Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 6
ISET 2024

ResNET-based underwater fish identification with Data Augmentation

Yichen Ding[1, *], Hangyu Zhou [2]

[1] College of Computer and Information Science, Southwest University, Chongqing, 400715, China

[2] College of Physical Science and Technology, Hubei University of Education, Wuhan, 430205, China

* Corresponding Author: ding0427@email.swu.edu.cn

Abstract. The underwater environment is intricate, and light stands as one of the foremost influences. When the model trained with fish pictures under ideal conditions is applied to real underwater recognition, the influence of light will challenge the recognition accuracy, and it is costly to collect new underwater fish datasets. In this paper, Data Augmentation is applied to enhance the model's accuracy for underwater fish recognition. This paper begins by introducing the Residual Network (ResNet) network and Data Augmentation techniques. It then presents the research idea including how to validate underwater effects on the model and the effectiveness of Data Augmentation to improve underwater fish identification accuracy, modeling, analyzes data on model performance above and underwater and finally draws conclusions of influences of Data Augmentation on underwater fish identification based on analysis. According to the experimental results, it is concluded that applying Data Augmentation to model training can improve model’s accuracy in underwater fish recognition. The significance of this method is that it not only saves the cost of a new datasets, but also makes the model better adapted to underwater fish recognition.

Keywords: Data Augmentation; underwater fish recognition; ResNet; accuracy.

1. Introduction

Underwater environments are complex and variable, and illumination is one of the biggest changing factors. The variation of illumination has a considerable impact on image recognition, which may affect the recognition and detection accuracy of the model trained under ideal conditions for images in real situations [1]. In this paper, we would like to incorporate the perturbation of illumination as well as fish image samples from real scenes into the model training to improve the accuracy in real scenarios (underwater exploration, etc.) [2]. However, collecting new datasets requires a lot of cost, and collecting fish image samples under underwater dim conditions requires even more cost and has higher difficulty [3]. And the raw data are mostly taken under ideal conditions, which are far from the images in the real environment. Therefore, by augmenting the existing datasets and adding it to different types of ResNet for training, we can not only save the cost of new datasets, but also better adapt to the purpose of image recognition under real conditions.

This paper firstly analyzes the structural characteristic of different types of ResNet networks, then introduces the advantages of Data augmentation technology and application scenarios, then introduces the general experimental idea of adding the transforms. RandomGrayscale function to add the perturbation of light, then models the fish recognition model with Data augmentation and presents and analyzes the results.

2. Theoretical basis analysis

2.1. ResNet

The main forms of ResNet include ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152, as depicted in Figure 1, each network consists of three primary sections: the input section, the output section, and the intermediate convolution section. The intermediate convolution section comprises four stages, as depicted in the Figure 1.

