An Image Classifier for Underwater Fish Detection using Classification Tree-Artificial Neural Network Hybrid

Vincent Jan D. Almero
Electronics and Communications Engineering Department
De La Salle University
Manila, Philippines
vincent_almero@dlsu.edu.ph

Ronnie S. Concepcion II
Electronics and Communications Engineering Department
De La Salle University
Manila, Philippines
ronnie_concepcionii@dlsu.edu.ph

Edwin Sybingco
Electronics and Communications Engineering Department
De La Salle University
Manila, Philippines
edwin.sybingco@dlsu.edu.ph

Elmer P. Dadios
Manufacturing Engineering and Management Department
De La Salle University
Manila, Philippines
elmer.dadios@dlsu.edu.ph

***Abstract*—Fish detection using imaging technologies and computer vision systems is considered as an effective tool in fish monitoring for increasing the production to satisfy future global demands. This persistent tasks, with image classification as one of its subtasks, encounters challenges due to the complex nature of underwater images. A proposed approach to address this subtask was to create a hybrid image classification model from classification tree and artificial neural network. The classification tree component performed feature selection to extract a reduced representation of the fundamental dataset, derived from a series of acquired and processed underwater images in a land-based aquaculture setup. This said representation was then fed to a feedforward artificial neural network to develop such model. The best configuration of this hybrid model was determined, based on learning time and cross entropy, and was compared to a classification tree and an artificial neural network, both developed from the fundamental dataset, based on training and testing accuracies. The best performing hybrid model, composed of 100 hidden neurons in the artificial neural network component, achieved training and testing accuracies of 93.6% and 78.0%, respectively, hence, providing a competitive solution to the image classification in fish detection problem.**

Keywords— Artificial Neural Network, Aquaculture, Binary Classifier, Classification Tree, Feature Selection, Fish Detection

I. INTRODUCTION

Object detection, a task in computer vision that determines the classification and location of such object in an image or a video frame, in underwater-based setups is a relatively new research area. A demand to produce researches from this interest is motivated by the relative lack of such researches in this setup, in comparison to the land-based setups, due to challenges of the complexities in the captured images due to the underwater illumination and the necessity for a real-time analysis due to the drastic changes in the environment [1].

Furthermore, fish detection, a special application of this computer vision task in an underwater environment, is seen one of the promising approaches to fish monitoring [2], providing a better understanding of fishes' growth cycle and interaction to the environment (e.g. the effect of the harsh and stressful conditions on the behavior of fishes [3], the drivers of the changes in the fish ecosystems [4]). Such derived information would efficiently increase fish harvest by reducing mortality rates due to diseases, mutations and erratic behaviors, addressing the continuous increase in demand for fish produce [5]; hence, further enhancing the motivation to partake in such studies.

The fish detection, much similar to object detection, can be further divided into several subtasks: (1) image classification, classifying whether or not an image contains a fish; (2) fish recognition, the identification of the fish within an image; and (3) fish localization, the determination of coordinates that determines the location of the fish within an image [6]. These subtasks can be solved through integration of image processing and machine learning algorithms. Typically, these algorithms observe data preprocessing, feature extraction, feature selection, modelling, matching and positioning. These phases are summarized through a pipeline shown below [1] ; some related studies that addresses this problem are as follows.

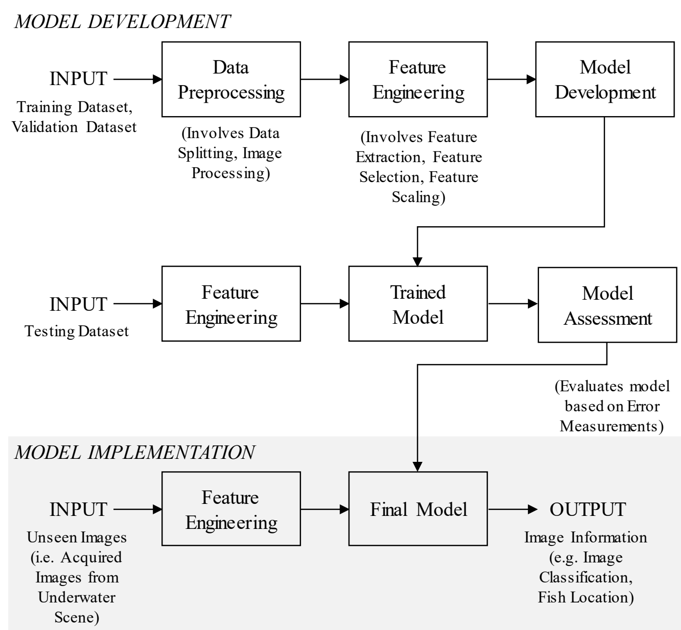


Fig. 1. The general framework of fish detection. In this sense, a model addresses some or all subtasks in the fish detection (i.e. Image Classification, Fish Recognition and Localization)

