Building a Machine-Learning Model to Classify Three Different Types of Fish

Rittik Saha
Energy and Electronic Engineering
University of Portsmouth
Portsmouth, UK
UP2038861@myport.ac.uk

***Abstract*—To preserve marine biology needs effective management system with quick response mechanism to tackle urgent issues which are damaging marine ecosystem. To tackle this problem and make management system more efficient, we are using machine learning tool known as convolutional neural network (CNN) for image recognition, so that the system can precisely classify different fishes, which it then can help to take timely decision and support better decision making for preserving marine ecology.**

Keywords—Machine learning, Marine Biology, Matlab, Deep Network Designer, Convolutional neural network (CNN), Image Recognition, Classification, VGG-16, AlexNet, ResNet-50, Transfer learning.

I. Introduction

Two-thirds of earth which is nearly 70% is covered with oceans. There are vast varieties of fishes one can observe in marine ecosystem. Conservation of marine biology is critical, so to address this problem and conserve marine biodiversity, the rise of using artificial intelligence, especially deep learning tool such as convolutional neural network (CNN) is an important revolution in field of computer vision and image data analyzation [1].

CNN models can classify any image near to perfection with the help of its different layers which performs different functions such as convolution layer, max-pooling layer, and fully connected layer. With the help of provided image datasets we can train our CNN models, which can later provide better accuracy in predicting different types of fishes.

To identify different types of fish based on provided data, we have used pre-trained CNN models as it is proved to be efficient in image recognition, less time consuming compared to its counterpart which is built from scratch, needs huge amount of data to be trained first which is expensive to manage and time consuming [1].

II. Literature Review

Deep CNN models have shown outstanding result in predicting accuracy of a model with loss function as low as possible. We have only used pre-trained CNN models in this paper as it is pretrained extensively with enormous amount of data with labeled images. These models are suitable for image processing, and classification of any sets of data once it has been trained with weights as well as bias in pre-trained phase. If necessary, new model can be built from these pre-trained models as it will be inexpensive and more valuable can be learnt [1].

The pre-trained CNN models which have shown great result in image recognition and analysis are VGG, ResNet, AlexNet, Inception, GoogleNet and SqueezeNet. In this research paper, we will discuss mainly about VGG-16, ResNet50 and AlexNet models.

A. VGG-16 Model:

The Visual Geometry Group (VGG) has two popular model variations which are VGG-16 and VGG-19 both are trained to handle massive amount of image recognition tasks. Specifically, the VGG-16 model is trained on ImageNet database which has around 14 million images divided in 2000 categories. This model is made up of 16 layers in which 5 are convolution blocks (each convolution block contains 2 convolutional layers), a max-pooling layers and the fully connected layer (FC) performs the main function which is classifying the images. One of the characteristics of VGG-16 model is with the increasing number of filters in convolutional layer after each max-pooling layer, increases volume of depth of the whole system [1].

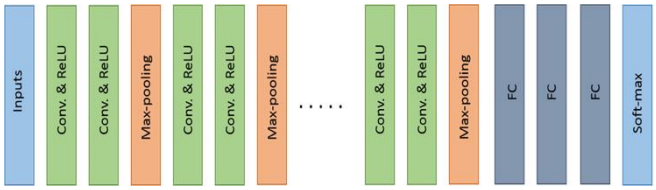

Fig.1. Basic building block of VGG-16 model

B. ResNet50 Model:

The ResNet Model was first developed by Microsoft research group [1]. This model is built to address the problem of vanishing gradient. The structure of ResNet50 model consists of 49 convolution layers and 1 fully connected layer connected at the end [2]. At first, it was assumed stacking of layers could possibly give better and deeper network, but the reality was different, it was providing large training error. To address this problem effective residual block mapping is necessary [2] which is shown in Fig.2.

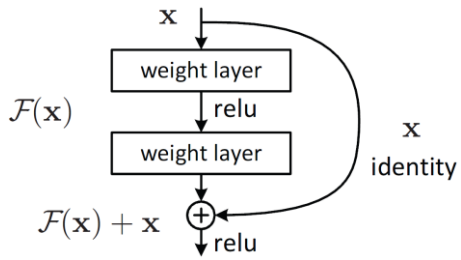

Fig.2. Residual Block

Since then, many advance architectures have been proposed with the combination of different Inception and Residual blocks [2].

C. AlexNet Model:

The AlexNet is first proposed by Alex Krizhevesky in 2012. It was a first big breakthrough in the field of computer vision which can perform the task of classification and visual recognition [2].

