Fish Species Identification in Unconstrained Underwater Environment

By

Nirmal Tariq

NUST201464114MSEECS61314F

Supervisor:

Dr. Faisal Shafait

A thesis submitted in partial fulfillment of the requirements for the degree of MSCS

School of Electical Engineering and Computer Science,

National University of Sciences and Technology (NUST), Islamabad,

Pakistan.

(August 23, 2016)

Dedication

I dedicate my work to my parents whose endless support brought me here and my husband for his encouragement throughout my research period.

Approval

It is certified that the contents and form of the thesis entitled "**Fish Species Identification in Unconstrained Underwater Environment**" submitted by **Nirmal Tariq** have been found satisfactory for the requirement of the degree.

Advisor: **Dr. Faisal Shafait**

Signature: ______________

Date: ______________

Committee Member 1: **Dr. Ahmad Salman**

Signature: ______________

Date: ______________

Committee Member 2: **Dr. Anis ur Rehman**

Signature: ______________

Date: ______________

Committee Member 3: **Dr. Asad Anwar Butt**

Signature: ______________

Date: ______________

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at National University of Sciences & Technology (NUST) School of Electrical Engineering & Computer Science (SEECS) or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: **Nirmal Tariq**

Signature: ________________

Acknowledgments

I would firstly like to thank my thesis advisor Dr. Faisal Shafait of School of electrical Engineering and Computer Science at National University of Sciences and Technology. His guidance throughout my research enlightened me to do well and his support during all this tenure encouraged me to gain the best out of it.

Then I would like to thank all my GEC members Dr. Ahmad Salman, Dr. Anis ur Rehman and Dr. Asad Anwar Butt for their expert guidance and feedback in time. I would alsoo like to acknowledge all the members of TUKL-NUST lab for their participation and valuable comments on the thesis.

Nirmal Tariq

Abstract

Preservation of biodiversity is of utmost importance and one of the major ecosystem which is highly in danger and a notable cause of disturbing biodiversity is freshwater ecosystem. Marine researchers and scientists along with government bodies are developing ways to monitor and preserve the underwater living organisms for a sustainable environment. Many techniques have been developed for underwater monitoring and classification but with advancement in knowledge the trends are changing rapidly. An automated system for underwater fish monitoring and classification is needed but the underwater environment is highly variable and complex which poses a great challenge for such automated systems to accurately monitor and identify the inhabitants. We proposed a variant of deep learning method namely Convolutional Neural Network for specie identification in unconstrained underwater environment. We demonstrated how the convolutional layers within a network can provide useful information if used appropriately. The classification was done using dataset obtained from University of Western Australia (UWA). We indicated many reasons that the proposed approach can potentially be used for further classification tasks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Create an accurate knowledge of the identity, geographical distribution and evolution of life is indispensable for a sustainable development of mankind and for the conservation of biodiversity. It has become evident that human existence on earth is possible only if we take care of the biodiversity. It is also important to conserve biodiversity for the sake of our own inquisitiveness. Biodiversity can also be called as life support system of our planet. We depend on it for every major aspect of life, the air we breathe, the food we eat, and the water we drink. Unfortunately, this basic information is not fully revealed to human beings, but to see a slow and often incomplete to ecosystems that holds the highest diversity. A notable cause and consequence of this knowledge is rare plant identification or live animals is usually impossible for the general public, and it is often hard work of professionals, such as farmers and fish farms or forests as natural disasters and specialists themselves. This gap is known taxonomic actually is one of the major challenges, requiring efforts to resolve is identified in 1992 in the Rio's United Nations Conference (Lif, 2014).

Fresh water ecosystems are the most threatened ecosystems worldwide. The declines in biodiversity in any affected terrestrial ecosystem is far less than the declines occurring in fresh water ecosystems (Sala et al., 2000). What are the factors that makes freshwater habitats and the biodiversity they support particularly sensitive to human activities and environmental variations? The main reason is that the abundance of inland water as a habitat for animals and plants is out of proportion (Dudgeon et al., 2006).

Every ecosystem has its own threats and the threats to fresh water ecosystems lies under the following five categories: (see Figure 1.1)

1. Over exploitation
2. Water pollution
3. Flow modification
4. Destruction or degradation of habitat
5. Invasion by exotic species.

Over exploitation essentially affects vertebrates, largely fish and reptiles while the other four categories of threats effects all freshwater biodiversity from microbes to megafauna. When a particular specie is over exploited it has tendency of becoming extinct so disturbing the ecosystem as a whole.

As opposed to traditional methods of manual photography or net-casting, underwater video capturing has been extensively used by marine researchers as it not only captures the images but also provide information regarding fish behavior and underwater environment. Although the video data generated daily is enormous and it is impractical for humans to analyze it manually as it will be error prone and require much time and concentration. An automated fish identification and classification system is of critical importance, which can estimate not only the fish existence and quantity but also recognize its specie for monitoring and its preservation.

1.1 Challenges

Under water environment is highly variable in terms of light intensity, water conditions and depth so taking account of organisms living there is quite a difficult job. There are thousands of fish species whose biomass estimation will be of utmost importance to marine scientists and certain fisheries as well. Many of the species would be in abundance at certain areas whereas near to extinction at others so change in

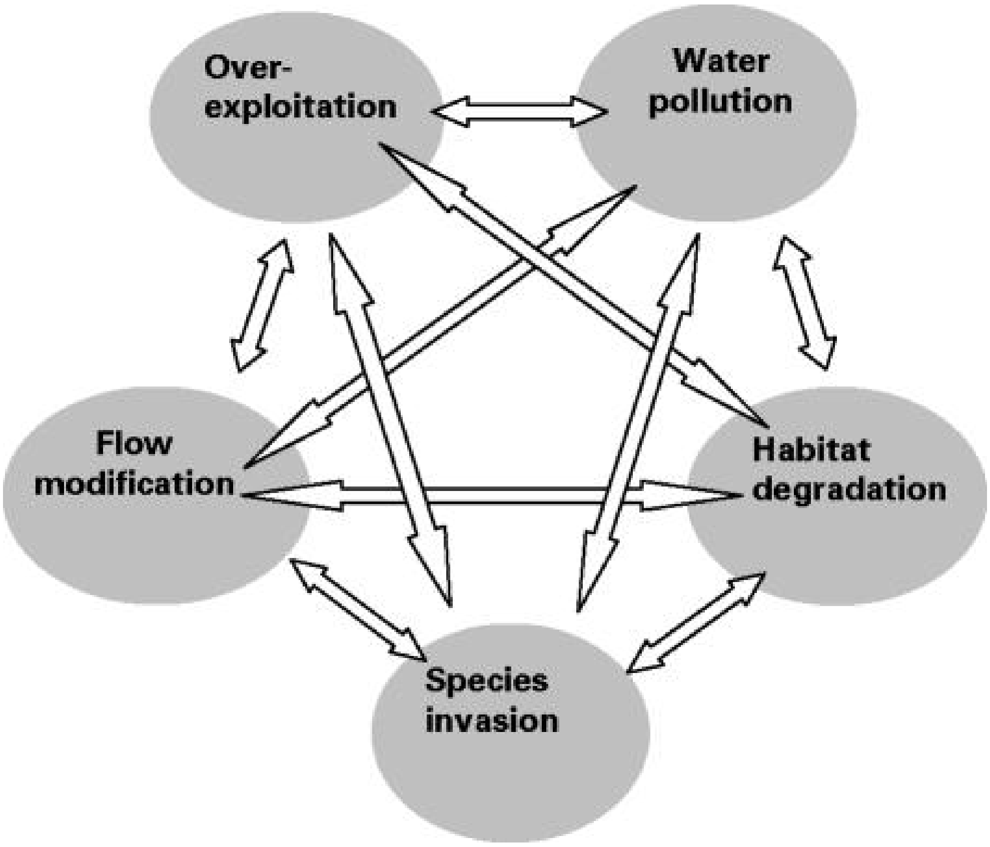

Figure 1.1: Five major threats categories of fresh water ecosystem

abundance is an important factor to be known for taking particular decisions like enforcing fishing bans at particular regions to preserve the specific specie. so the regular sampling is needed to keep record of the changing trend (Jennings and Kaiser, 1998).

Most of the work for monitoring of these species is done manually which is not only expensive in terms of tools and labor but also a danger to fish life causing the mortality. For this purpose underwater videos and still cameras are being used extensively by scientists and researchers to measure the fish count and floral abundance in such environment. But still the classification of fish specie is done manually after taking the records from videos and cameras which is the most tedious and exerting thing to do.

In Pakistan the number of unique fish specie has reached to 34 which of them mostly are endangered. The task of monitoring fish specie as well as keeping their count is so far done marginally and manually. Therefore an automated fish identification and classification system is urgently needed to replace the present error-prone procedures (Lee et al., 2003).

1.2 Objectives

Developing an automated fish identification and classification system will have many worldwide applications. For example, this system will help marine researchers in continuous monitoring of underwater inhabitants and can enforce fishing bans if a certain specie reaches its margin of extinction. In this way the system can be used for preservation of fresh water biodiversity and studying changing trends in underwater environment. The major goals can be summarized as follows:

1. The proposed project makes use of technology to overcome the problems of sampling for fish biodiversity estimation. Current methods in use are either destructive, or so expansive that they are not feasible for use in Pakistan.

2. This projects aims at providing a low cost solution, which causes no harm to fish habitats.

3. With minimal training, data can be collected on endangered species and their habitats to carry out preventive actions to save them.

4. Our goal is to commercialize the product for maximum use, and we will also provide the required training.

1.3 Main Contributions

Fish Species identification in unconstrained underwater environment is a task that a fish is being detected in the image and a bounding box has been created around it and it needs to be classified in one of the species defined. Many researchers have used different techniques in order to create an automated system. These classification systems were based on texture, shape combined with texture and then further combined with color but all these feature based classifications have their own limitations. So the researchers have diverted their attention to a new paradigm which is machine learning algorithms for such complex tasks of classification.

Machine learning is a trending field and researchers have used many algorithms for the classification tasks such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) etc but all such systems were used under controlled environment and unconstrained underwater environment poses a lot more complexity to the task including noise, light intensity variation and changes in orientation.

Deep learning in recent years have been able to gain attention of researchers as it offers multiple layer architecture which can cope all the complexities produced due to uncontrolled underwater environment. Deep Convolutional Neural Networks (CNN) has proved to show drastic improvement in classifying fish specie in unconstrained underwater environment. Features which are learned using CNN are proved to be robust against environmental variability. The CNN application fulfills two major aims:

- Learning fish specie dependent features.
- Applying a feasible classification approach on learned features.

