recognition of fish species in underwater images

Hanane Abourifa , Oumayma Essarhi , Mouhamed Amine Kamili

Faculty of Sciences Ben M'sik, Hassan II University, Casablanca,Morocco

Abstract: The exploration of underwater ecosystems unveils a mysterious world, where fish species play a central role. Recognizing these species in underwater images blends marine biology, image technology, and artificial intelligence. Despite challenges in understanding marine biodiversity, recent advancements in underwater photography and image processing offer new perspectives. Accurate fish species identification not only enhances marine life comprehension but also finds practical applications in ecosystem preservation and environmental monitoring.

Recent progress in artificial intelligence, particularly in computer vision, enables sophisticated species recognition systems. Leveraging deep learning algorithms on extensive visual datasets, this fusion of marine biology and technology opens new possibilities for profound ocean exploration and effective marine conservation. This abstract introduces a journey into the depths, exploring the symbiosis between marine biology and technology in fish species recognition, emphasizing its pivotal role in preserving marine ecosystems. using resnet50 and mobilenetv2.

Keywords: ***recognition Fish, ResNet50 ,MobileNetV2***

1. INTRODUCTION

When we talk about fishing, it's not about picking or mining; it's about hunting. The seas, lakes, and rivers are filled with human traps leading to the death of billions of animals each year. Fishing boats are becoming more sophisticated, equipped with tools that leave no chance for fish, both in terms of capture techniques and the multitude of devices used to spot their prey. It is estimated that over 1 trillion fish are caught annually.

Preserving aquatic resources is crucial for ensuring the sustainability of fishing activities without jeopardizing fish populations. The project focused on recognizing fish species in underwater images plays a crucial role in this management by collecting comprehensive data on fish communities. This information is crucial for determining appropriate catch levels, ensuring the maintenance of healthy and sustainable fish populations in the long term.

The marine world covers more than 70% of our planet, retaining its infinite complexity, especially concerning marine species With the advancements in artificial intelligence and computer vision, numerous studies have been conducted to better understand and explore the mysteries of the marine world. Our project falls within this category of research.

The main goal of this project is to develop an automatic fish detection system from underwater photos using advanced image processing and machine learning techniques, particularly in the field of computer vision. This system should be capable of recognizing and classifying different types of fish in images. The ultimate objective is to provide a precise and effective solution for monitoring marine life. This solution can be valuable for marine ecosystem monitoring, studying relationships between marine species, gaining a deeper understanding of marine food chains, and enhancing sustainable fishing practices.

2. RELATED WORK

This study [1]introduces an approach utilizing VGG-16 with a deep fish architecture for improved feature extraction and accuracy. The method involves fish localization and classification, employing VGG16 for feature mapping and DeepFish architecture for categorization. The combination of support vector machines and a random forest classifier achieves an impressive accuracy of 99.47%.

Focusing on local and coastal fish classification, this work[2] uses a deep learning approach based on convolutional neural networks (CNN). The modified AlexNet achieves a satisfactory classification accuracy of 98.33% across twelve different fish species.

This article[3] presents a two-step deep learning approach for detecting and classifying temperate fish using YOLO for object detection and CNN with SE architecture for classification. Transfer learning is employed to enhance classification accuracy, reaching a peak accuracy of 99.27%.

Focused on classifying various fish species, this work[4] utilizes a CNN with Keras and TensorFlow. With 10

classes and 560 fish images, the model achieves a notable 95% classification accuracy on the test dataset.
Targeting automatic fish species classification, this study[5] employs a modified AlexNet on QUT fish and LifeCLEF-15 datasets, achieving a classification accuracy of 90.48% across six fish species.
Focused on fish species classification, this work[6] uses ResNet-50 on a diverse dataset, reaching an impressive 99% classification accuracy across 20 fish species in various underwater habitats.
Leveraging GMM and Optical Flow, this study[7] addresses fish detection and classification, demonstrating robust results. The combination of ResNet50 and YOLOv3 handles complex scenarios, achieving reliable performance.
Using Mask R-CNN for fish identification, this study[8] employs two branches for localization/classification and tracking, showcasing satisfactory performance, albeit with occasional limitations.
Comparing YOLOv3 and Mask R-CNN for fish detection, this work [9] highlights the strengths and weaknesses of each technique. Challenges include manual data labeling and data insufficiency.
This study [10] proposes YOLOv7 and FD_Net for fish detection and species recognition, achieving high performance with MobileNetV3 for object detection. However, real-world applicability may be affected by the pristine nature of the utilized dataset.