The architecture of AlexNet is given in Fig.3. This model is divided in 7 layers. In the first convolutional layer, the

given data is convoluted and max-pooled with Local Response Normalization (LRN) of 96 different receptive filter which are 11×11 in size and max-pooling uses 3×3 filters with stride size of 2. In second layer, system uses 5×5 filters and the next 3 consecutive convolutional layers use 3×3 filters with 384, 384 and 256 feature maps respectively. These layers then connected with two fully connected layers (FC) followed by Soft-max layer [2].

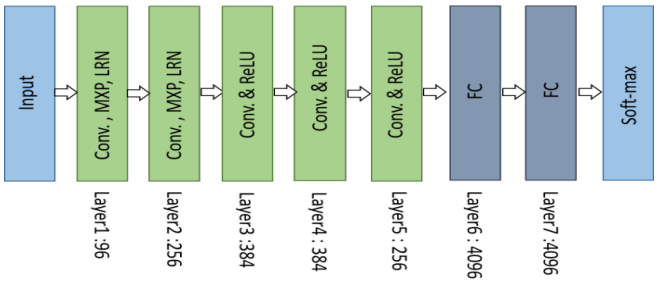

Fig.3. Architecture of AlexNet

III. Materials and Methods

In this paper, we have identified three different models VGG-16, ResNet50 and AlexNet using MATLAB's deep network designer tool to identify different types of fish based on provided data. The steps we need to follow beforehand to get a clear picture of the algorithm which is more efficient.

A. *Data-set*

First, we need to import provided dataset that contains three different types of fish data which are Dace, Perch, and Roach in our Matlab workspace. In total, we have 252 sets of images in three different types of fish data which is not enough for training a model for convolutional neural network (CNN). Image augmentation can be a solution for lack of availability of data. Using this image augmentation technique, we can generate extra images. We can achieve that by changing the reflection axis, or by rotating and rescaling the image [3].

In three different models VGG-16, ResNet50 and AlexNet, we have used random reflection axis as 'X', minimum and maximum random rotation are -90 degrees and 90 degrees respectively and for random rescaling we have used minimum 1 and maximum 2 to generate those extra images for training. Then, split the training data set between training and validation in 70:30.

B. *Methodology*

We have used pre-trained convolutional neural network as it is cheap and quick to build compared to training new model from scratch. In this paper we will use transfer learning for classification and image processing. Transfer learning is a process of using pre-trained models which were already been trained on big datasets, along with connected new layer of defined weights and biases. This kind of structure makes our network cheaper and efficient which can speed up our training process [2].

In three different models VGG-16, ResNet50 and AlexNet, we have used the same weight and bias as 10 in fully Connected (FC) layer. To train the system, we take initial learning rate as low as possible 0.0001 because if we take a larger learning rate then, the network will diverge and will not reach the minima, so it is recommended to take learning rate as low as possible to converge at the minima [2]. Then, train the network after setting the Validation frequency to 3, Max Epochs to 10 and Mini Batch size to 26.

IV. Results

We have used Matlab different deep network models such as VGG-16, ResNet50 and AlexNet for classification of different kind of fishes such as Dace, Perch, and Roach (not by using fish images provided in dataset). In this way, we will compare which one of these models is providing us the best accuracy with minimum loss.

A. *VGG-16 Model:*

This model is providing us a validation accuracy of 85% (Fig.4) with validation loss 0.3458 (Fig.5) with 41 layers.

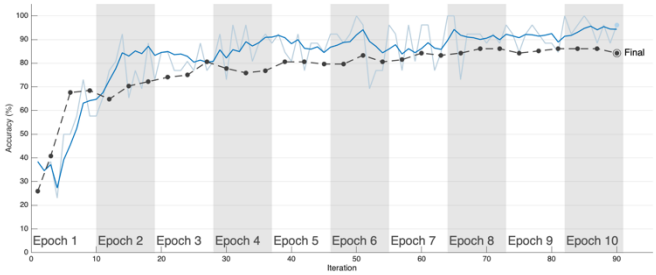

Fig.4. Validation Accuracy (%)

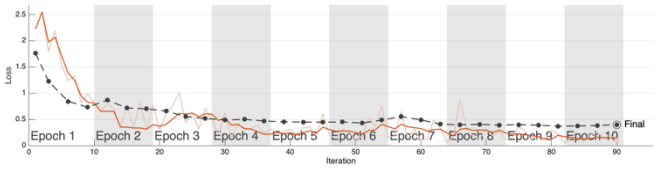

Fig.5. Validation Loss

Now we will measure how well this model can perform the task of classification and image recognition. For that reason, we have downloaded some pictures of fishes from the internet (not from the given datasets) to classify their name using the model along with the percentage of the model rightly guessed the image.

Fig.6. Classification of Images using VGG-16

From the above Fig.6., it is clearly proven that the VGG-16 model classify all fishes name correctly.