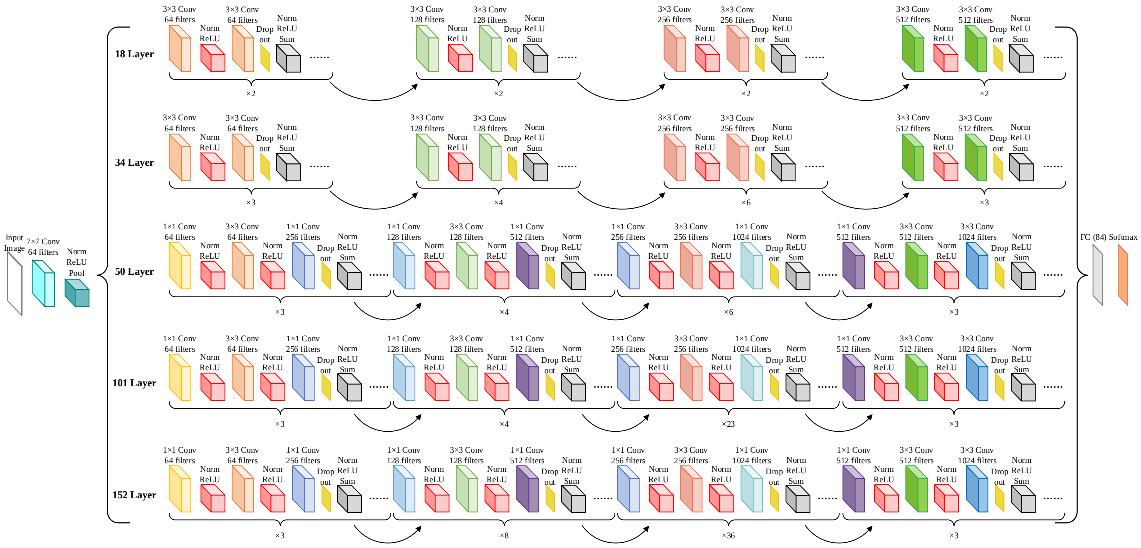

Figure 1. Schematic diagram of different kinds of ResNet structures

While ResNet variants exhibit a diversity of architectures, they all adhere to the structural characteristics outlined above. The distinctions between the networks primarily lie in the parameters of the blocks and the quantity of intermediate convolutional sections [4].

Network Input Section: In all ResNet networks, the input section features a substantial convolutional kernel with specification of 7x7, utilizing 64 filters and a stride of 2, followed by a max pooling operation with dimensions of 3x3 and a stride of 2. This initial processing step reduces the storage size substantially by transforming a 224x224 input image into a 56x56 feature map [5].

Intermediate convolution part of the network: The intermediate convolution section, depicted in the curly-bracketed portion in the accompanying figure, is responsible for information extraction by stacking various convolutional layers. The specifics of the block parameters and the quantity of intermediate convolution sections vary across each network [6]. The notation x2, x3, and so on in the figure signify the number of repetitions of the block stacks.

Residual block: Each residual block contains two or three convolutional layers, and a direct connection across the Each residual block consists of either two or three convolutional layers, with a direct connection established between these layers. For a two-layer residual block (called a basic block), the input is first passed through a convolutional layer (usually a 3x3 convolution), then through a ReLU activation function, then through another convolutional layer, and finally added to the original input (dimensionally matched by a constant mapping or a 1-by-1 convolution) [7]. In three-layer residual networks, an additional 1x1 convolutional layer is incorporated in the middle to decrease the channel count and consequently reduce the computational workload.

2.2. Data Augmentation

2.2.1. Concept of Data Augmentation

Data Augmentation is a widely employed method for data processing, which is mainly used to solve the issue of insufficient data or data imbalance in machine learning and deep learning. The core idea of Data Augmentation is to generate new data samples by transforming the existing data to a certain extent, so as to expand the size of the datasets and the diversity of the samples in order to improve the generalization ability of the model [8].

2.2.2. The main advantages of Data Augmentation

Improve the generalization ability of the model by generating more samples, it allows models to better grasp the distribution characteristics of the data, thus improving its performance on unseen data [9].

Dealing with the problem of insufficient data in practical applications, data collection and labeling is a very time-consuming and laborious process, and Data Augmentation has the potential to enhance model performance when data is scarce [10].

2.2.3. Application in this paper

this paper applies the Data Augmentation technique to underwater fish identification by simulating the underwater dim environment through Data Augmentation and adding perturbations to the model training, which not only doesn't need to spend too much cost to collect a new dataset, but also makes the trained model more experimental for fish detection in the actual underwater environment.

3. Experimental Idea Design

In this paper, the function transforms. Random Grayscale is used as a variable to be added or removed in valid set and train set for experimental purposes. See 3.1, 3.2, 3.3, 3.4 for details. The rough flowchart is as follows in figure 2.