3. METHODOLOGY

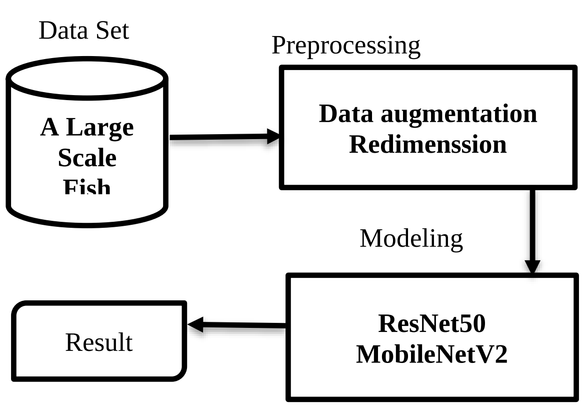

Figure1 : diagram of our approach

3.1 DATA SET

In this paper, we utilize a widely recognized and extensively cited dataset known as the "Large-Scale Fish Dataset." This dataset has been referenced in scientific papers on 40 occasions, highlighting its significance in the research community. It is publicly available and comprises a collection of images representing nine distinct types of seafood, sourced from the fish counter of a local grocery store. To capture the images for this dataset, two camera models were employed: the Kodak Easyshare Z650 and the Samsung ST60. This dataset contains 50 distinct fish images per each of seven classes as follows: red mullet, gilt-head bream, horse mackerel, sea bass, red sea bream, black sea sprat, and striped red mullet, and 30 distinct images are captured for trout and shrimp. To ensure the dataset's authenticity and representativeness, only fresh fish was used during the image acquisition process. The fish were positioned in various orientations and displacements to capture their natural diversity. However, the lighting conditions remained relatively consistent throughout the entire procedure to minimize any drastic shifts in illumination. Furthermore, in order to make the dataset applicable and practical to real-life scenarios, we intentionally opted for a blue and noisy background rather than a pristine white background. This decision was made to simulate the conditions typically encountered in real-life environments. Following the collection of images, the dataset underwent augmentation and balancing processes to enhance its efficiency, utility, and suitability for implementation. Detailed discussions about these processes will be provided in the subsequent preprocessing phase. Consequently, the dataset has been expanded to encompass a total of 9000 samples, with an evenly distributed allocation of 1000 samples per class, as visually depicted in the accompanying figure 2.

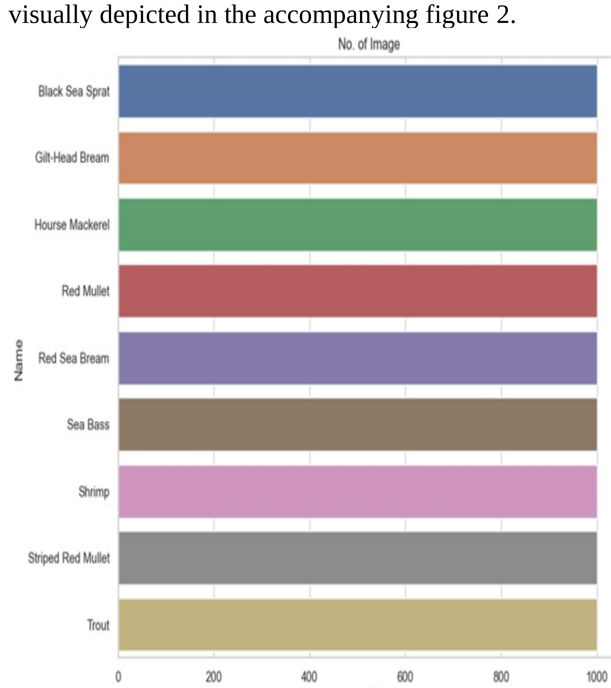

Figure2 : classe of data set large scale

3.2 PREPROCESSING

In our approach, we applied data augmentation to increase the quantity of available data. Additionally, we resized the images to adjust their dimensions :