A synthesis of pixel-level background modelling and frame-level blob detection algorithms, was used as a part of fish species recognition system of [7] by Balance-Guaranteed Optimized Tree (BGOT). This fish recognition algorithm achieved an average accuracy of 61.37%. Also, [7] reported that support vector machine (SVM) is intended for binary classification tasks and implied that fish recognition using SVM should be feasible and should perform well.

The use of different methodologies for fish detection such as background subtraction, morphological operations, edge detection and geometric algorithms was proposed by [8], to establish the presence or absence of fish in an image. Accordingly, this presence or absence, can be further validated by utilizing an artificial neural network (ANN) or an SVM classification, based on the features extracted from fish species; so this supports the report of [7] on the feasibility of SVM for fish detection.

Another method for fish detection is the use of Gaussian mixture models (GMM), as in [9] and in [2], to estimate the background and subtract this background to the captured image for fish detection. The system performance of [9] was reported at 73%. On the other hand, some erosion and dilation operators were appended at fish detection algorithm of [2], to improve the system performance from 69.92% to 83.26%.

As observed, the integration of machine learning approaches and image processing algorithms had promising performances. This research paper introduces another competitive and substantial approach for solving the image classification applied specifically to aquaculture-based fish detection systems, by supporting the report of [8], using an artificial neural network, incorporating this approach to a classification tree, and letting this hybrid model work in conjunction with image processing algorithms. Specifically, the objectives of this research study are: (1) to develop a binary image classifier for fish detection, that differentiates images with fish and images without fish, applied in an aquaculture setup, using a classification tree – artificial neural network hybrid, and (2) to evaluate the performance of this developed model.

II. METHODOLOGY

The proposed model, a binary classifier that recognizes the images with fish from images without fish, in aquaculture setup, is shown below.

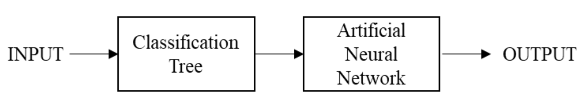


Fig. 2. The block diagram of the classification tree-artificial neural network hybrid classification model

This sequential hybrid is composed of machine learning algorithms of different approaches, that can act autonomously and independently: a classification tree and an artificial neural network; some output knowledge derived from the classification tree serves as an input of the artificial neural network.

The classification tree, is a symbolic machine learning approach that induces the classification of objects by developing rules based on the values of the features of objects. This machine learning structure is composed of internal nodes and root node, contains a certain feature of the objects, the leaf nodes, contains the classification of the objects, and branches, represents a test of the values of a certain feature of the objects; each branch connects a root node or an internal node to another internal node or a leaf node. A dataset, composed of features and classification of the objects, is necessary to construct such structure. In developing the relationship between these nodes, a specific feature of the dataset is selected to be tested for the establishment of the partitions between the dataset. Then, this step is performed again using the specific partition, instead of the dataset. This series of steps is recursively done until all the members of the partition are in the same classification; this classification results to a leaf node of a classification tree.

In this hybrid approach, the classification tree aims to perform feature selection, the process of obtaining reduced representation or determining the most relevant (input) features out of the fundamental dataset to decrease internal complexity of a model [10]. This reduced dataset is then fed to the artificial neural network, another component of this hybrid.

The artificial neural network, with the structure of such component is shown on the next page, is a connectionist machine learning approach that determines a profound relationship between inputs and outputs using connection of concertedly running neurons [11], [12]. A neuron is the basic computing element of this approach. The simplest model of a neuron is a perceptron, a computing structure with adjustable weights and bias as an input and an output, with summation function for the multiplied weights and inputs and bias, and an activation function, that limits the output of this summation, usually a large number, into a specific range, and is shown at a figure below.