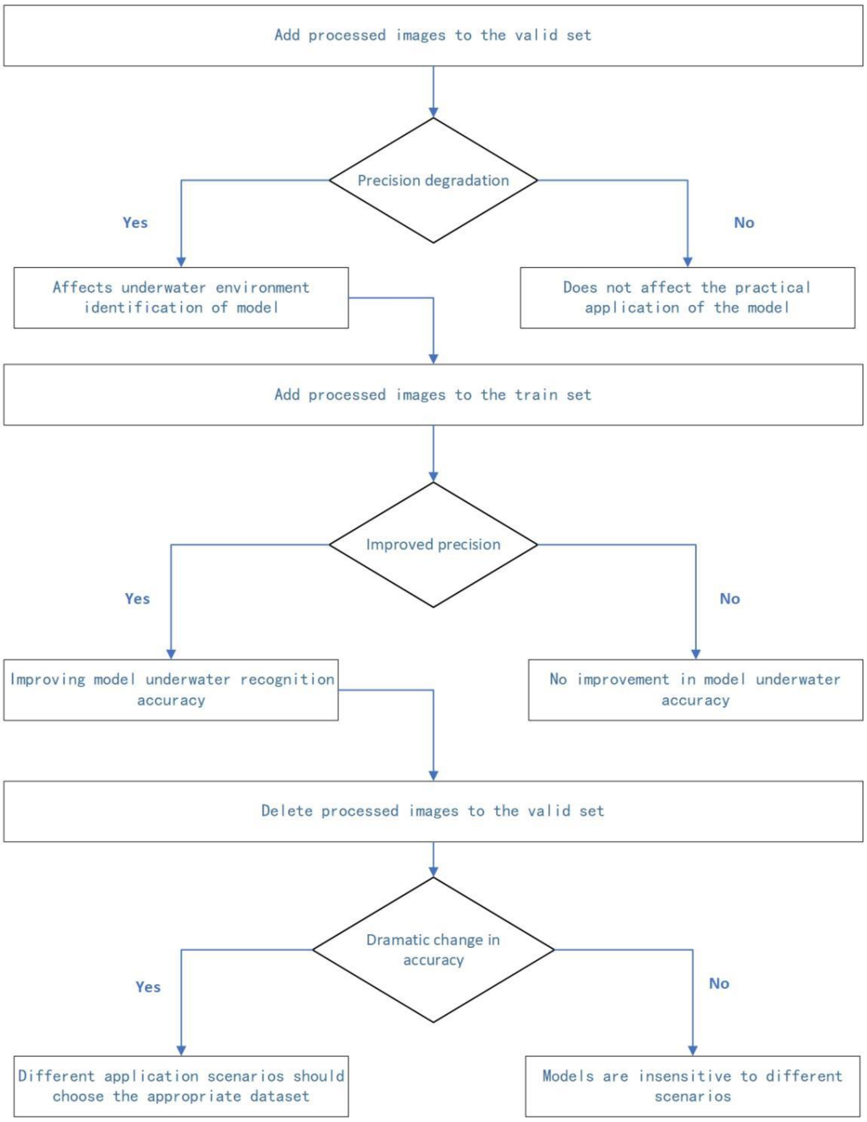

Figure 2. Flowchart of experimental ideas

3.1. Simulates underwater darkness by Data Augmentation

Data Augmentation using the transforms. Random Grayscale function can simulate dim conditions in underwater environments seen in figure 3, and this technique can provide underwater fish data that is closer to the real situation for the training of fish detection models.

Figure 3 Image under transforms Random Grayscale processing

Underwater, light is affected by water refraction and reflection, making it difficult for light to effectively penetrate the water surface, resulting in low light intensity and dim underwater environments, which will have an impact on the performance of the actual image recognition task. By randomly converting the original datasets into a grayscale image through this function, this light change can be effectively simulated, making the datasets more relevant to the real underwater environment.

3.2. Differences in the effectiveness of models trained with the original datasets

Random Grayscale function is added to the valid set, while the train set maintains the original images under ideal conditions, on which the model is trained. The method simulates whether the model under the training of the original ideal datasets, put into the real environment of the image classification situation will appear a decline in accuracy, if the accuracy of the decline, it indicates that the lack of the real dim environment of the datasets trained model in the actual image classification, there will be a decline in accuracy, resulting in the decline of the detection ability. If there is no change in the accuracy, it means that the model trained from the fish datasets under ideal conditions can be directly commissioned to the underwater dim environment for fish recognition work.