3.2.1 DATA AUGMENTATION

In this work, data augmentation employed a transformation to generate numerous copies of frames with varying variances. Rotating, shifting, flipping, and zooming were some of the augmentation techniques used. This strategy did not alter the target class, but it gave a new viewpoint on catching objects in new frames [11].
To eliminate the class imbalance in the data, the data augmentation approach was also employed as an oversampling method. It not only expanded the data set but also introduced some variety, allowing the model to generate better classifiable predictions on previously unknown data. Furthermore, when trained on new, slightly modified frames, the models became more resilient.
Some of the augmentation strategies used in this study are listed below.
1. Random rotation: This is an augmentation approach that allows the model to be object orientation invariant [12]. Data augmentation in the form of a rotation range allows for random rotation of frames across any degree between 0 and 360. Some pixels migrate outside the picture when the frame rotates, leaving empty spaces that could be filled.
2. Random shifts: This is a method used to compensate for the fact that objects are not constantly in the center of the frame. This issue can be resolved by moving the frame pixels horizontally or vertically. The height shift range is used for vertical frame shifts, whereas the width shift range is utilized for horizontal shifts.
3. Random Flip: One of the augmentation strategies used to flip frames is random flip [13]. The horizontal flip is used to flip the frame horizontally, and the vertical flip is used to flip the frame vertically. However, for the model to generate classification results based on the class, the augmentation strategy utilizing random flip must have a frame that is symmetrical with the item in the original frame.
4. Random Zoom: This is an augmentation method that uses a zoom range to randomly increase or shrink the picture size.

3.2.2 RESCIZING IMAGE

Resizing input images is a crucial step when using image classification models like MobileNet and ResNet50. This ensures that images are compatible with the expected input dimensions of the model, facilitating consistent and effective data processing and contributing to better overall model performance.
Models such as MobileNet and ResNet50 are typically pre-trained on large datasets like ImageNet, where images have been resized to specific dimensions for training purposes. Resizing input images ensures that all images are of the same size, which simplifies model processing and learning.
Typical input dimensions for MobileNet and ResNet are usually 224x224 pixels or 299x299 pixels, depending on the model version used. Similarly, ResNet50 commonly expects input images to be of size 224x224 pixels. The figure below shows the result of the preprocessing.

Figure 3 : recizing images

3.3 MODELING

In this approach, ResNet and MobileNet were used :

Resnet50:

ResNet was first introduced in 2015, where it has also won ILSVRC competition with an error rate of 3.57%. ResNet's high accuracy rate can be mainly attributed to the introduction of residual layers that allow the network to be designed deeper compared to the previous popular network architectures. The residual layer or also known as identity mapping mitigates the problem of diminishing gradient in training a deep network, where the previous layer is fed to the later layers. The idea was to overcome the reduction of input features from the original learning feature that produces zero features [14].

MobileNetV2:

MobileNet is a deep learning architecture that focuses on the mobile platform where the computational resource is limited. An improved version, which is called MobileNet V2 is then introduced by Google with slight modifications to the original version. The basis of the network still remains the same, which is separable convolution. MobileNet version 2 previously trained on ImageNet datasets has been used to extract fruit image features in . The paper claimed that the parameters used have reduced from 4.24 millions to just 3.47 millions, but with better accuracy [15].