This artificial neural network can be constructed with an input layer, which takes each input and feeds it into a neuron, and an output layer, takes the output of all the preceding neurons and feeds it to the neurons to generate an output. However, a hidden layer, composed of neurons, is added to act as an intermediary between the input layer and the output layer. This arrangement of the layers of neurons that significantly improves the classification performance is also known as multilayer perceptron [13].

To develop the profound relationship between the input and outputs in a dataset, backpropagation algorithm is used. The backpropagation adjusts the weights and biases (inputs) of each neuron; the adjustments are done with a certain algorithm, based on comparison between the actual and the target outputs.

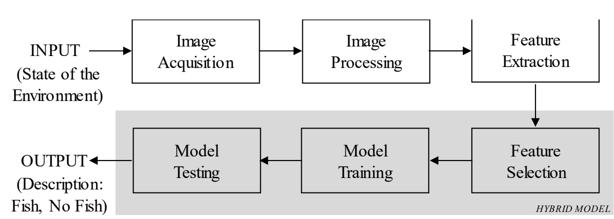


Fig. 3. The process flow of development of the hybrid binary classifier model.

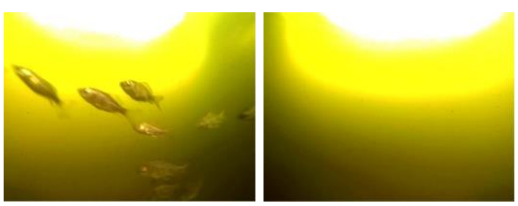
A. Development

The figure on the previous page presents the development of the binary image classifier using the classification tree-artificial neural network hybrid. First, the underwater images of a land-based fishpond with common carp (Cyprinus carpio) were acquired and processed. Second, some features of these processed images were considered as a fundamental dataset for the development of the hybrid model. Then, the features from this dataset were assessed based on relevance. Next, this reduced dataset was subdivided as training dataset and testing dataset. Lastly, these datasets were used to train and test the model that provides a description of the state of the (aquaculture) fishpond, whether the captured scene has fish in it or otherwise. Each process is explained further, as follows.

1) Image Acquisition: Underwater images from the common carp ecosystem in an aquaculture setup was acquired during the first 4 months of the fishes' growth, through the extraction of frames of underwater videos. These videos were captured by a GoPro 4 Hero Black Camera, situated at approximately 22'' from the water surface.

2) Image Processing: As expected, these underwater images suffered from low quality due to the complex illuminations produced by the interactions of the natural sunlight and the water particles, attributed to its turbidity [14]. So, image processing algorithm of [15], which are composed of Dark Channel Prior and Linear Contrast Enhancement, were applied to enhance these images.

(a)



(b)

Fig. 4. (a) A sample of captured underwater images (b) A sample of enhanced underwater images.

3) Feature Extraction, Selection: The color features from the RGB and HSV channels of the images, which is one of the low-level features of an image, tend to be the most intuitive, most obvious and easiest to calculate, in comparison to the point-, line- and region-based features; hence, the consideration of the averages and standard deviations of the R, G, and B, and H, S, and V levels, to the development of the proposed image classification model.

To expound this consideration further, the RGB color space is a common source of features in model development. However, each channel in this color triplet is closely correlated; the isolation of a certain channel in this color space may misrepresent the information in an image [16], so other significant color space which closely represents human vision systems, such as HSV is also considered [7].

These color level quantities were derived from the group of processed images to compose the input portion of the fundamental dataset. A binary value that corresponds to the to the derived set of color features and that describes the presence and absence of fish in an image (1, image with fish; 0, image without fish), was added as the output portion of the fundamental dataset.

To determine the relevant (input) features, a classification tree was generated using the fundamental dataset. In the generation of this classification tree, a specific (input) feature, deemed as a relevant feature, was selected to govern the partitions in the nodes. To create these partitions, the Gini's diversity index, a measure that determines the sparsity of a dataset through accounting the average differences from all of the members of the dataset [17], was observed. After the recursive processes of selection of relevant features and establishment of partitions to the dataset, these relevant (input) features in the internal nodes and root node were only considered to create the reduced dataset.

4) Model Training, Validation, Testing: This reduced dataset was subdivided into training dataset and testing dataset, and was used to develop the artificial neural network component of the hybrid; the composition of this dataset is shown below.