Using Deep Convolutional Neural Network for such complex task like fish species classification has opened doors to a new era of research and this method with certain changes can prove to get drastically improved results and we propose to use the CNN architecture to learn local features in the image that can prove to get better results.

Chapter 2

Related Work

Fresh water ecosystems are threatened due to the over-exploitation of different living organisms in different habitats. The abundance of one type of species in any area does not mean that the particular specie is in abundance overall. Fishing is the major source of over-exploitation and it has direct as well as indirect effects on seabed communities, theirs habitats, structure and mass production. In some areas these effects are mostly vivid and are long lasting in the areas where natural disturbances do not occur frequently. initiation of fishing and increasing its intensity creates a dramatic change in the structure of fish community in particular area. So the regular sampling of fish population in underwater systems is of utmost importance so as to monitor the relative count and abundance of key species in any area (Jennings and Kaiser, 1998).

2.1 Underwater Monitoring Techniques

The sampling in underwater environment should be done in a way as it would not harm the living population in the seabeds. Hence, many non-destructive sampling techniques have been discovered and has grasped the attention of researchers which provides them the required information without paying any cost to biodiversity. The non-destructive technique involving different configurations for both stereo video cameras as well as single cameras have been used widely ever since where the image quality has drastically improved in last few decades and digital technology have replaced the older technology (Gibson et al., 2016).

While discussing the underwater video monitoring techniques there have been many of them which are being used widely under different circumstances. These techniques are easily and readily accessible to local researchers for which they can use to estimate the biomass distribution in marine protected areas and to monitor the subsequent changes in that area (Cappo et al., 2003). Underwater monitoring techniques varies indefinitely in terms of their application (Mallet and Pelletier, 2014) which includes:

- Behavioral Study of animals
- Effect of human-induced interference
- Spatial and temporal patterns of fish biomass
- Habitat mapping
- Benthos monitoring at species level

So the underwater video surveillance techniques have mastered based on above application areas.

2.1.1 Remote Underwater Video (RUV)

For behavior related research RUV (Remote Underwater Video) has been used extensively and recently it was used to record species response to habitat and environmental conditions.

2.1.2 Baited Remote Underwater Video (BRUV)

While particularly for fish size estimation and their species responses the BRUV (Baited Remote Underwater Video) technique has been used recently.

2.1.3 TOWed Video (TOWV)

For habitat mapping TOWV (TOWed Video) has been used and focused the study on benthic environment.

2.1.4 Diver Operated Video (DOV)

Then there is this divers' based video gaining technique DOV (Diver-Operated Video) which supported multiple applications such as estimation of fish abundance, habitat mapping and monitoring fish behavior as well.

2.2 Cost Effectiveness of Monitoring Techniques

The underwater video monitoring techniques have proven to be cost effective and efficient means of fish assembling and monitoring (Murphy and Jenkins, 2010). Cost estimation accounts not only the manufacturing cost of equipment but also the total time spent in the field and in the laboratory. Mostly the techniques which provides better results in detecting and identification of fish require more time in image analysis but this cannot be considered as a disadvantage unless the technique is giving better results. Also note that time spent at sea is always more expensive than in laboratory, so a good monitoring technique is always cost efficient and provides a way of repeatable sampling.

Once the data is collected from underwater environment either in the form of videos or still images, it has to be analyzed. Manual processing of resulting data will create a time lag and effect the cost efficiency of the monitoring techniques and cause a hurdle in the uptake of this technology. Recent developments show that researchers aim to completely eliminate human involvement in the process of identifying and counting the fish population in the acquired underwater videos. An automated system will deal with the hourly long stereo videos without any error and will make the technology more usable and improve the efficiency of image analysis (Shortis et al., 2013).

2.3 Automated System for Underwater Imagery Monitoring

An automated system dealing with stereo-video imagery acquired from underwater environment can be used for the following two tasks:

1. Detection of fish in the video frames.

2. Classification of detected fish according to the species.

Former deals with differentiating a fish from non-fish substances present in the video. The non-fish category involves the background of videos including coral reefs, sea grass beds and other sea plants such as mangroves, marine algae and also other vertebrates living on the sea beds. Whereas latter aims to deal with distinguishing fishes on the basis of their species related information.

Keeping in view the fundamental importance of fish species recognition, since last two decades many techniques have been developed to identify the species of fish in an image. Based on the scope the techniques can be divided into three main categories:

- Identifying the specie of dead fish under controlled environment (e.g Conveyor belt).

- Identifying the specie of live (swimming) fish under controlled underwater environment (e.g aquafarming relocation).

- Identifying the specie of live fish swimming in the unconstrained underwater environments (e.g living in their natural habitats).

2.3.1 Species Identification of Dead Fish

Based on the first category of identifying specie of dead fish was applicable for commercial and research use as sorting out fish in different vessels according to their use. Earlier work done in this direction primarily focused on shape descriptor features of fish (Strachan et al., 1990), (Strachan and Kell, 1995). The color features were also examined for the identification purposes (Strachan, 1993). Another approach used was to use laser light and a camera to create 3-D model of fish and extract features based on the height, width and thickness to identify the species of fish (Storbeck and Daan, 2001). Comparative high accuracy was also obtained recently by the use of conveyor belt systems where the illumination was controlled and fishes were passed on conveyor belts and sorted out based on their shape discriminant features (White et al., 2006). Apart from shape-based and color-based features for fish species identification, texture-based features are also being explored (Larsen et al., 2009). All these techniques provided a high ratio of accuracy as they were applied in controlled environment.

2.3.2 Species Identification of Live Fish under Controlled Environment

Aquafarming allows farming of different species of fish and other aquatic beings in ponds under controlled environments which can be contrasted with harvesting of wild fishes. So the identification in such environments is of high importance and (Ruff et al., 1995) and (Harvey and Shortis, 1995) initiated the technique of deploying stereo camera systems in such environments. Manual processing of these videos for species identification and length measurements were being done in laboratory for which researchers developed some automated methods (Zion et al., 1999), (Zion et al., 2000). Later for improving the accuracy and performance of these systems researchers developed a computer vision system in which fishes were to pass a narrow channel with cameras installed in the side taking the profile view of fishes. With this system the researchers were able to get high accuracy in real time (Zion et al., 2007). After this technique another method was developed for fish recognition and size measurements in dams where fishes were to swim through a narrow fish passage, fish ladder in dams (Lee et al., 2008). The color features were used to extract fish from images as controlled illumination setup was used. Then they performed contour matching to identify the species of fishes.

2.3.3 Species Identification of Live Fish in Unconstrained Underwater Environment

All of the techniques discussed above produced good results as they were performed under controlled environments whereas real life underwater environment is much more complex and variable. Therefore fish species classification in unconstrained underwater environments is more challenging. The factors involved in increasing the complexity of underwater environments are:

- Lighting variation
- Murkiness of water
- Background distraction due to underwater plants and coral reefs
- Orientation variation of freely swimming fishes

All the automated fish species classification systems assume that the fish is already available and detected in the image. For fish detection many techniques are present (Shortis et al., 2013), (Walther et al., 2004), (Spampinato et al., 2008), (Nadarajan et al., 2011). Underwater videos are usually captured through digital cameras and when deploying the cameras the underwater environment is highly uncertain. So the major problem is caused by freely swimming fish directions which are quite difficult to capture and then identified. Considering these issues the research in unconstrained underwater environment has only been addressed recently as the technology is rapidly improving.

2.3.4 Feature-based Approaches

Based on the texture-feature initial efforts for fish species identification in unconstrained underwater environment were made (Rova et al., 2007). They classified fishes but the approach was limited to fish exhibiting rich texture. Another approach involving shape information along with texture features was proposed by (Spampinato et al., 2010). The problem in this approach was that the highly variable background creates confusion and segmenting out fish from the image was a great challenge. The other issue involved the orientation of free swimming fish gives much larger shape variation than the fish moving in profile view (owing to the camera direction).

2.3.5 Machine Learning-based Approaches

Machine learning is a branch of computer science that deals with self learning when exposed to new data. Which means the algorithms designed for machine learning techniques learn the labels from trained data and differentiate classes based on prediction from learned features. So for fish classification machine learning is now being adopted to get more accurate results.

2.3.5.1 Principal Component Analysis (PCA)

Initial machine learning algorithms for fish species classification involved Principal Component Analysis (PCA) (Turk and Pentland, 1991) or Linear Discriminant Analysis (LDA) (Scholkopft and Mullert, 1999).

Principal component analysis (PCA) is a technique which is used to extract strongly related patterns in the dataset and highlight the variations in them. When large multivariate datasets are to be analyzed, it is often required to reduce their dimensions. PCA derives q variables, principal components, which are smaller in number than p original variables and these q variables are linear combinations of p, original variables. Most of the time these q variables are able to depict and retain the variability in p variables.

These techniques work in linear manner. They assume that the species are linearly independent in terms of appearance that is fishes belonging to one species will have no similarity with fishes belonging to other species and same is assumed for background. Whereas in reality these suppositions do not hold as the fishes resemble to each other in both size and shape and also because highly contrasted backgrounds create vagueness.

2.3.5.2 Sparse Representation based Classification (SRC)

Lately for fish species identification in Taiwan coral reef ecosystems Sparse Representation based Classification (SRC) combined with PCA has been used (Hsiao et al., 2014).

The sparse representation was first developed in signal processing community which used it for the reconstruction of a sparse signal by using its structure of sparsity. In terms of object recognition it is known that samples collected from a single class can be found as a linear subspace. Hence the test sample belonging to a class can be depicted as a linear combination of training samples from the same class as that of test sample. Meantime it can also be said that the linear representation of a test sample offered by training samples from genuine class of this test sample is more concrete than by any other training samples in the set. This is the discerning nature of SRC method.