3.3.1 Architecture

ResNet50 :

```
['conv5_block3_2_relu[0][0]']
 D)

 conv5_block3_3_bn (BatchNo  (None, None, None, 2048)     8192
['conv5_block3_3_conv[0][0]']
 rmalization)

 conv5_block3_add (Add)      (None, None, None, 2048)     0
['conv5_block2_out[0][0]',
                                                                     'conv5_block3_
3_bn[0][0]']

 conv5_block3_out (Activati  (None, None, None, 2048)     0
['conv5_block3_add[0][0]']
 on)

 global_average_pooling2d (  (None, 2048)                 0
['conv5_block3_out[0][0]']
 GlobalAveragePooling2D)

 dense (Dense)               (None, 9)                    18441
['global_average_pooling2d[0][
                                                                    0]']

==================================================================================
==============
Total params: 23606153 (90.05 MB)
Trainable params: 23553033 (89.85 MB)
Non-trainable params: 53120 (207.50 KB)
__________________________________________________________________________________
```

Figure 4 : Archtecture ResNet50

There are a total of 23,606,153 parameters in the architecture of ResNet50.

MobileNetV2:

```
Model: "sequential"

_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 feature_extraction_layer (  (None, 1280)              4226432
 KerasLayer)

 dense (Dense)               (None, 9)                 11529

=================================================================
Total params: 4237961 (16.17 MB)
Trainable params: 11529 (45.04 KB)
Non-trainable params: 4226432 (16.12 MB)
_________________________________________________________________
```

Figure 5 : Archtecture MobileNetV2

There are a total of 4,237,961parameters in the architecture of MobilNetV2.

3.3.2 Hyperparaùeter

Model	*optimiseur*	*epochs*
ResNet50	**adam**	**12**
MobileNetV2	**adam**	**5**

In the ResNet model, we used the Adam optimizer with 12 epochs, while in MobileNetV2, we also utilized Adam with just 5 epochs.

3.4 Resualt

ResNet50 :

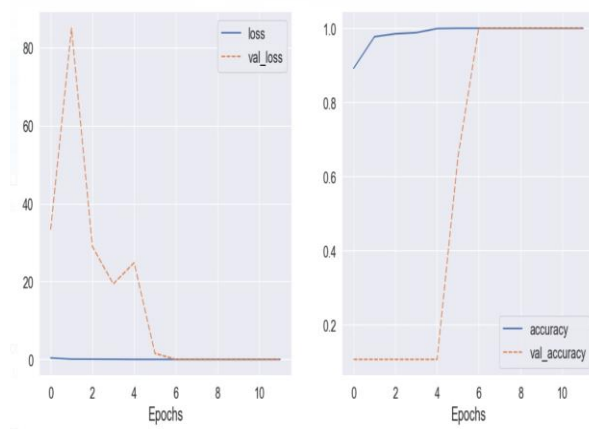

Figure 6 : resualt graph of ResNet50

	precision	recall	f1-score	support
Black Sea Sprat	1.00	1.00	1.00	195
Gilt-Head Bream	1.00	1.00	1.00	211
Hourse Mackerel	1.00	1.00	1.00	179
Red Mullet	1.00	0.99	0.99	191
Red Sea Bream	1.00	1.00	1.00	203
Sea Bass	1.00	1.00	1.00	207
Shrimp	1.00	1.00	1.00	209
Striped Red Mullet	0.99	1.00	1.00	204
Trout	1.00	1.00	1.00	201
accuracy			1.00	1800
macro avg	1.00	1.00	1.00	1800
weighted avg	1.00	1.00	1.00	1800

Figure 7 : resualt performance of ResNet50

MobileNetV2:

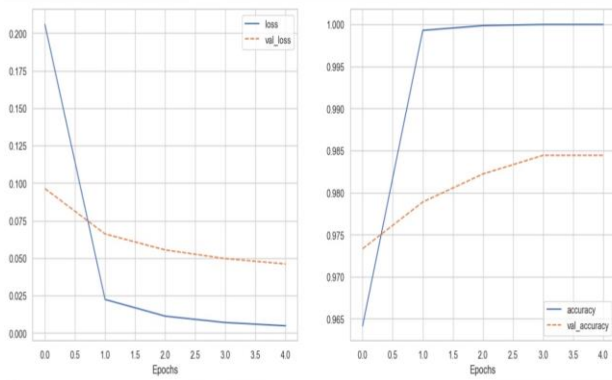

Figure 8 : resualt graph of MobileNetV2

	precision	recall	f1-score	support
Black Sea Sprat	1.00	1.00	1.00	91
Gilt-Head Bream	1.00	0.99	0.99	90
Hourse Mackerel	1.00	1.00	1.00	88
Red Mullet	1.00	1.00	1.00	79
Red Sea Bream	1.00	1.00	1.00	70
Sea Bass	0.92	1.00	0.96	90
Shrimp	1.00	1.00	1.00	104
Striped Red Mullet	1.00	1.00	1.00	132
Trout	1.00	0.96	0.98	156
accuracy			0.99	900
macro avg	0.99	0.99	0.99	900
weighted avg	0.99	0.99	0.99	900

Figure 9 : resualt performance of MobileNetV2

According to the results obtained, the performance of both models, ResNet50 and MobileNetV2, on the large-scale fish dataset is remarkably high, reaching up to 99%. This dataset consists of 9000 images distributed across 9 balanced classes of fish. Moreover, the application of data augmentation on the images has increased the quantity of available data for training.

The graphs depicting the error rates show that the training proceeded effectively, without the presence of overfitting on both models. Ultimately, for the fish recognition problem, both ResNet50 and MobileNetV2 stand out as the most performant models. Their accuracies surpass those reported in related works in this field.

4 Discussion

The results of our study demonstrate the effectiveness of the proposed approach in enhancing fish classification accuracy. By leveraging data augmentation techniques and deep learning models, we achieved improved classification results compared to traditional approaches. Data augmentation played a crucial role in our methodology by increasing the quantity and diversity of available data. The augmentation techniques, including random rotation, random shifts, random flip, and random zoom, allowed us to generate multiple copies of frames with varying variances. This approach provided a broader perspective on capturing fish objects in new frames, enabling the models to learn robust features for accurate classification. Additionally, data augmentation served as an oversampling method to address class imbalance issues, contributing to better classifiable predictions on previously unknown data. The resizing of input images was another important preprocessing step in our approach. It ensured compatibility with the expected input dimensions of the deep learning models, such as MobileNet and ResNet50. Resized images facilitated consistent and effective data processing, leading to improved overall model performance. The utilization of deep learning models, specifically ResNet50 and MobileNetV2, further enhanced the classification accuracy. ResNet50, with its ability to handle deep network architectures effectively, proved effective in capturing intricate features of fish species. On the other hand, MobileNetV2, designed for resource-constrained platforms, achieved comparable accuracy with reduced parameters, making it suitable for practical implementations. The combination of data augmentation techniques and deep learning models resulted in better generalization and handling of previously unseen fish species. The diverse augmented data allowed the models to learn more representative features and make more accurate predictions on the test data. It is important to note that the performance of the proposed approach may vary depending on the specific dataset and the characteristics of the fish species involved. Further research could explore additional augmentation strategies and evaluate the applicability of the approach to different fish classification tasks. In conclusion, our study highlights the effectiveness of data augmentation and deep learning models in enhancing fish classification accuracy. The combination of these techniques provides a powerful framework for accurate classification, enabling better generalization and handling of diverse fish species.

Reference

[1] F. Syreen et K. Merriliance, « DEEP CONVOLUTIONAL NETWORKS FOR UNDERWATER FISH LOCALIZATION AND SPECIES CLASSIFICATION », p. 91-100, nov. 2020, doi: 10.34218/IJARET.11.11.2020.009.