TABLE I. DATASET FOR THE DEVELOPMENT OF THE HYBRID MODEL

Type	Composition		
	Images with fish	***Images without fish***	***Total***
Training	150	50	250
Testing	50	50	100

This artificial neural network observed a simple feedforward architecture, formed and arranged as follows: an input layer, a hidden layer, and an output layer; the connections between neurons go from one layer to the next and does not feedback to the previous layer. Also, tangent sigmoid, a non-linear activation function for the neurons at hidden layer,

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (1)$$

and, logistic sigmoid, a non-linear activation function for the neurons at output layer,

$$f(x) = \frac{1}{1 + e^{-2x}} \quad (2)$$

was considered. As reported in [13], these activation functions were capable of handling binary classification. In addition, the backpropagation algorithm used for the training this artificial neural network was the scaled conjugate gradient.

B. Evaluation

To obtain the best configuration of this artificial neural network, the number of neurons in the hidden layers were adjusted (from 10 to 1000), and the metrics necessary to assess the performance of this component of the hybrid model: (1) learning time, the duration of the model training, measured in ms, (2) cross entropy, a measure of performance of the trained model, determining the difference between the two probability distributions that represents the actual value and the target value, (3) accuracy, the correctness in classifying the data, during its training phase and testing phase, were recorded and compared. The best performing configuration should achieve the least training time and least cross-entropy during training and highest accuracies during training and testing.

To assess the effectiveness of the feature selection of the classification tree component of this hybrid model, the training and testing accuracies of this hybrid model developed with reduced dataset, an artificial neural network of best configuration (obtaining this configuration is similar to the process stated earlier), and a classification tree, both developed with fundamental dataset, was compared.

III. Results and Discussions

The fundamental dataset was acquired from the processed images. Then, this dataset was used to train the classification tree component of the hybrid. The rules of the classification tree is shown below.

TABLE II. CLASSIFICATION TREE COMPONENT OF THE HYBRID MODEL

Node	Rules[b]
1	if H_sd<0.152687 then node 2 elseif H_sd>=0.152687 then node 3 else 1
2	if H_sd<0.152438 then node 4 elseif H_sd>=0.152438 then node 5 else 1
3	if B_sd<0.522578 then node 6 elseif B_sd>=0.522578 then node 7 else 1
4	class = 1
5	if H_sd<0.15244 then node 8 elseif H_sd>=0.15244 then node 9 else 1
6	if S_mean<1.34372 then node 10 elseif S_mean>=1.34372 then node 11 else 0
7	if S_mean<1.42077 then node 12 elseif S_mean>=1.42077 then node 13 else 1
8	class = 0
9	if S_sd<0.995801 then node 14 elseif S_sd>=0.995801 then node 15 else 1
10	if V_mean<2.7212 then node 16 elseif V_mean>=2.7212 then node 17 else 1
11	if R_sd<1.78776 then node 18 elseif R_sd>=1.78776 then node 19 else 0
12	if H_mean<1.30696 then node 20 elseif H_mean>=1.30696 then node 21 else 1
13	class = 0
14	if S_sd<0.995012 then node 22 elseif S_sd>=0.995012 then node 23 else 1
15	class = 1
16	if R_mean<1.40497 then node 24 elseif R_mean>=1.40497 then node 25 else 0
17	if R_sd<1.14539 then node 26 elseif R_sd>=1.14539 then node 27 else 1
18	class = 0
19	class = 1
20	class = 1
21	if H_mean<1.32745 then node 28 elseif H_mean>=1.32745 then node 29 else 1
22	class = 1
23	class = 0
24	class = 0
25	class = 1
26	if R_mean<1.03302 then node 30 elseif R_mean>=1.03302 then node 31 else 0
27	if H_sd<0.153132 then node 32 elseif H_sd>=0.153132 then node 33 else 1
28	if H_mean<1.32661 then node 34 elseif H_mean>=1.32661 then node 35 else 1
29	class = 1
30	class = 1
31	class = 0
32	if S_sd<0.994118 then node 36 elseif S_sd>=0.994118 then node 37 else 1
33	class = 1
34	class = 1
35	class = 0
36	if R_mean<1.2433 then node 38 elseif R_mean>=1.2433 then node 39 else 1
37	class = 1
38	class = 0
39	class = 1