These machine learning algorithms due to their linear nature are not able to cater nonlinear differences among the backgrounds as well as the fish species. Complex and variable backgrounds in unconstrained underwater environments can be seen in the

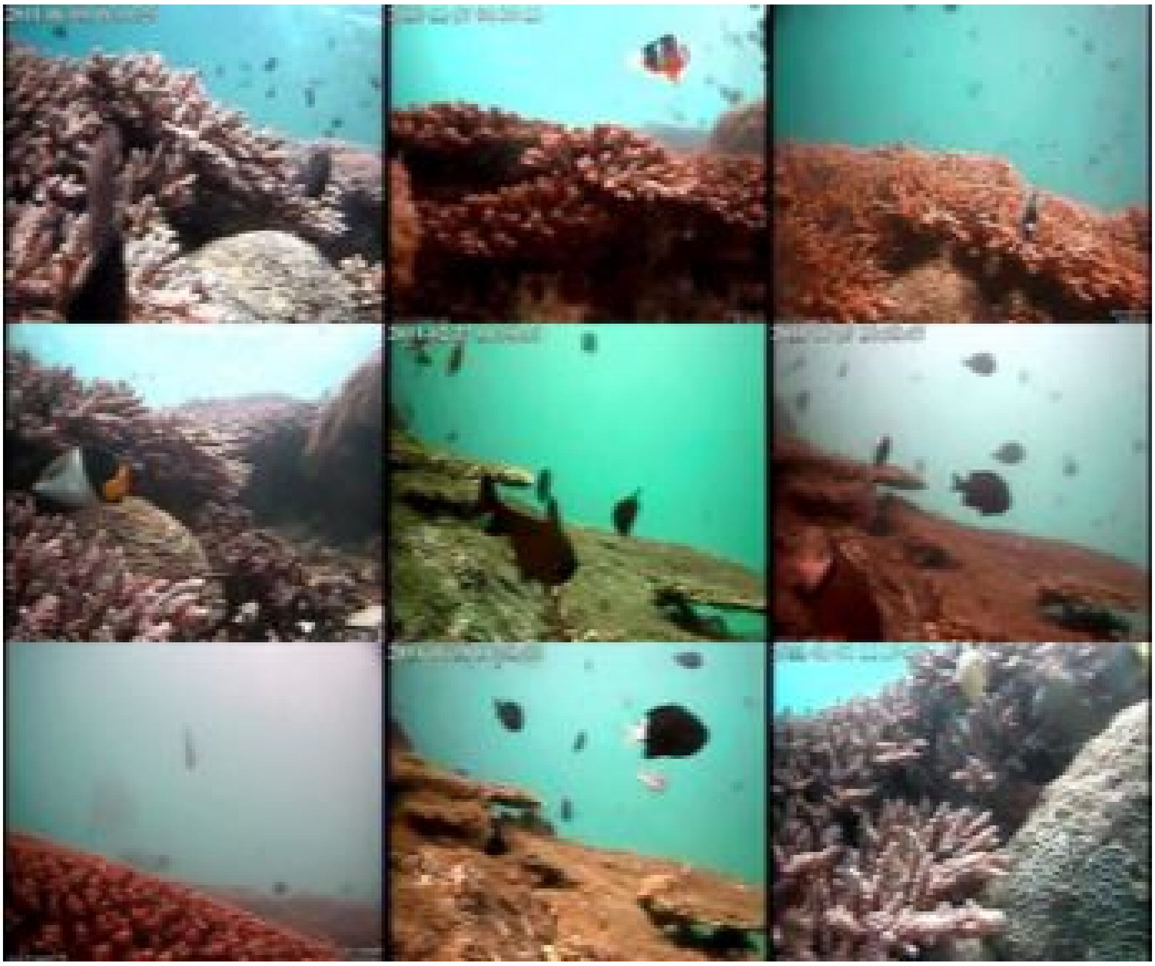

Figure 2.1: Highly complex and variable background in underwater environment of Taiwan Reef

Figure 2.1 .

2.3.5.3 Support Vector Machines (SVM)

Another machine learning technique for fish species classification in underwater environment used hierarchical trees with Support Vector Machines (SVMs) (Huang et al., 2015).This technique was developed to reduce dependency on shape features and used ML for developing feature descriptor trained on input image features. This technique was used for Live fish dataset (see Figure 2.2). The results produced were

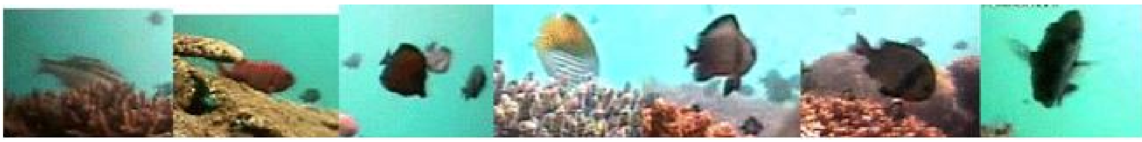

Figure 2.2: Hard fish examples in Live fish dataset.

much more improved than PCA with standard SVM classification.

SVM is a supervised learning technique used for classification. It was initially developed to solve binary linear classification in which the sample was put under one of the two main categories or class labels after classification. But this method can also be used effectively for multi-class problems by using the kernels, which maps the inputs to higher dimensional feature space.

2.3.6 Deep Learning-based Approaches

Deep learning has emerged as new era of machine learning and is providing extraordinary capability to improve image classification eliminating the shortcomings in conventional machine learning algorithms. Image data is not always clean it has a series of issues involved like variations in lighting, intrusion in image due to noise and poor quality, variety of backgrounds and changes in orientation of object of interest (Bengio, 2009). In such images it is difficult for conventional machine learning algorithms to get trained successfully and give accurate results. Underwater images of fish in unconstrained environment includes all of the above discussed challenges. There must a mathematical function which is non-linear and specifically designed to address the complex features in image data. Deep neural networks with multiple layer architecture can provide such features which in the presence of such variability and distortions can extract rare and vigorous fish dependent features.

Recently deep Convolutional Neural Networks (CNNs) alongwith a classifier have been proposed to extract invariant and robust features from fish image datasets for the task of fish species identification (Salman et al., 2016). They trained and tested

Figure 2.3: Example of various fish images showing one specie per row in LCF-14 (above) and LCF-15 (below) datasets .

their proposed network on benchmark datasets LifeCLEF'14 and LifeCLEF'15. Images from datasets used can be seen in Figure 2.3. With same and cross dataset training/testing they have shown that machine learning algorithms provided better performance when training set has more distorted and noisy images and test set was comprised of clean images as compared when both training and test set consisted of clean images. They used CNN to train their network and used last layer features as input for classification and proved that employing this technique for the task of fish species identification is best suited algorithm for the said dataset.

Chapter 3

Proposed Approach

The proposed approach is inspired by the deep learning methods used for fish species identification in unconstrained underwater environment. It mainly comprises of three parts which can be seen in the block diagram (see Figure 3.1). The 3 main parts are:

1. Pre-processing
2. Main Algorithm
3. Classifier training/testing

Pre-processing deals with image processing tasks that make data more valuable and classifier training/testing part is used to make decisions and classify specie of fish in the image. The main algorithm as can be seen from Figure 3.1 comprises of 2 further parts: First is the extraction of local features using predefined AlexNet with dimensionality reduction and then Cross-Convolutional layer pooling on those extracted feature to generate image representation. AlexNet and Cross-Convolutional layer pooling are described in the sub-sections below.

3.1 AlexNet

In 2012, AlexNet was proposed during ILSVRC2012. The model won the award for generating higher accuracy as compared to non-DNN method and producing comparatively better results. This model has got 5 convolutional layers followed by 3 fully connected layers. Some kind of randomness is introduced during data augmentation

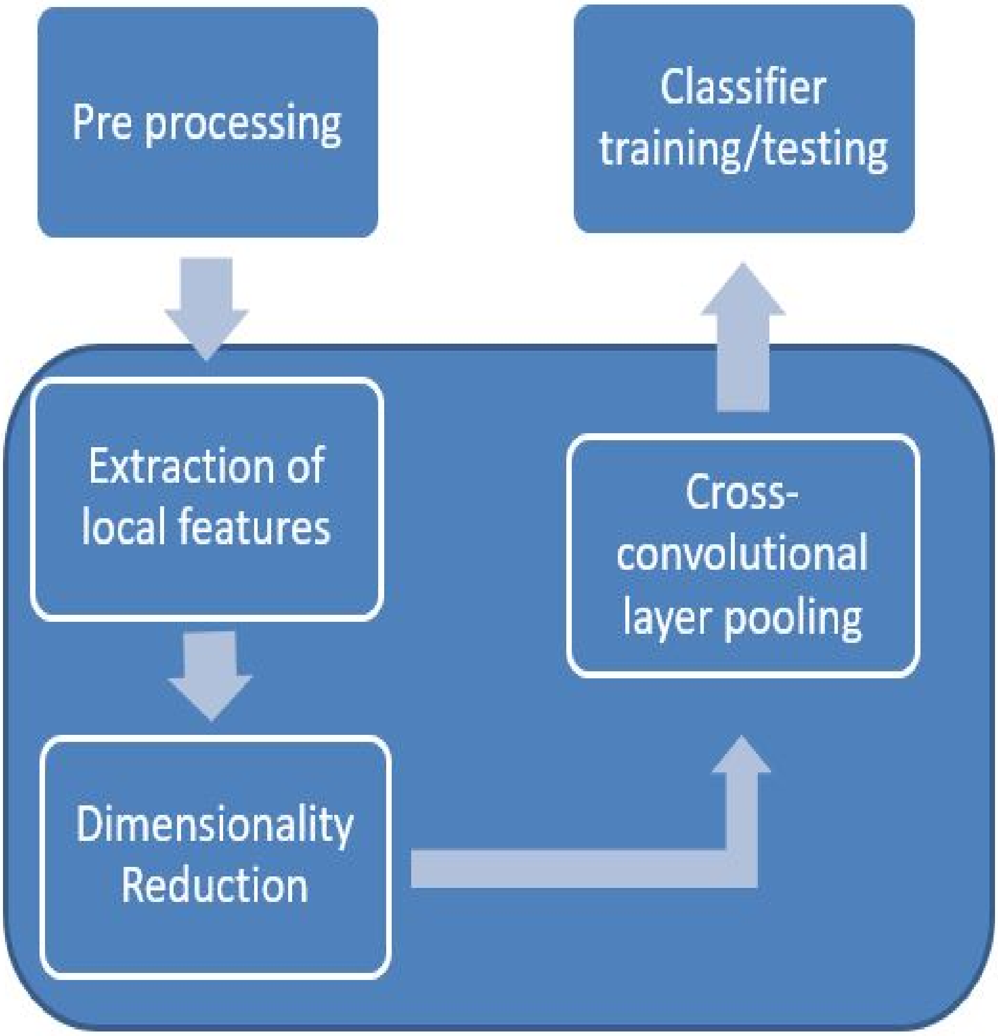

Figure 3.1: Block diagram of proposed approach.