3.3. The effect of adding data-augmented images

This paper first tests the effect of adding data-augmented images under dim conditions in the train set for training on the recognition effect under real dim conditions. Random Grayscale function is also added to the train set and trained on the basis of the dim underwater environment that has been simulated for actual detection in 3.2. This method simulates the performance of the model trained under Data Augmentation when it is put into operation to recognize fish in the actual underwater environment. If there is an improvement in the accuracy compared to the data in 3.3, it means that the model trained with the Data Augmentation has an improvement in recognizing fish in real dim conditions, which is conducive to improving the performance of the model when it is put into the real recognition work. If there is no change in the accuracy, it means that the model trained with the Data Augmented to simulate the underwater dim environment added to the datasets has no improvement in the accuracy of underwater fish identification.

Then the effect of adding data-augmented images under dim conditions in the train set for training on the recognition of high-brightness, high-resolution images has been tested. The datasets under ideal conditions in the valid set is unprocessed, while the transforms.RandomGrayscale function is added to the train set. This method simulates the performance of fish identification under ideal conditions (fish identification on the water) for the model trained with the data added under dim conditions. If a large difference is obtained from the model trained with the original data in both the train and valid sets (in-water fish identification model trained with fish data under ideal conditions), it means that the model trained with different data sets under different conditions should be selected according to different fish identification application scenarios. If there is no large difference, it indicates that according to different fish recognition application scenarios, the models trained from datasets under different conditions should be selected to have less impact on the recognition accuracy.

4. Modeling and Experimental Results and Analysis

4.1. Modeling

Firstly, build the datasets Train set and valid set, see 4.2 for details, by adding the transforms.RandomGrayscale function to the datasets can simulate different scenarios of fish recognition applications; import ResNet (18, 34, 50, 101, 152) from models; then initialize the model, freeze the parameters of ResNet and change only the model output layer, then set the optimizer and create a PT file to record the time elapsed, train loss,valid loss, train acc and valid-acc as well as optimizer rate for each Epoch during training; start training the model by placing the data and labels into the GPU, followed by 20 Epochs and display the results; finally add training, unfreeze all the layers of the model then load the parameters trained in the previous 20 Epochs, and then train for another 20 Epochs. The operation flow chart is as follows in figure 4.

