	10	99.016%	94.389%	39.380
2	50	25	99.575%	97.415%	131.543
2	100	25	99.609%	96.271%	105.916
2	200	25	99.797%	96.685%	92.170
2	50	50	99.782%	97.753%	262.581
2	100	50	99.842%	94.100%	210.820
2	200	50	99.738%	94.429%	181.103
2	50	100	99.807%	95.003%	527.902
2	100	100	99.985%	95.092%	421.459
2	200	100	99.822%	92.863%	360.943

5. Conclusion

Underwater fish classification based on machine intelligence has significant importance in the field of biodiversity and study of aquaculture. However, it is challenging as deep learning requires a large amount of training data and obtaining this data is complex and costly. This paper proposed two deep convolutional neural networks models for fish classification which were trained with images from the Ground - Truth dataset consisting of 24971 fish images belonging to 5 fish species which were collected from a video. The accuracy achieved by the proposed framework was 98.122%. In future work, the CNN architecture can be enhanced to improve the performance and the proposed method can be extended for real time underwater images dataset that has more categories of underwater species.

References

[1]. De Souza, E. N., Boerder, K., Matwin, S., & Worm, B. (2016). Improving fishing pattern detection from satellite AIS using data mining and machine learning. *PloS one*, *11*(7).

[2]. Butt, M. M., Latif, G., Iskandar, D. A., Alghazo, J., & Khan, A. H. (2019). Multi-channel Convolutions Neural Network Based Diabetic Retinopathy Detection from Fundus Images. *Procedia Computer Science*, *163*, 283-291.

[3]. Alghmgham, D. A., Latif, G., Alghazo, J., & Alzubaidi, L. (2019). Autonomous Traffic Sign (ATSR) Detection and Recognition using Deep CNN. *Procedia Computer Science*, *163*, 266-274.

[4]. Shaikh, E., Mohiuddin, I., Manzoor, A., Latif, G., & Mohammad, N. (2019, October). Automated Grading for Handwritten Answer Sheets using Convolutional Neural Networks. In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)* (pp. 1-6). IEEE.

[5]. Chouiten, M. (2013). Underwater Real-Time Fish Recognition by Image Processing. *Computer Graphics and Imaging / 798: Signal Processing, Pattern Recognition and Applications*. doi:10.2316/p.2013.798-116

[6]. Pelletier, S., Montacir, A., Zakari, H., & Akhloufi, M. (2018, May). Deep Learning for Marine Resources Classification in Non-Structured Scenarios: Training vs. Transfer Learning. In *IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)* (pp. 14). IEEE.

[7]. Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y. H. J., Fisher, R. B., & Nadarajan, G. (2010, October). Automatic fish classification for underwater species behavior understanding. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams* (pp. 45-50). ACM.

[8]. Zhuang, P., Wang, Y., & Qiao, Y. (2018, October). WildFish: A Large Benchmark for Fish Recognition in the Wild. In *ACM Multimedia Conference on Multimedia Conference* (pp. 1301-1309). ACM.

[9]. Hossain, E., Alam, S. S., Ali, A. A., & Amin, M. A. (2016, May). Fish activity tracking and species identification in underwater video. In *5th International Conference on Informatics, Electronics and Vision (ICIEV),* (pp. 62-66). IEEE.

[10]. Hnin, T. T., & Lynn, K. T. (2016). Fish Classification Based on Robust Features Selection Using Machine Learning Techniques. In *Genetic and Evolutionary Computing* (pp. 237-245). Springer.

[11]. Jin, L., & Liang, H. (2017, June). Deep learning for underwater image recognition in small sample size situations. In *OCEANS 2017Aberdeen* (pp. 1-4). IEEE.

[12]. Villon, S., Chaumont, M., Subsol, G., Villéger, S., Claverie, T., & Mouillot, D. (2016, October). Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between Deep Learning and HOG+ SVM methods. In *International Conference on Advanced Concepts for Intelligent Vision Systems* (pp. 160-171). Springer, Cham.

[13]. Siddiqui, S. A., Salman, A., Malik, M. I., Shafait, F., Mian, A., Shortis, M. R., & Harvey, E. S. (2018). Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. *ICES Journal of Marine Science*, *75*(1), 374-389.

[14]. Fouad, M. M., Zawbaa, H. M., Gaber, T., Snasel, V., & Hassanien, A. E. (2016). A fish detection approach based on BAT algorithm. In *The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015)*, Beni Suef, Egypt (pp. 273-283). Springer, Cham.

[15]. Rodrigues, M. T., Padua, F. L., Gomes, R. M., & Soares, G. E. (2010, September). Automatic fish species classification based on robust feature extraction techniques and artificial immune systems. In *BioInspired Computing: Theories and Applications (BIC-TA), 2010 IEEE Fifth International Conference on* (pp. 1518-1525). IEEE.

[16]. Miyazono, T., & Saitoh, T. (2018). Fish Species Recognition Based on CNN Using Annotated Image. In *IT Convergence and Security* (pp. 156-163). Springer, Singapore.

[17]. Ding G., Song Y., Guo J., Feng C., Li G., He B., Yan T. (2017). Fish recognition using convolutional neural network. *OCEANS 2017 – Anchorage*, (pp. 1-4). Anchorage, AK: IEEE.

[18]. Kratzert, F. and Mader, H., Fish species classification in underwater video monitoring using Convolutional Neural Networks, 2018.

[19]. Latif G, Alghazo J, Alzubaidi L, Naseer MM, Alghazo Y. (2018). Deep Convolutional Neural Network for Recognition of Unified Multi-Language Handwritten Numerals. In *IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition*, (pp. 90-95).