[b] Rules are presented in level-first approach, from left to right

As observed from the nodes of the classification tree, the relevant features were the averages of R and standard deviations of R and B of the RGB color space and the averages of H, S, and V and standard deviations of H and S of the HSV color space. The features of G channel were eliminated to create the reduced dataset. This suggests that these features are irrelevant, because these features are inherent in an image; the 'greenness' is common to all images. Also, the standard deviation of V was excluded to produce the reduced dataset. According to [16], V is the only channel in the HSV color space that is affected by the illuminations; the removal of the such complex feature in the dataset should improve the performance of this hybrid model. In addition, almost all of the features derived from the HSV color space are relevant; supports [7] that states HSV as a significant color space.

After employing this significant information gathered from the generation of the classification tree, the reduced dataset was used to train and test the artificial neural network component of this hybrid model. The metrics that reflect the performance of the configurations of this artificial neural network, gathered from the adjustments done in the number of neurons at the hidden layers, is presented at the table shown on the next page.

TABLE III. THE PERFORMANCE OF ARTIFICIAL NEURAL NETWORK COMPONENT OF THE HYBRID MODEL

Number of Neurons	**Metrics**	
	Learning Time (ms)	***Cross-entropy***
10	7.2620	0.1162
25	23.7820	0.1848
50	34.4430	0.1404
75	3.7660	0.1769
100	3.5030	0.1204
250	5.3730	0.1407
500	8.0670	0.1578
750	6.3240	0.1888
1000	7.2540	0.1747

Based on the cumulative weights of learning time and cross-entropy, as the neural network operators during the training stage, an artificial neural network with 100 neurons in the hidden layer results to the best performing configuration.

A confusion matrix, a visual representation of the performance of a classifier, was generated for the training and the testing phase, and accuracy was presented as a part of a confusion matrix, all shown in the figures on this page. The testing accuracy provides more correlation to the actual performance of the developed model than the training accuracy, so, in this assessment, the testing accuracy was only considered. The proposed model differentiates images without fish from images with fish 73.3% of the time and correctly identifies the images with fish 88.0% of the time. These results were competitive since the intent of this hybrid classifier is to detect the presence of fish from the group of captured images.

Then, this developed hybrid model was compared with an artificial neural network (best configuration is at 500 neurons in the hidden layers) and a classification tree, trained and tested with the fundamental dataset. The training and testing accuracies are shown below.

TABLE IV. THE PERFORMANCE OF THE BINARY CLASSIFIERS

Methodology	**Accuracy**	
	Training	***Testing***
CT	88.4%	51.0%
ANN	94.8%	60.0%
Hybrid (CT-ANN)	93.6%	78.0%

As observed, all methodologies shown inconsistent training and testing accuracies, a sign of overfitting. The imbalance in the dataset attributed to the observed overfitting of all the methodologies. The development of a model using imbalanced dataset dictates the said model to be biased at a certain class. Also, overfitting is shown in machine learning models developed with noisy dataset.

All methodologies were trained and tested with the imbalanced dataset. However, the hybrid model achieved the highest training and testing accuracies and exhibited lesser degree of inconsistency between the training and testing accuracies, compared to other methodologies. In this experimentation, the hybrid model was considered the best performing model; the feature selection of the classification tree component was successful in reducing this imbalance and noise in the dataset, hence, presenting a significant effect to the performance of the hybrid model.

(a)

(b)

Fig. 5. The confusion matrix of the best configuration of the hybrid model (a) at training (b) at testing

IV. Conclusions and Recommendations

The classification tree-artificial neural network hybrid was used for generating a model that can identify images with fish and images without fish, one of the persistent tasks in underwater fish detection. The classification tree component performed feature selection on the dataset gathered from the captured underwater images in a land-based fishpond. Some features were eliminated from the dataset to produce a model that can perform better than models developed with an unaltered dataset. Also, the best configuration of the artificial neural network was selected in the development of this model, through the adjustment in the number of neurons in the hidden layers. Each of the components of this model contributed to the competitive training and testing accuracies.

As observed, the quality of the developed model is dependent on the quality of dataset. An observance to the balance of this data through further data preprocessing can be observed, to develop a more reliable classification model. Also, deep learning techniques that reduces the variability of the feature selection and engineering can be considered for the generation of a model. In addition, a quantitative assessment to deeply understand the feature selection of a dataset can be of interest. To emphasize, this binary classifier is limited to addressing the image classification subtask of the fish detection. Hopefully, this model seeks to be integrated in other approaches that solves the fish recognition and localization, the other subtasks in the fish detection.