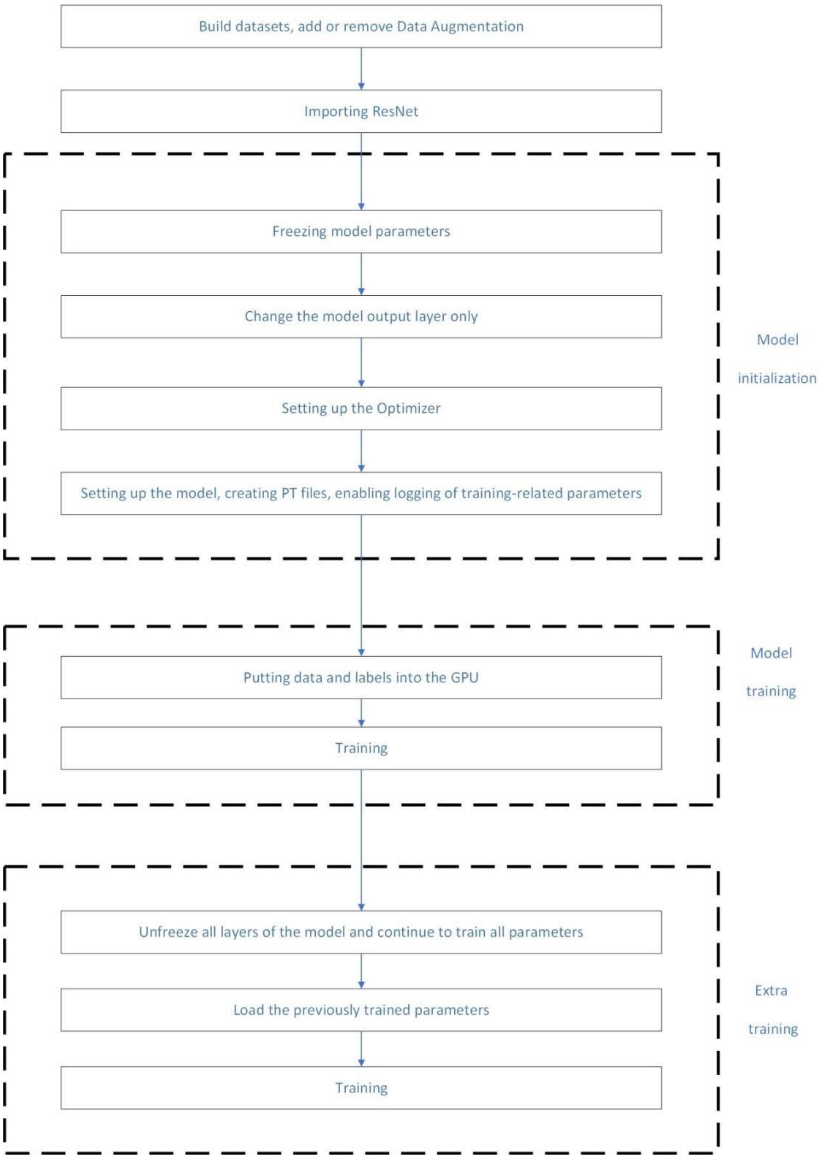

Figure 4. Flowchart of modeling code

4.2. Introduction to the dataset

The datasets are from kaggle at https://www.kaggle.com/datasets/markdaniellampa/fish-dataset. The total number of images in this datasets is 11,165, of which the train set contains 8,509 and the valid set contains 2,656, containing the following 31 species Fish (Bangus, Big Head Carp, and Black Spotted Barb,etc.) [11].

4.3. Experimental data and analysis

4.3.1. Performance of the model trained on the original datasets applied in practice

The transforms.RandomGrayscale function is added to the valid set, while the train set maintains the original images under ideal conditions, on which the model is trained. The method simulates the performance of the model under the training of the original ideal datasets, put into the real environment, and compares it with the model trained with the original datasets (simulated fish recognition on water with the ideal condition datasets).The test consisted of 4 scenarios as follows, Scenario1: The data set is not processed and only the output layer is changed; Scenario2: Data sets are not processed and all layers are unfrozen (training is added on top of changing only the output layer); Scenario 3: The valid set randomly reduces the luminance p=0.5 and only changes the output layer; Scenario 4: Valid set randomly reduces brightness by p=0.5 and unfreezes all layers (adding training on top of changing only the output layer). The test results are presented in Table 1.

Table 1. Comparison of the accuracy between processed valid set and unprocessed train set

	Scenario1	Scenario2	Scenario3	Scenario4
ResNet18	0.404580	0.639040	0.299164	0.464922
ResNet34	0.430752	0.681934	0.331516	0.541985
ResNet50	0.513631	0.679389	0.382770	0.565613
ResNet101	0.503453	0.529262	0.383497	0.476190
ResNet152	0.521628	0.657943	0.389676	0.502363

And the diagram is shown in figure 5.

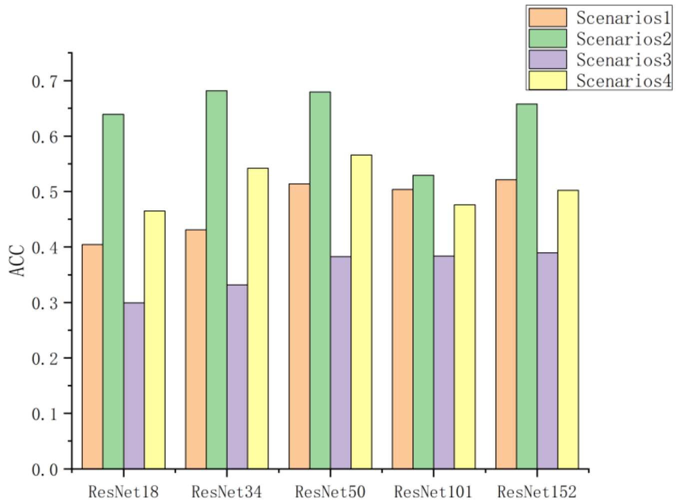

Figure 5. Bar chart of Table1

Based on the information provided in Table 1 and Figure 5, this paper can derive the following observations and conclusions:

(1) Regardless of whether the datasets are processed or not, or which ResNet network is used, the accuracy of the model trained using the unfrozen ResNet surpasses the accuracy achieved before unfreezing, particularly when dealing with datasets that have not processed, the accuracy of the model trained by the ResNet34 improves most obviously, up to 25%. This suggests that with this fish dataset, the models trained by the unfrozen ResNet network are more effective for fish recognition, regardless of the environment in which they are located.

(2) The models trained with ResNet50 gave roughly the best results regardless of whether the datasets were processed or not, and regardless of whether the ResNet was unfreezing or not. This suggests that with these datasets, model training for fish identification is not positively correlated with the number of ResNet network layers, and that using a neural network with too many layers can lead to problems such as overfitting and high computational costs, thus allowing the use of ResNet50 with a moderate number of layers.

(3) Regardless of whether ResNet is unfrozen or not, the accuracy of the model trained with the valid set processed is significantly less than the accuracy of the model trained without the processed datasets. This indicates that when the model trained from the datasets under ideal conditions is applied to the actual dim underwater environment, the accuracy will be significantly degraded under the interference of the lighting factor.

4.3.2. The effect of adding data-augmented images in both the train and valid sets

RandomGrayscale function in both the train and valid sets, this method simulates the performance of the model trained by adding fish images under real dark conditions and transporting them underwater, and compares the accuracy of the trained model (which simulates fish identification under ideal conditions) with that of the model trained by adding only the transforms.RandomGrayscale function in the valid set. The test consisted of 4 scenarios as follows, Scenario1:The valid set randomly reduces the luminance p=0.5 and only changes the output layer;Scenario2:The valid set randomly reduces brightness by p=0.5 and unfreezes all layers (adding training on top of only changing the output layer);Scenario3:The train and valid sets are trained with random brightness reduction samples p=0.5 and only the output layer is changed;Scenario4:Both train and valid sets are trained with random brightness reduction samples p=0.5 and all layers are unfrozen. The test results are presented in Table 2 below.

Table 2. Comparison of accuracy of simultaneously processed train and valid sets

	Scenario1	Scenario2	Scenario3	Scenario4
ResNet18	0.299164	0.464922	0.354417	0.573610
ResNet34	0.331516	0.541985	0.393312	0.628499
ResNet50	0.382770	0.565613	0.444202	0.628499
ResNet101	0.383497	0.476190	0.447110	0.474009
ResNet152	0.389676	0.502363	0.458379	0.506725

And the diagram is shown in figure 6.

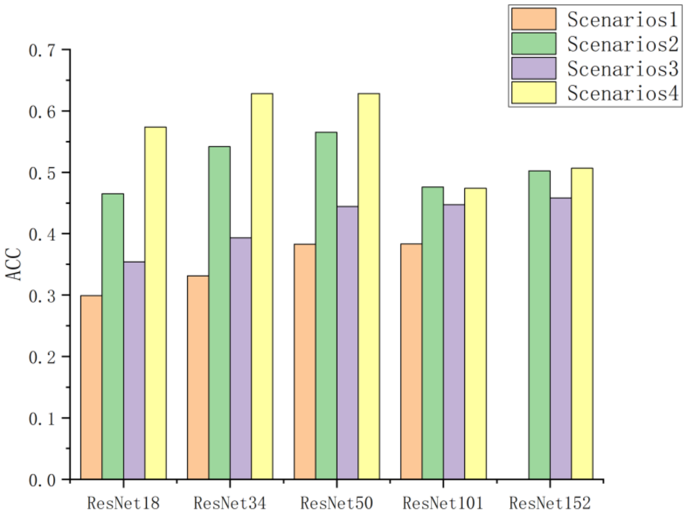

Figure 6. Bar chart of Table2

Based on the information provided in Table 2 and Figure 6, we can derive the following observations and conclusions:

(1) The accuracy of the model trained by adding the transforms.RandomGrayscale function to both the train set and the valid set is improved compared to the model trained by adding the transforms.RandomGrayscale function to the valid set only. RandomGrayscale function in the valid set. This shows that the model trained with the augmented datasets has the effect of improving the recognition in real dim conditions, which is conducive to improving the effectiveness of the model to be put into the real recognition work.