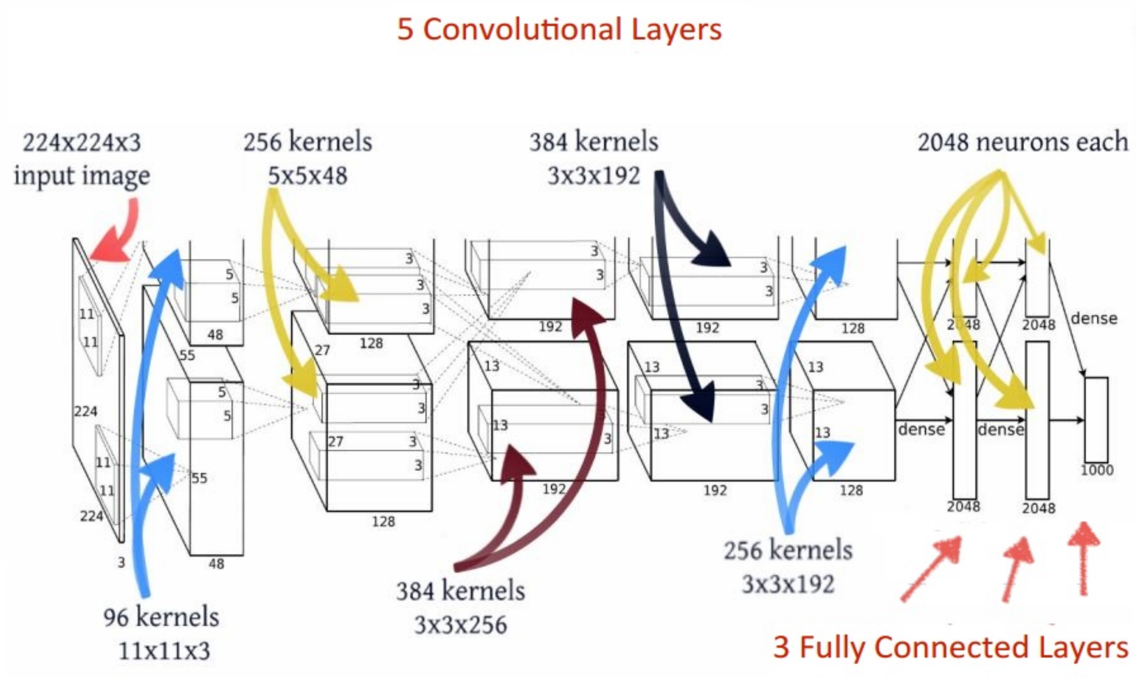

Figure 3.2: Architecture of AlexNet

process and dropout layer in phase of training. The respected architecture details can be seen in the Figure 3.2. These parameters need to be defined as provided by Caffe in configuration profile. Data path and mean profile path need to be defined as required and according to the placement of these files. The current support is for one group only. For current requirement with one GPU the RAM is enough to hold data (Xiao et al., 2014).

Multi-view classification is supported in Alexnet. Image is cropped into below 5 patches:

1. left-upper
2. left-down
3. right-upper

4. right-down

5. horizontal flipped version

It can be used as feature extractor for other computer vision application once the training is completed. Features are extracted and then stored as matrix (feature vector) to be used for further processing or classification. The size of matrix depends on number of images and number of features extracted. In our approach we have extracted features of 4th and 5th layers to create our final feature vector.

3.2 Cross-Convolutional Layer Pooling

Before understanding the concept of cross-convolutional layer pooling one must have the basic knowledge of convolutional neural networks as discussed below.

3.2.1 Convolutional Neural Network

This is an area associated with machine learning that learn in a multi-level artificial neural networks with the help of some strong analytical techniques. These techniques are used for extracting different features that are helpful in training phase. Hidden features are extracted and this is so far considered to be the best available technique for features extraction, Since feature extraction is considered to be the main and important module in field of machine learning, till now it is done through deep learning. Same is the case for us where we need to extract features from available images. While dealing with images convolutional neural network (CNN) is the kind of network that we used for processing. This architecture is designed specifically keeping in mind for images. Simple neural network contains an input and an output layer. These layers are connected using multiple hidden layers. All the perceptrons are connected via neuron that has got some weight. A simple neural network structure is shown in Figure 3.3. Considering same architecture for images, for 200x200 pixel image, initial

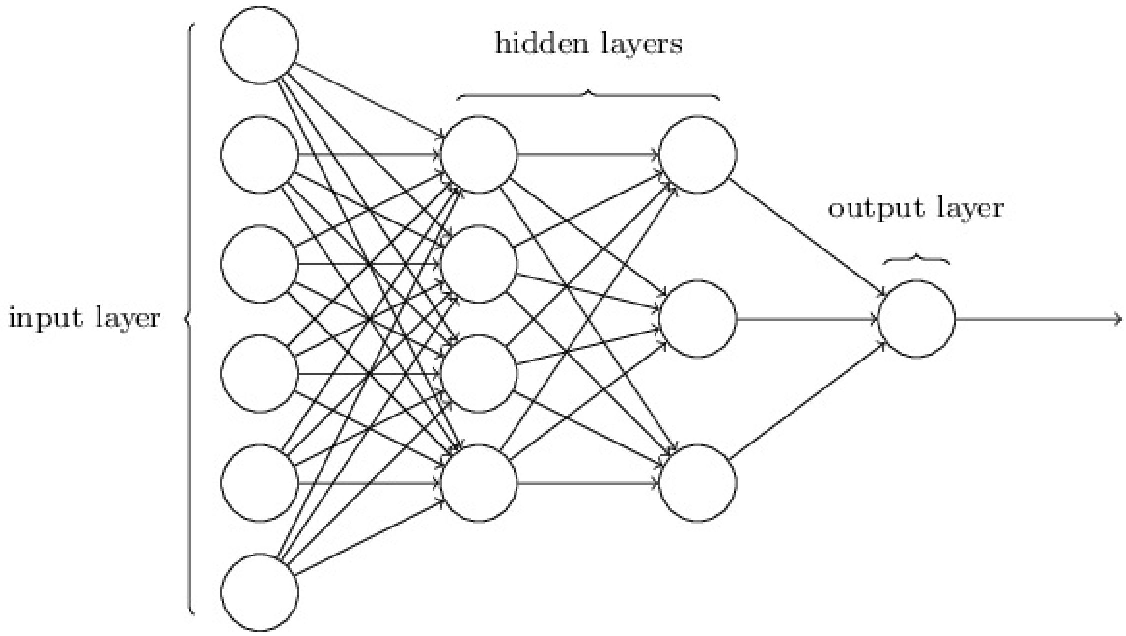

Figure 3.3: Fully connected Neural Network

layer will be having 200x200x3 = 120,000 number of weights for the connecting neurons of first layer. This number is too huge and computationally expensive as well. This is directly proportional to number of images and as the number will grow the amount will increase as well which will make it tough to manage. Now to address this problem CNN architecture has been designed keeping in mind and all the layers are not fully connected. The size decreases as we move forward and number of images increase as shown in Figure 3.4.

A CNN consists of following layers:

- An input layer comprising of the pixels from input image with same dimensions as of the image.
- Convolutional layer work like series of filters. These filters can then be applied to connect the local regions of the input. Dot product of local regions and weights is computed for the network.
- Down sampling is performed in Pooling layer that work on spatially connected

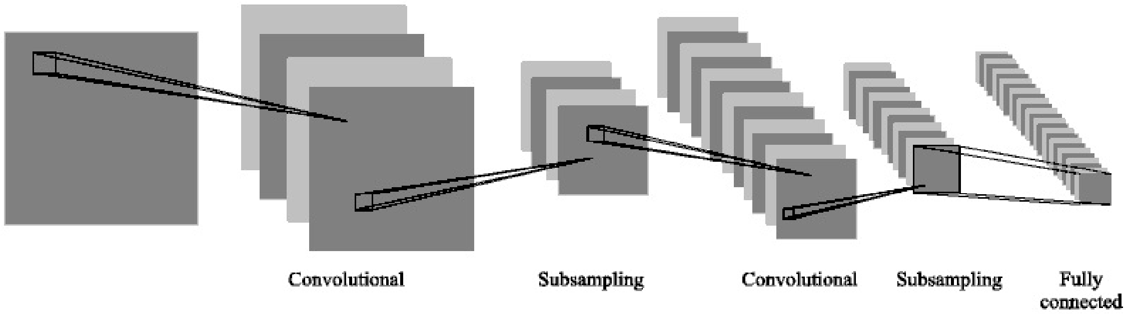

Figure 3.4: Architecture overview of Convolutional Neural Network

local regions from the input layer. Max pooling is considered to be the common type of pooling used.

- In fully connected layer,all input from previous layer are connected to all inputs of next layer. In CNN, last layer is a fully connected layer. Class scores are computed in last layer. Number of classes determine the number of units in this layer.

ImageNet is used for training the model that tries to retain the knowledge gained on large datasets and then used it for further processing. This is also known as transfer learning technique. A pre-trained CNN model named Imagenet Caffe-Alex was used which is publicly available for learning purposes.

3.2.2 Why Convolutional Layers?

As we now know that convolutional neural network is designed specially for image data and that all of the layers have local information stored in them. Most of the studies and researches being done recently are using the activations of fully connected layers of CNNs as feature vectors for classification task and image descriptors as it is believed that convolutional layers are less discerning. But it is now being discussed that the convolutional layers, if used properly, can give us a powerful representation of images (Liu et al., 2015).

One big contrast in using convolutional layers rather than fully connected layers is that the latter although saturated with semantic information lacks spatial information. The convolutional layer activations can be represented as 2-D array of D-dimensional local features, each of the feature describes a local region. The fully connected layer then use these convolutional layer activations and changes them to a feature vector representing the whole image. While going through this process the spatial information concentrated in convolutional layers is lost and it can not be recovered from the feature vectors of a fully connected layer.

To extract multiple regional descriptors we can use subset of activations located in convolutional layers which represent a local region. We can use these subsets of activations which coincides to a set of subarrays of convolutional layer activations and are considered as local features.Figure 3.5 exhibit the extraction of local features. We can extract 121 (9 x D)-dimensional local features by scanning 13 x 13 feature maps by using this method. As we opt to use activations of two consecutive convolutional layers since we believe that feature vector extracted from a single layer might not be eloquent enough to describe a visual pattern in a local region.

3.2.3 Cross-Convolutional-Layer Pooling

To get image-level representation one can perform simple max pooling or sum pooling on the extracted features. We on the other hand use a technique of cross convolutional layer pooling that out performs the other techniques (Liu et al., 2015). This technique was inspired from parts-based pooling method in which multiple regions-of-interest (ROIs) are first computed and then a pooled feature vector is obtained by pooling the local features that lie into each ROI. In cross-convolutional layer pooling, the feature maps of (k+1)th convolutional layer are used as indicator maps for creating pooling channels of k-th convolutional layer.