[2] S. Das, S. A. Shammi, et D. M. Raza, « A Substantial Deep Learning Approach for Classification of Local and Coastal Fish », in *Advanced Computing*, vol. 1781, D. Garg, V. A. Narayana, P. N. Suganthan, J. Anguera, V. K. Koppula, et S. K. Gupta, Éd., in Communications in Computer and Information Science,

vol. 1781. , Cham: Springer Nature Switzerland, 2023, p. 362-373. doi: 10.1007/978-3-031-35641-4_29.

[3] « Temperate fish detection and classification: a deep learning based approach | Applied Intelligence ». Consulté le: 2 février 2024. [En ligne]. Disponible sur: https://link.springer.com/article/10.1007/s10489-020-02154-9

[4] « Recognition Of Fish Categories Using Deep Learning Technique | IEEE Conference Publication | IEEE Xplore ». Consulté le: 2 février 2024. [En ligne]. Disponible sur: https://ieeexplore.ieee.org/document/8824916

[5] « Automatic Fish Species Classification Using Deep Convolutional Neural Networks | Semantic Scholar ». Consulté le: 2 février 2024. [En ligne]. Disponible sur: https://www.semanticscholar.org/paper/Automatic-Fish-Species-Classification-Using-Deep-Iqbal-Wang/880b8e93064544a265f6e424685d2ed985e8c5b2

[6] « [2008.12603] A Realistic Fish-Habitat Dataset to Evaluate Algorithms for Underwater Visual Analysis ». Consulté le: 2 février 2024. [En ligne]. Disponible sur: https://arxiv.org/abs/2008.12603

[7] A. Jalal, A. Salman, A. Mian, M. Shortis, et F. Shafait, « Fish detection and species classification in underwater environments using deep learning with temporal information », *Ecol. Inform.*, vol. 57, p. 101088, mai 2020, doi: 10.1016/j.ecoinf.2020.101088.

[8] « Automated detection and classification of southern African Roman seabream using mask R-CNN - ScienceDirect ». Consulté le: 2 février 2024. [En ligne]. Disponible sur: https://www.sciencedirect.com/science/article/pii/S1574954122000425

[9] « Frontiers | Automated Detection, Classification and Counting of Fish in Fish Passages With Deep Learning ». Consulté le: 2 février 2024. [En ligne]. Disponible sur: https://www.frontiersin.org/articles/10.3389/fmars.2021.823173/full

[10] H. Malik, A. Naeem, S. Hassan, F. Ali, R. A. Naqvi, et D. K. Yon, « Multi-classification deep neural networks for identification of fish species using camera captured images », *PLOS ONE*, vol. 18, nº 4, p. e0284992, avr. 2023, doi: 10.1371/journal.pone.0284992.

[11] A. Mumuni et F. Mumuni, « Data augmentation: A comprehensive survey of modern approaches », *Array*, vol. 16, p. 100258, déc. 2022, doi: 10.1016/j.array.2022.100258.

[12] C. Shorten et T. Khoshgoftaar, « A survey on Image Data Augmentation for Deep Learning », *J. Big Data*, vol. 6, juill. 2019, doi: 10.1186/s40537-019-0197-0.

[13] J. Nalepa, M. Marcinkiewicz, et M. Kawulok, « Data Augmentation for Brain-Tumor Segmentation: A Review », *Front. Comput. Neurosci.*, vol. 13, déc. 2019, doi: 10.3389/fncom.2019.00083.

[14] S. Z. M. Zaki, M. A. Zulkifley, M. M. Stofa, N. A. M. Kamari, et N. A. Mohamed, « Classification of tomato leaf diseases using MobileNet v2 », *IAES Int. J. Artif. Intell. IJ-AI*, vol. 9, nº 2, Art. nº 2, juin 2020, doi: 10.11591/ijai.v9.i2.pp290-296.

[15] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. de la T. Díez, et I. Ashraf, « Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network », *Diagnostics*, vol. 12, nº 5, Art. nº 5, mai 2022, doi: 10.3390/diagnostics12051280.