B. *ResNet-50 Model:*

This model is providing us a validation accuracy of 78.7037% (Fig.7) with validation loss 0.5667 (Fig.8).

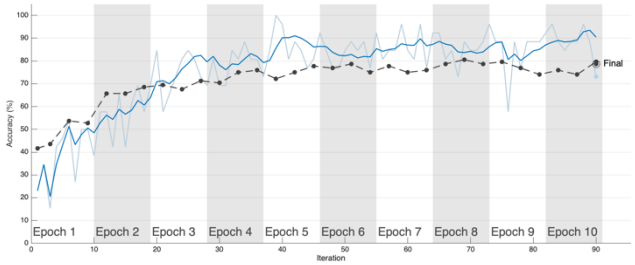

Fig.7. Validation Accuracy (%)

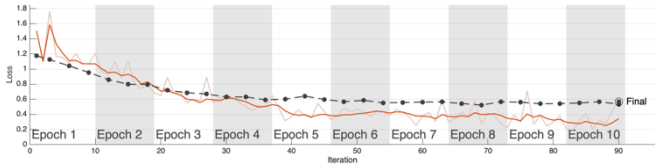

Fig.8. Validation Loss

Now we will measure again how well this model can perform the task of classification and image recognition.

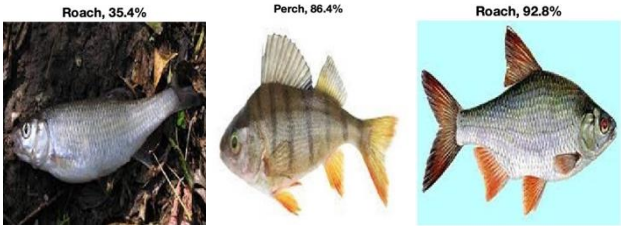

Fig.9. Classification of Images using ResNet-50

We can clearly observe from Fig.9 that this model has made a mistake in image recognition and classification for the first image Roach (35.4%), which is an image of Dace. So, from this above analysis it is proven that this model is not well suitable for this exercise.

C. AlexNet Model:

This model is providing us a validation accuracy of 74.0741% (Fig.10) with validation loss 0.6781.

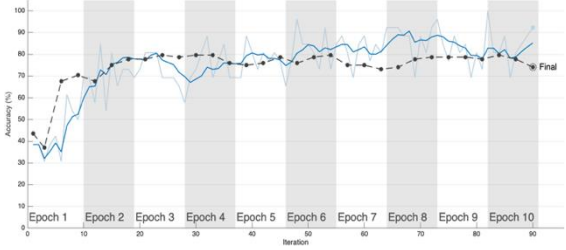

Fig.10. Validation Accuracy (%)

Now we are going to test the model in terms of its classification and image recognition.

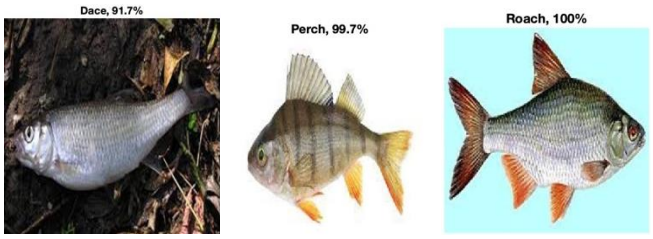

Fig.11. Classification of Images using AlexNet

It is proven from the above experiment that, AlexNet image recognition and classification skill is up to the mark. This model classifies all the images perfectly.

From the above discussion, we have observed that VGG-16 and AlexNet are two of the models which have shown more accuracy in image recognition and classification, but VGG-16 is the most efficient convolutional neural network because it has large percentage of validation accuracy with a smaller number of Validation Loss.

V. Conclusion

Machine learning is a broad subject where still many areas left to discover. In this project, we have used pre-trained convolutional neural network (CNN) for image classification and try to show how computer vision works with an experiment. After the experiment, we get an idea that validation of accuracy does not matter always for computer's vision but short and effective layering with right weight and bias can make a huge difference in learning. In my future project, I would like to research more on these issues, so that I can build a better model next time.

Reference

3. [illegible]

1. Agarwal, A. K., Tiwari, R. G., Khullar, V., & Kaushal, R. K. (2021, August). Transfer learning inspired fish species classification. In *2021 8th International conference on signal processing and integrated networks (SPIN)* (pp. 1154-1159). IEEE.
2. Alom, Md Zahangir, et al. "The history began from alexnet: A comprehensive survey on deep learning approaches." *arXiv preprint arXiv:1803.01164* (2018).
3. Ahmed, Tawsin Uddin, et al. "Facial expression recognition using convolutional neural network with data augmentation." *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. IEEE, 2019