In this technique we first obtain feature vectors by extraction of the locally spatial

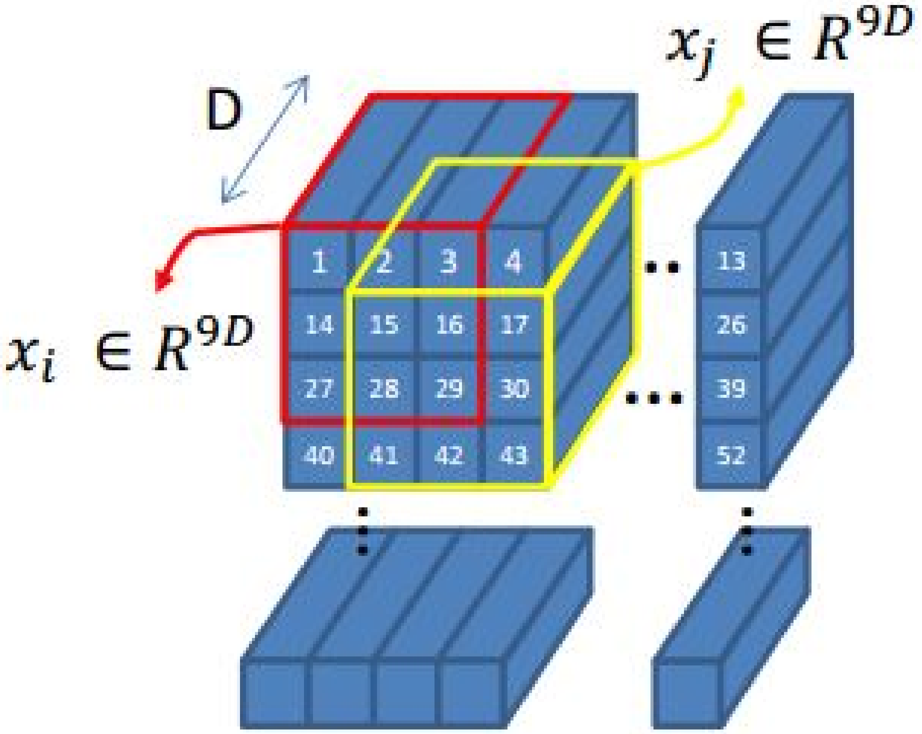

Figure 3.5: Extraction of local features from a convolutional layer.

units of a convolutional layer. Then by using cross layer pooling method the feature vectors are pooled where the feature vectors of k-th convolutional layer are multiplied with the corresponding activations from (k+1)-th convolutional layer and then added together to make a single pooled vector. This process continues for all pooling channels and then one large feature vector is obtained by concatenating all the pooled vectors. Then the feature vector gained are used to train a classifier like SVM.

In our approach we extracted features from 4th and 5th convolutuional layer and then cross pooled them to get feature vector for classifier.

3.3 Proposed Algorithm

In the proposed algorithm we first pre-processed all the images in dataset. Since all the images in dataset were of very low contrast (will be discussed in next chapter) we performed image processing technique to make the object of interest visible in the image. We simply converted the images from RGB to YCbCr domain and only Y-component is kept on which contrast stretching is applied. In YCbCr image Y is the luminance component and CbCr is the chrominance component of image. Only the Y-component is kept as most of the information is concentrated in it where the red-blue difference of chrominance component only contained color information. Moreover, we added Gaussian noise with 0 mean 0.01 variance to the training images to double the size of training data. The result of process can be seen in Figure 3.6.

After pre-processing the main part of algorithm is to prepare the dataset for training. A deep learning classification technique is being used for this task. The main algorithm further comprises of two parts: First is the extraction of local features and then cross pooling of features to get a feature vector for training.

For the extraction of local features we aimed at using a pre-trained network. For this purpose the preprocessed dataset was to be fed to a deep convolutional neural network and trained. We used Caffe (Yangqing Jia, 2014), which is a C++ library for

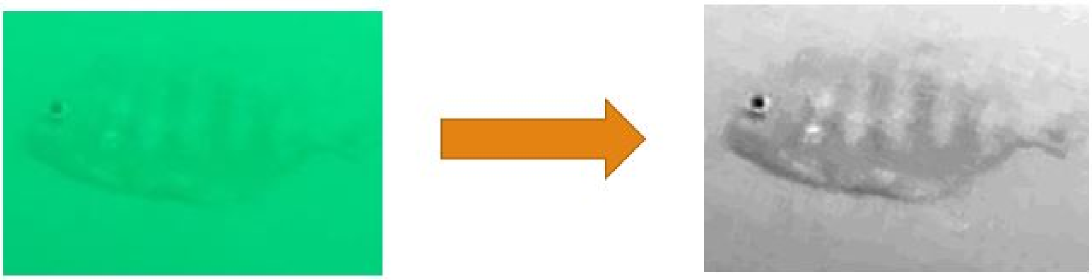

Figure 3.6: Result of contrast stretching applied to Y-component of YCbCr converted image

deep learning, for implementation and training of all the networks. Several networks which were tried are as follows:

1. A pretrained (on ImageNet dataset) GoogLenet (Szegedy et al., 2015) is fine-tuned on RGB images without any pre-processing from the UWA dataset.

2. An AlexNet (Krizhevsky et al., 2012) is trained on the RGB images without any preprocessing.

3. An AlexNet trained on Y-component of training images doubled in number by adding noise.

4. An AlexNet trained on Y-component of training images which was doubled with noise and stretched contrast.

The accuracy results of all the tried approaches can be seen in the graph in .Figure 3.7. It is proved by the experiments that AlexNet with training images doubled and stretched contrast provided highest accuracy so it is best suited for the extraction of local features for cross pooling.

So AlexNet was used to extract CNN activations from 4th and 5th convolutional layers. The extracted features were high in dimension so we opted to down sample the

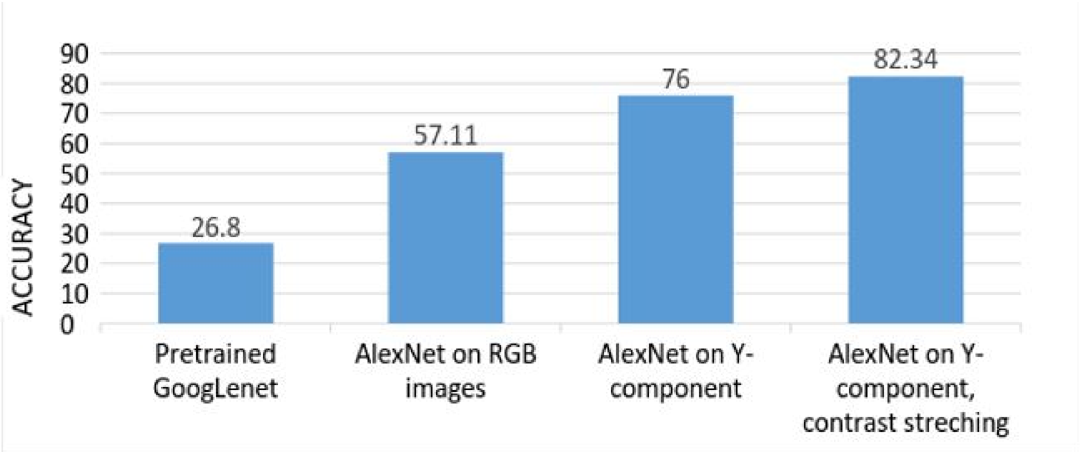

Figure 3.7: Training results of different networks on UWA dataset

extracted features by 500 using PCA (Principal Component Analysis) before cross-layer pooling of the extracted activations. Reducing the dimensionality using PCA increases the performance as well as it is faster and simple. After extraction and reduction of features a final feature vector was produced by cross-convolutional layer pooling method (discussed above).

In the last step of algorithm the feature vector was then fed to a classifier, we used SVM (Support Vector Machine) with pre-computed linear kernels. The reason of using pre-computed linear kernels is that its calculations can be implemented in parallel. So when the dimensionality of data is high most of the computational time is taken by kernel matrix calculation. Thus in order to save computational time and accelerate the process, it is appropriate to use parallel computing.

Chapter 4

Implementation and Results

In this chapter we describe the details of the dataset used in the research. After that we present the results of several experiments done during the research.

4.1 Dataset

The dataset used for classification was taken from the University of Western Australia. The complete dataset consists of 4412 images from which 3288 images belong to 16 species of interest and the rest of 1124 images belong to other's class which are of no interest. These images have directly been cropped from underwater video sequences and are of extremely low contrast and also saturated with intrusion either from background or from camera rig itself. Moreover, when thinking of several features that can be used for classification, color is a very bad choice because of sunlight or any other factor the color of a fish from one specie differs from the color of another fish of the same specie. So color is not a good feature which can be used for classification in this dataset. Similarly the images are of very low contrast and this is why texture-based classification is also not possible. Figure 4.1 gives an idea of images present in the dataset. So we converted images from RGB to YCbCr domain and stretched the contrast using Y-component as discussed in the previous chapter.

Out of all 4412 images in the dataset, 70% were used for training and rest of the 30% of the images were used for testing. Table 4.1 gives a categorical view of UWA Dataset according to number of sample images in each specie.

Figure 4.1: Sample images of various fish species (one per row) in UWA dataset.

UWA Species	**No. of Images**
Abudefduf_bengalensis	201
Carangoides_fulvoguttatus	217
Choerodon_cyanodus	221
Choerodon_rubescens	201
Coris_auricularis	283
Lethrinus_atkinsoni	211
Lethrinus_nebulosus	211
Lethrinus_sp	197
Lutjanus_carponotatus	207
Pagrus_auratus	203
Pentapodus_emeryii	185
Pentapodus_porosus	207
Plectropomus_leopardus	195
Scarus_ghobban	189
Scombridae_spp	179
Thalassoma_lunare	203
Other	1126

Table 4.1: Population division of different species in UWA dataset.

Algorithms	**Testing Error**
SRC	50.94%
SVM	75.53%
PCA with KNN (100 components)	90%
PCA with KNN (1000 components)	89.53%
CNN	49.88%

Table 4.2: Resullts of applying conventional Machine Learning algorithms on UWA dataset.

4.2 Experimental Results

The proposed approach consists of 3 main steps as shown in Figure 3.1 and thus several experiments were done at each stage to get better results. In the beginnig before implementing the main algorithm, preprocessed images dataset was trained and tested on several machine learning algorithms for comparison with proposed approach. Table 4.2 provides the rate of accuracy obtained. We can see that all the algorithms provided us with very high error rate at testing time which clearly shows that conventional methods cannot be directly applied on UWA dataset and some other advance method is required for better performance.