Acknowledgment

This research work, a contribution to the Smart Farm Project of the Intelligent Systems Laboratory, Gokongwei College of Engineering, De La Salle University – Manila, is supported by the Department of Science and Technology – Engineering Research and Development for Technology (DOST - ERDT).

References

[1] H. Yang, P. Liu, Y. Z. Hu, and J. N. Fu, "Research on underwater object recognition based on YOLOv3," *Microsystem Technologies*, 2020.

[2] A. Salman, S. Maqbool, A. H. Khan, A. Jalal, and F. Shafait, "Real-time fish detection in complex backgrounds using probabilistic background modelling," *Ecological Informatics*, vol. 51, pp. 44–51, May 2019.

[3] M. Saberioon, A. Gholizadeh, P. Cisar, A. Pautsina, and J. Urban, "Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues - Saberioon - 2017 - Reviews in Aquaculture - Wiley Online Library," *Reviews in Aquaculture*, vol. 0, pp. 1–19, 2016.

[4] J. Radinger *et al.*, "Effective monitoring of freshwater fish," *Fish and Fisheries*, vol. 20, pp. 729–747, 2019.

[5] B. W. Brooks and J. L. Conkle, "Commentary: Perspectives on aquaculture, urbanization and water quality," *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, vol. 217, pp. 1–4, Mar. 2019.

[6] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11. Institute of Electrical and Electronics Engineers Inc., pp. 3212–3232, 01-Nov-2019.

[7] P. X. Huang, B. J. Boom, and R. B. Fisher, "Underwater Live Fish Recognition Using a Balance-Guaranteed Optimized Tree," pp. 422–433, 2013.

[8] M. R. Shortis, M. Ravanbakhsh, F. Shafait, and A. Mian, "Progress in the Automated Identification, Measurement, and Counting of Fish in Underwater Image Sequences," *Marine Technology Society Journal*, vol. 50, no. 1, pp. 4–16, Jan. 2016.

[9] E. Lantsova, T. Voitiuk, T. V. Zudilova, and A. Kaarna, "Using low-quality video sequences for fish detection and tracking," in *Proceedings of 2016 SAI Computing Conference, SAI 2016*, 2016, pp. 426–433.

[10] C. Turchetti and L. Falaschetti, "A manifold learning approach to dimensionality reduction for modeling data," *Information Sciences*, vol. 491, pp. 16–29, Jul. 2019.

[11] R. S. Concepcion and L. C. Ilagan, "Application of Hybrid Soft Computing for Classification of Reinforced Concrete Bridge Structural Health Based on Thermal-Vibration Intelligent System Parameters," in *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)*, 2019, pp. 207–212.

[12] J. C. Puno, E. Sybingco, E. Dadios, I. Valenzuela, and J. Cuello, "Determination of soil nutrients and pH level using image processing and artificial neural network," in *HNICEM 2017 - 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, 2017, vol. 2018-January, pp. 1–6.

[13] R. A. R. Bedruz, A. Fernando, A. Bandala, E. Sybingco, and E. Dadios, "Vehicle Classification Using AKAZE and Feature Matching Approach and Artificial Neural Network," *TENCON 2018 - 2018 IEEE Region 10 Conference*, pp. 1824–1827, Oct. 2018.

[14] M. Jian *et al.*, "The extended marine underwater environment database and baseline evaluations," *Applied Soft Computing*, vol. 80, pp. 425–437, Jul. 2019.

[15] K. Xie, W. Pan, and S. Xu, "An Underwater Image Enhancement Algorithm for Environment Recognition and Robot Navigation," *Robotics*, vol. 7, no. 1, p. 14, Mar. 2018.

[16] R. Prados, R. Garcia, N. Gracias, L. Neumann, and H. Vagstol, "Real-time fish detection in trawl nets," in *OCEANS 2017 - Aberdeen*, 2017, pp. 1–5.

[17] S. Goswami, C. A. Murthy, and A. K. Das, "Sparsity measure of a network graph: Gini index," *Information Sciences*, vol. 462, pp. 16–39, Sep. 2018.