(2) On the basis of the conclusion obtained in (1), the Augmentation effects of ResNet18, ResNet34, and ResNet50 are more obvious, while the Augmentation effects of ResNet101 and ResNet152 are smaller. This indicates that on the basis of this datasets, the Data Augmentation with transforms.RandomGrayscale function to simulate the images of real underwater environment has a better Augmentation effect for the ResNet network with low number of layers.

(3) The models trained with ResNet50 give roughly the best results regardless of whether the datasets are processed or not, and regardless of whether ResNet is unfrozen or not. This suggests that the method of Data Augmentation based on this datasets, using the transforms.RandomGrayscale function to simulate images of real underwater environments, is more suitable for ResNet50.

4.3.3. The effect of adding data-augmented images in the train sets

The datasets under ideal conditions in the valid set is unprocessed, while the transforms. RandomGrayscale function is added to the train set. This method simulates the performance of fish recognition under ideal conditions (over-water fish recognition) for the model trained with the Data Augmentation under dim conditions. The test consisted of 8 scenarios as follows, Scenario1:No Data Augmentation and only the output layer is changed; Scenario2: No Data Augmentation and unfreeze all layers;Scenario3:Train set with Data Augmentation (p=0.5) and only changes the output layer of network; Scenario4:Train set with Data Augmentation (p=0.5) and unfreezing all layers of network; Scenario5: Train set with Data Augmentation (p=0.8) and only changes the output layer of network; Scenario6: Train set with Data Augmentation (p=0.8) and unfreezing all layers of network; Scenario7: Train set with Data Augmentation (p=0.2) and only changes the output layer of network; Scenario8: Train set with Data Augmentation (p=0.2) and unfreezing all layers of network. The test results are presented in Table 3.

Table 3 Comparison of the accuracy between processed train set and unprocessed valid set

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
ResNet18	0.404580	0.639040	0.408579	0.571792	0.346419	0.530353	0.426390	0.637950
ResNet34	0.430752	0.681934	0.434387	0.654308	0.378408	0.599418	0.449291	0.718284
ResNet50	0.513631	0.679389	0.492548	0.648855	0.429662	0.646674	0.515449	0.692475
ResNet101	0.503453	0.529262	0.498728	0.564158	0.423119	0.554707	0.523083	0.469284
ResNet152	0.521628	0.657943	0.513268	0.600872	0.451836	0.468920	0.528899	0.566703

And the diagram is shown in figure 7.

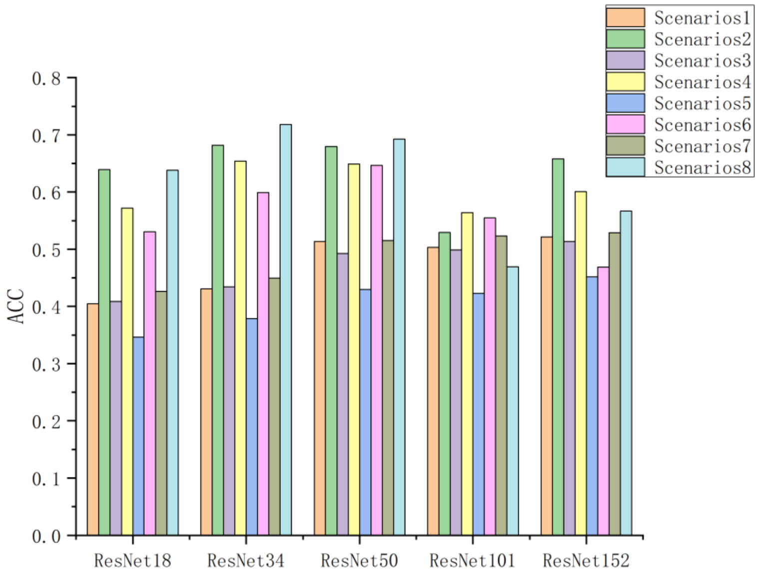

Figure 7. Bar chart of Table3

Based on the information provided in Table 3 and Figure 7, we can derive the following observations and conclusions:

(1) When the probability of transforms.RandomGrayscale function p=0.2, the model trained by ResNet network with lower number of layers has a good lifting effect. This indicates that, based on this datasets, the over-water fish recognition model can be trained using a ResNet network with a lower number of layers and adding Data Augmentation of the transforms.RandomGrayscale function to improve the accuracy of over-water fish recognition.

(2) When the probability of the transforms.RandomGrayscale function is p=0.5 and 0.8, the accuracy of the model trained by the ResNet network in the over-water fish identification is generally decreasing. This suggests that the model applied to fish recognition on water should choose a smaller value of p when using the Data Augmentation of transforms.RandomGrayscale as a way to enhance the accuracy of recognition. Although the model trained at p=0.5 has an Augmentation effect on fish recognition in underwater dim environment, the fish recognition accuracy on water decreases,

therefore, in different fish recognition scenarios, the model trained under different conditions should be selected to adapt to the interference of the environment on the fish recognition accuracy.

5. Conclusion

Due to the differences in lighting conditions, applying models trained using fish images under ideal conditions on water directly to real dim environments underwater can lead to a decrease in recognition accuracy. Data Augmentation using the transforms.RandomGrayscale function to simulate fish images under underwater dim conditions is an effective strategy. In this way, the model can be exposed to more data similar to the real underwater environment in the training phase, thus improving its accuracy in recognizing fish in underwater dim environment in real applications. Compared with collecting a large number of underwater fish images in real environments, this Data Augmentation method is not only less costly, but also achieves similar results, providing a more convenient and economical option for practical applications. In addition, when dealing with the fish recognition task on water, the transforms.RandomGrayscale function with lower p-value can still be utilized for Data Augmentation to improve the accuracy of fish recognition on water. Finally, in different fish recognition scenarios, the models trained under different conditions should be selected to adapt to the interference of the environment on the fish recognition accuracy.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Mangeruga M, Cozza M, Bruno F. Evaluation of underwater image enhancement algorithms under different environmental conditions. Journal of Marine Science and Engineering, 2018, 6(1): 10.

[2] Nair R S, Agrawal R, Domnic S, et al. Image mining applications for underwater environment management-A review and research agenda. International Journal of Information Management Data Insights, 2021, 1(2): 100023.

[3] Mahmood R, Lucas J, Acuna D, et al. How much more data do i need? estimating requirements for downstream tasks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 275-284.

[4] Zhang L, Schaeffer H. Forward stability of ResNet and its variants. Journal of Mathematical Imaging and Vision, 2020, 62: 328-351.

[5] Jafar A, Lee M. High-speed hyperparameter optimization for deep ResNet models in image recognition. Cluster Computing, 2023: 1-9.

[6] Zhu Z, Zhai W, Liu H, et al. Juggler-ResNet: A flexible and high-speed ResNet optimization method for intrusion detection system in software-defined industrial networks. IEEE Transactions on Industrial Informatics, 2021, 18(6): 4224-4233.

[7] He F, Liu T, Tao D. Why resnet works? residuals generalize. IEEE transactions on neural networks and learning systems, 2020, 31(12): 5349-5362.

[8] Wong S C, Gatt A, Stamatescu V, et al. Understanding data augmentation for classification: when to warp?. 2016 international conference on digital image computing: techniques and applications (DICTA). IEEE, 2016: 1-6.

[9] Li P, Li D, Li W, et al. A simple feature augmentation for domain generalization. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 8886-8895.

[10] Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings, 2022, 3(1): 91-99.

[11] Lampa M D, Librojo R C, Calamba M M. “Fish dataset.” Kaggle, [online] Available: https://www. kaggle. com/datasets/markdaniellampa/fish-dataset/data, 2022.