4.2.1 SRC Results

In Sparse Representation based Classification, a dictionary learning technique is used from which sparse linear combination of main functions are found out to make subspace. To find the structure of this subspace enough training is required and this subspace representation is effective for image classification and recognition. When there is little training data available the formation of subspace becomes difficult and this is the main problem that this machine learning algorithm could not perform well on our UWA datasest.

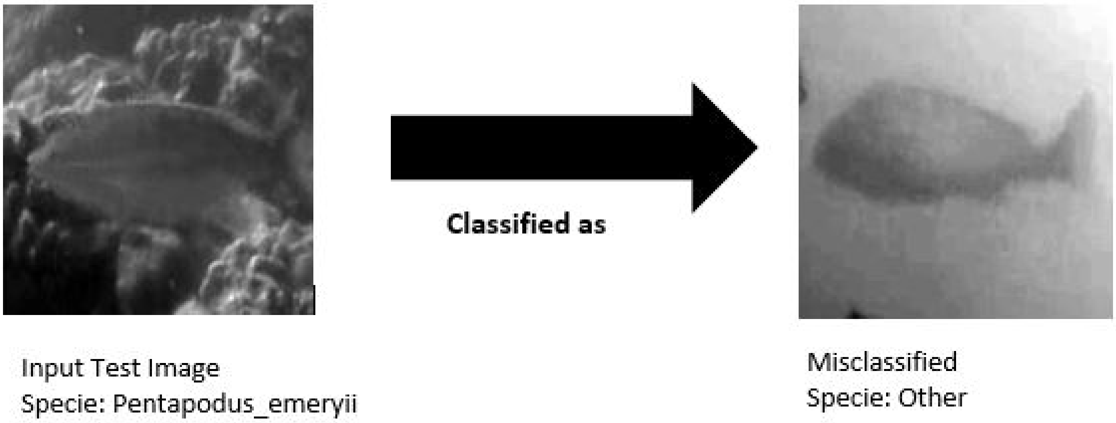

Figure 4.2: Example of wrong classification in SRC

In Figure 4.2, we can see that the neighboring subsample created in SRC caused the mis-classification. The main reason is lack of training data in the provided dataset.

4.2.2 SVM Results

The biggest limitation of using support vector machines lies in the wrong choice of kernel and regularisation parameters. We can say that parameters for model selection plays a great role in classification process and wrong selection of kernel models can cause over-fitting.

In our dataset SVM produced very low accuracy which is below 25% (as in Table‘4.2) and the result of mis-classification as shown in Figure 4.3.

4.2.3 PCA with KNN Results

Principal components Analysis extract the strongly related components or features from the training dataset and by using lesser number of variables allow classifier to optimally find the solution. We used PCA with KNN (K-Nearest Neighbors) for classification and got very high error rate with varying size of components (can be seen in Table 4.2). Big disadvantage of PCA is its linear learning behavior and with

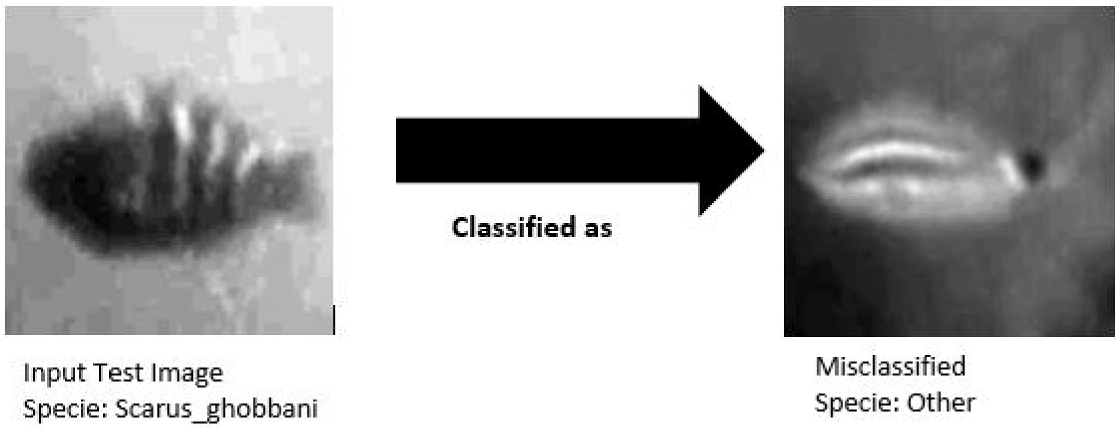

Figure 4.3: Example of wrong classification in SVM

KNN it could not perform any better because of the variability in dataset. The result of mis-classification can be seen in Figure 4.4.

4.2.4 CNN Results

We trained our UWA dataset using Convolutional Neural Network having 3 convolutional layers with 200 number of epochs and 34 as batch-size. We got a low training error of 17.9412% which then lead to 49.8822% testing error, which clarifies the fact that our dataset is highly variable and complex in terms of lightining and texture features. CNN was unable to extract the features able enough to correctly classify the fish species, this result motivated us to use extracted local features as feature vector for training instead of using fully connected layer resulting feature vector as applied in the tested approach.

4.2.5 Robustness of trained AlexNet

We trained different CNNs with different set of images to see which of them afford to have highest accuracy and will be suitable for our dataset (Refer to Figure 3.7). We performed another experiment to check the robustness of our trained and selected

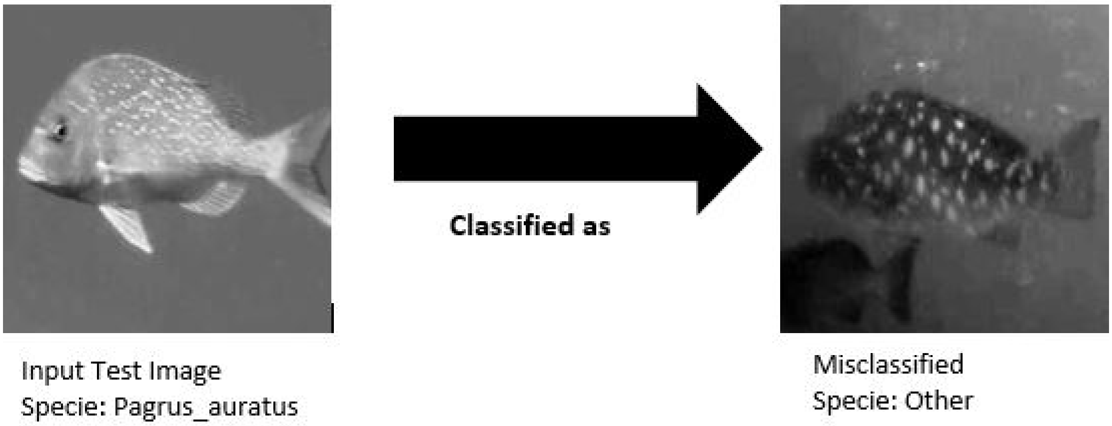

Figure 4.4: Example of wrong classification in PCA with KNN (100 components)

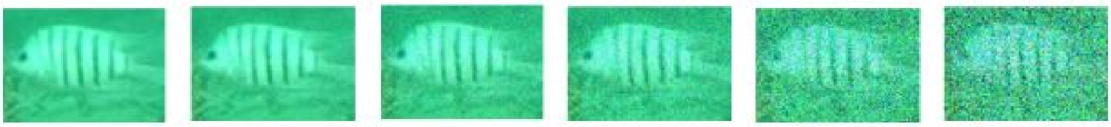

Figure 4.5: Sample images with different level of induced noise.

CNN which was ALexNet with Y-component images stretched and doubled with noise. We tested the trained AlexNet with images from UWA dataset with different level of noise. Different level of noise added to images can be seen in Figure 4.5. We can see the results in graph in Figure 4.6 which shows that our network performs very well and classified accurately to 82% with images having induced noise of variance up to 0.02.

4.2.6 AlexNet Caffe Implementation

In our proposed approach, as previously mentioned, we used pre-trained AlexNet for extraction of CNN activations from 4th and 5th convolutional layer of the network. We set the resolution size of 4th and 5th layer to 13 x 13 spatial units, it is the

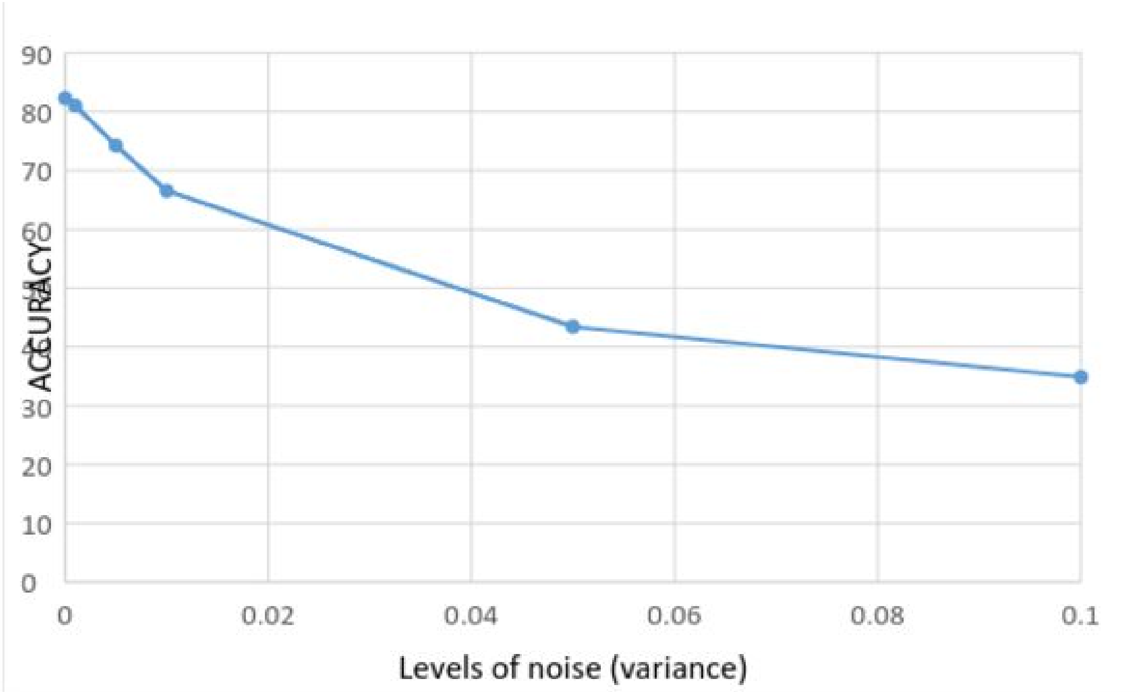

Figure 4.6: Classification accuracy on levels of induced noise.

default parameter size in Caffe implementation. We extracted the features and without applying PCA for dimensionality reduction we trained a classifier say LibSVM with pre-computed linear kernels and used the model for validation and testing. The confusion matrix showing accuracy results for training, validation and testing can be seen in below Tables 4.3, 4.4, 4.5.

As it can be seen from confusion matrix in Table 4.3 that training yields 100% accuracy as 1309 images out of 1309 images set for training were classified accurately. We then used a validation dataset to validate the training results and we obtained 62.0253% accuracy as 147 images out of 237 were correctly classified as shown by confusion matrix in Table 4.4. When the model was tested on test dataset with same number of species provided us with 65.7617% accuracy. There were total of 663 among which 436 items were correctly classified while other being distributed among rest of the specie.

One primary observation that we can make from confusion matrices is that most of the items got mis-classified as 'Other' specie. This is because the data items under 'other' class label are large in number and of diverse type. The model trained with such variety under one label made the classification vague and imprecise.

Another reason for less accuracy is that our model reported 100% accuracy which means that it was over-fitted due to small number of training set and large number of parameters while classifier training. More training data is required for optimal training of model which can result in better test accuracy.

4.2.7 Cross-Convolutional Layer Pooling Results

The second sub-step in main algorithm is to cross-pooled the activations of 4th and 5th convolutional layers in order to get final feature vector for classification of fish species. For our implementation as stated previously we used pretrained AlexNet to extract features (i.e, CNN activations) from convolutional lyers.

Firstly the activations from 4th convolutional-layer were extracted and it provided us with 169 feature vectors, each having 3456 features, for every image. Then these feactors were to be pooled using activations from 5th convolutional layer which was of depth 256. This resulted in a total of 3456*256=884736 features for every image. Since this resulted in an extremely large feature vector to be trained on so we used PcA for reducing the dimensionality of our resulting feature vector.

We applied PCA on the local feature vectors extracted from 4th convolutional layer and then pooled the resulted reduced feature vectors according to the activations from 5th convolutional layer. As it can be seen from the latency plot in Figure 4.7 that 94% energy (information) was concentrated in the first 500 principal components and thus we used them for transforming our feature vectors. We reduced the dimensionality by using 500 features instead of 3456 and then these 500 features pooled with 256 features from 5th layer gave us feature vector having 500*256=128000 features.

Lastly for the classification SVM was trained on these feature vectors with linear

	Pentapodus porosus	**Pentapodus emeryii**	**Choerodon rubescens**	**Abudefduf bengalensis**	**Choerodon cyanodus**	**Plectropomus leopardus**	**Lethrinus nebulosus**	**Carangoides fulvoguttatus**	**Lutjanus carponotatus**	**Scarus ghobban**	**Scombridae spp**	**Lethrinus sp**	**Thalassoma lunare**	**Coris auricularis**	**Other**	**Pagrus auratus**	**Lethrinus atkinsoni**
Pentapodusporosus	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pentapodusemeryii	0	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Choerodonrubescens	0	0	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Abudefdufbengalensis	0	0	0	56	0	0	0	0	0	0	0	0	0	0	0	0	0
Choerodoncyanodus	0	0	0	0	59	0	0	0	0	0	0	0	0	0	0	0	0
Plectropomusleopardus	0	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0
Lethrinusnebulosus	0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0
Carangoidesfulvoguttatus	0	0	0	0	0	0	0	64	0	0	0	0	0	0	0	0	0
Lutjanuscarponotatus	0	0	0	0	0	0	0	0	63	0	0	0	0	0	0	0	0
Scarusghobban	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0
Scombridaespp	0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0
Lethrinussp	0	0	0	0	0	0	0	0	0	0	0	46	0	0	0	0	0
Thalassomalunare	0	0	0	0	0	0	0	0	0	0	0	0	58	0	0	0	0
Corisauricularis	0	0	0	0	0	0	0	0	0	0	0	0	0	91	0	0	0
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	421	0	0
Pagrusauratus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	52	0
Lethrinusatkinsoni	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56

Table 4.3: Train data confusion matrix

	Pentapodus porosus	Pentapodus emeryii	Choerodon rubescens	Abudefduf bengalensis	Choerodon cyanodus	Plectropomus leopardus	Lethrinus nebulosus	Carangoides fulvoguttatus	Lutjanus carponotatus	Scarus ghobban	Scombridae spp	Lethrinus sp	Thalassoma lunare	Coris auricularis	Other	Pagrus auratus	Lethrinus atkinsoni
Pentapodusporosus	9	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0
Pentapodusemeryii	1	10	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
Choerodonrubescens	0	0	9	0	0	1	0	0	0	1	0	0	0	0	11	0	0
Abudefdufbengalensis	0	0	0	11	0	0	0	0	0	0	0	0	1	0	0	0	0
Choerodoncyanodus	0	0	0	0	7	0	0	0	0	1	0	0	0	0	3	0	0
Plectropomusleopardus	0	0	1	0	0	12	0	0	0	0	0	0	1	0	8	0	0
Lethrinusnebulosus	0	1	0	0	0	0	7	0	0	0	0	0	0	0	3	0	0
Carangoidesfulvoguttatus	0	0	0	0	0	0	0	4	0	0	0	0	0	0	6	0	0
Lutjanuscarponotatus	0	0	0	0	0	0	0	0	7	0	0	0	0	0	3	0	0
Scarusghobban	0	0	0	0	0	0	0	1	0	1	0	1	0	0	10	0	0
Scombridaespp	0	0	0	0	0	0	0	0	0	0	2	0	0	0	5	0	0
Lethrinussp	0	0	0	0	0	0	0	0	0	0	0	9	0	0	4	1	1
Thalassomalunare	0	0	1	0	0	0	0	0	0	0	0	0	8	1	2	0	0
Corisauricularis	0	0	1	0	0	0	0	0	0	0	0	0	0	10	0	0	0
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0
Pagrusauratus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0
Lethrinusatkinsoni	0	0	0	0	0	1	4	0	0	0	0	0	0	0	3	0	8

Table 4.4: Validation data confusion matrix

	Pentapodus porosus	Pentapodus emeryii	Choerodon rubescens	Abudefduf bengalensis	Choerodon cyanodus	Plectropomus leopardus	Lethrinus nebulosus	Carangoides fulvoguttatus	Lutjanus carponotatus	Scarus ghobban	Scombridae spp	Lethrinus sp	Thalassoma lunare	Coris auricularis	Other	Pagrus auratus	Lethrinus atkinsoni
Pentapodusporosus	21	0	0	0	0	0	0	0	0	0	0	0	0	2	8	0	0
Pentapodusemeryii	2	21	0	0	0	0	0	1	0	1	0	0	0	1	9	0	0
Choerodonrubescens	0	0	20	0	0	1	0	0	0	0	0	1	0	0	9	0	2
Abudefdufbengalensis	0	0	0	28	0	0	0	0	0	0	0	0	0	0	4	0	0
Choerodoncyanodus	0	0	0	0	28	0	0	1	0	4	0	0	0	0	5	1	1
Plectropomusleopardus	0	0	0	0	0	15	1	0	0	1	0	0	1	1	13	0	0
Lethrinusnebulosus	0	0	0	0	0	1	22	2	0	0	0	1	0	0	16	0	1
Carangoidesfulvoguttatus	0	0	0	0	0	0	0	22	0	1	0	1	0	0	10	0	0
Lutjanuscarponotatus	0	0	0	0	1	0	0	0	21	2	0	0	0	0	6	0	0
Scarusghobban	0	0	0	0	1	0	0	3	0	19	0	0	0	0	8	0	0
Scombridaespp	0	0	0	0	0	0	0	8	0	0	13	0	0	0	9	0	0
Lethrinussp	0	0	0	0	1	0	2	2	0	0	0	12	0	0	15	1	4
Thalassomalunare	0	0	0	0	0	0	0	0	0	1	0	0	27	0	3	0	0
Corisauricularis	0	0	1	0	0	0	0	0	0	0	0	0	0	30	7	0	0
Other	0	1	1	0	1	4	2	3	0	1	8	0	1	0	92	0	1
Pagrusauratus	0	0	1	0	1	0	0	0	0	0	0	0	0	0	8	28	0
Lethrinusatkinsoni	0	0	0	0	1	0	3	0	0	1	0	0	0	0	11	0	17

Table 4.5: Test data confusion matrix

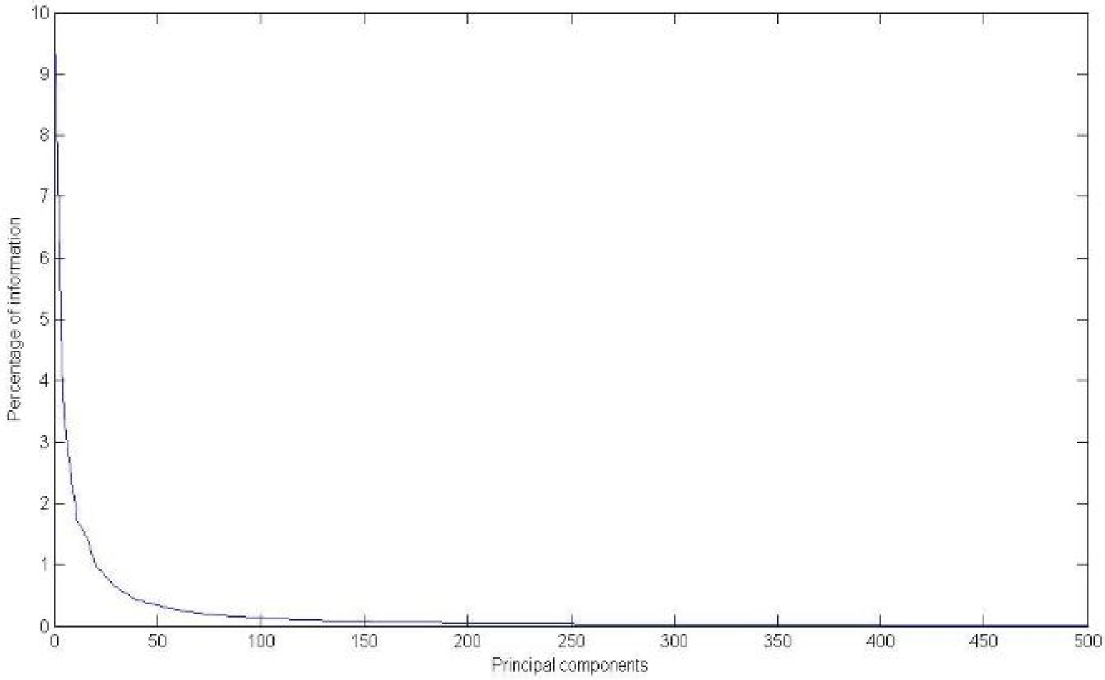

Figure 4.7: Latency plot for applied PCA on 4th convolutional layer

kernels but the classifier resulted in very small accuracy of about 30% which was the result of overfitted model.

With the following results we can say that cross-convolutional layer pooling method did not work for this dataset. One prabable reason could be the size of dataset since the dataset was very small for accurate training. A total of 2798 images for training and even the images are in stereo pairs which makes a total of 1399 images for training. While on the other side the said technique have been used for large datasets (Liu et al., 2015).

Chapter 5

Discussion

Preservation of every ecosystem in order to maintain a sustainable biodiversity is inevitable. Fresh water ecosystems are in danger and the liable factors include climate change, pollution and primarily over fishing. Over exploitation of certain species of fish population underwater certainly disturbs the ecosystem and marine conservatives must pay attention to the addressed problem. Thus continuous monitoring and sampling of fish population underwater is very important for coastal conservation status. Lately the task of sampling and tagging of fish population was being done manually which required both time and labor but it is not of good practice if we keep using the old system in today's era of real time monitoring. Use of advanced computer vision systems and cameras for underwater monitoring is being rapidly adopted. Automatic system of fish specie classification has its effects on the studying and monitoring underwater environment. Monitoring of fresh water environments is difficult where most people of the community depend on fishery for their earning. But failure in monitoring of such coastal areas will lead to the extinction of certain specie and this decline will eventually disturbs the whole marine ecosystems. Therefore repeated monitoring is of utmost need in areas where such threats prevail and government bodies should deploy these efficient fish identification systems in the areas for timely warning and preservation.

We have proposed to use deep learning approach for fish species classification on UWA dataset. Automatically recognizing and classification of fishes from videos/images in the presence of highly variable background and distortions poses a great challenge for machine learning community. The factors such as variable environment, distortions and noise in the image plays a major role in decreasing the performance of machine learning algorithms. These factors effects the linearity of data and make it non-linear where most of the machine learning algorithms are linear. Hence it effects the performance of such algorithms and emphasizes the researchers to move towards the use of non-linear architectures such as multi-layer neural networks such as Convolutional Neural Networks. Such multi-layer architectures are highly complex as every hidden layer is input to the next hidden layer. The depth of such automatic non-linear learning systems depends on the non-linearity of the input data. Therefore we opted to use deep convolutional neural systems for our task which is classification of fish species in unconstrained underwater environment.

Conventional machine learning algorithms such as KNN and SVM failed to perform well when trained and tested on our dataset (as shown in results). It is either because they cannot handle the non-linearity of input data or over-fit to certain environment. We tried training on a much highly non-linear and complex architecture which is using activations from hidden convolutional layers to make feature vector for training and testing. Our technique being the latest (as per our knowledge) to be used for this task of fish species identification is the first try on any dataset.

We could not achieve the best results with our technique as the dataset we opted for training and testing have had quite a number of issues that made the classification even harder. Firstly the dataset was very small as the images were in stereo pairs and the background was highly diverse making it very less for training. Thus our training over-fit to the training data and provided very less accuracy when tested on images. And the technique of cross-layer pooling on the other hand has been used for very large datasets such as MIT dataset which contain a total of 15620 images under 67 indoor categories.

Another important factor that could be a possible reason of failure is that the images in the dataset were of very low contrast which made the features almost invisible and impossible to be classified. And some of the images were unusable because of high intrusion from camera and murkiness of water. The 'other' class in the dataset contain a large number of images which are of no interest and prove to be a disturbing factor in training.

Chapter 6

Conclusion and Outlook

Fish identification and monitoring in unconstrained underwater environment is of high interest not only for marine researchers and biologists but for geographical and biodiversity related committees as well, which work to maintain a sustainable environment on earth. The manual techniques which were being used in past posed danger for living organisms in the sea beds. But now with machine learning and deep learning, many automated methods have been discovered for monitoring in benthic zone. These techniques lower the probability of extinction and help in survival of fish and other organisms present in the underwater environment.

To conclude we have proposed and implemented a new technique of using activations from convolutional layers as feature vectors instead of using activations from fully connected layer as feature vector for classification. This novel approach has performed quite well for large datasets used for other classification tasks but for our task of fish species identification and with a new dataset, which is relatively small, it could not perform well. Main reason of not performing up to the mark is its size limitation and existence of 'other' class which creates confusion in classification.

In future we aim to use the technique of cross-convolutional layer pooling for some other datasets of fish species recognition in unconstrained underwater environment for comparative study. In addition we aim to refine the UWA dataset to get better accuracy.

References

(2014). The new lab of clef dedicated to life media.

Bengio, Y. (2009). Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127.

Cappo, M., Harvey, E., Malcolm, H., Speare, P., et al. (2003). Potential of video techniques to monitor diversity, abundance and size of fish in studies of marine protected areas. *Aquatic Protected Areas-what works best and how do we know*, pages 455–464.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., Stiassny, M. L., et al. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews*, 81(02):163–182.

Gibson, R., Atkinson, R., and Gordon, J. (2016). A review of underwater stereo-image measurement for marine biology and ecology applications. *Oceanography and marine biology: an annual review*, 47:257–292.

Harvey, E. and Shortis, M. (1995). A system for stereo-video measurement of sub-tidal organisms. *Marine Technology Society Journal*, 29(4):10–22.

Hsiao, Y.-H., Chen, C.-C., Lin, S.-I., and Lin, F.-P. (2014). Real-world underwater fish recognition and identification, using sparse representation. *Ecological Informatics*, 23:13–21.

Huang, P. X., Boom, B. J., and Fisher, R. B. (2015). Hierarchical classification with reject option for live fish recognition. *Machine Vision and Applications*, 26(1):89–102.

Jennings, S. and Kaiser, M. J. (1998). The effects of fishing on marine ecosystems. *Advances in marine biology*, 34:201–352.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Larsen, R., Olafsdottir, H., and Ersbøll, B. K. (2009). Shape and texture based classification of fish species. In *Scandinavian Conference on Image Analysis*, pages 745–749. Springer.

Lee, D., Redd, S., Schoenberger, R., Xu, X., and Zhan, P. (2003). An automated fish species classification and migration monitoring system. In *Industrial Electronics Society, 2003. IECON'03. The 29th Annual Conference of the IEEE*, volume 2, pages 1080–1085. IEEE.

Lee, D.-J., Archibald, J. K., Schoenberger, R. B., Dennis, A. W., and Shiozawa, D. K. (2008). Contour matching for fish species recognition and migration monitoring. In *Applications of Computational Intelligence in Biology*, pages 183–207. Springer.

Liu, L., Shen, C., and van den Hengel, A. (2015). The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4749–4757.

Mallet, D. and Pelletier, D. (2014). Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). *Fisheries Research*, 154:44–62.

Murphy, H. M. and Jenkins, G. P. (2010). Observational methods used in marine spatial monitoring of fishes and associated habitats: a review. *Marine and Freshwater Research*, 61(2):236–252.

Nadarajan, G., Chen-Burger, Y.-H., Fisher, R. B., and Spampinato, C. (2011). A flexible system for automated composition of intelligent video analysis. In *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*, pages 259–264. IEEE.

Rova, A., Mori, G., and Dill, L. M. (2007). One fish, two fish, butterfish, trumpeter: Recognizing fish in underwater video. In *MVA*, pages 404–407.

Ruff, B., Marchant, J., and Frost, A. (1995). Fish sizing and monitoring using a stereo image analysis system applied to fish farming. *Aquacultural engineering*, 14(2):155–173.

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Jackson, R. B., Kinzig, A., et al. (2000). Global biodiversity scenarios for the year 2100. *science*, 287(5459):1770–1774.

Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., and Harvey, E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*.

Scholkopft, B. and Mullert, K.-R. (1999). Fisher discriminant analysis with kernels. *Neural networks for signal processing IX*, 1(1):1.

Shortis, M. R., Ravanbakskh, M., Shaifat, F., Harvey, E. S., Mian, A., Seager, J. W., Culverhouse, P. F., Cline, D. E., and Edgington, D. R. (2013). A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. In *SPIE Optical Metrology 2013*, pages 87910G–87910G. International Society for Optics and Photonics.

Spampinato, C., Chen-Burger, Y.-H., Nadarajan, G., and Fisher, R. B. (2008). Detecting, tracking and counting fish in low quality unconstrained underwater videos. *VISAPP (2)*, 2008:514–519.

Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y.-H. J., Fisher, R. B., and Nadarajan, G. (2010). 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*, pages 45–50. ACM.

Storbeck, F. and Daan, B. (2001). Fish species recognition using computer vision and a neural network. *Fisheries Research*, 51(1):11–15.

Strachan, N. (1993). Recognition of fish species by colour and shape. *Image and Vision Computing*, 11(1):2–10.

Strachan, N. and Kell, L. (1995). A potential method for the differentiation between haddock fish stocks by computer vision using canonical discriminant analysis. *ICES Journal of Marine Science: Journal du Conseil*, 52(1):145–149.

Strachan, N., Nesvadba, P., and Allen, A. R. (1990). Fish species recognition by shape analysis of images. *Pattern Recognition*, 23(5):539–544.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86.

Walther, D., Edgington, D. R., and Koch, C. (2004). Detection and tracking of objects in underwater video. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–544. IEEE.

White, D., Svellingen, C., and Strachan, N. (2006). Automated measurement of species and length of fish by computer vision. *Fisheries Research*, 80(2):203–210.

Xiao, T., Zhang, J., Yang, K., Peng, Y., and Zhang, Z. (2014). Error-driven incremental learning in deep convolutional neural network for large-scale image classification. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 177–186. ACM.

Yangqing Jia, B. (2014). Caffe.

Zion, B., Alchanatis, V., Ostrovsky, V., Barki, A., and Karplus, I. (2007). Real-time underwater sorting of edible fish species. *Computers and Electronics in Agriculture*, 56(1):34–45.

Zion, B., Shklyar, A., and Karplus, I. (1999). Sorting fish by computer vision. *Computers and Electronics in Agriculture*, 23(3):175–187.

Zion, B., Shklyar, A., and Karplus, I. (2000). In-vivo fish sorting by computer vision.

Aquacultural Engineering, 22(3):165–179.