Deep Learning based Classification Framework to save Right Whales

Author

Muhammad Ali

Fall 2015-MS-85(CE)00000119183

Supervisor

Dr. Arslan Shaukat

Co-Supervisor

Dr Usman Akram

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

August, 2019

Deep Learning based Classification Framework to save Right Whales

Author

Muhammad Ali

Fall 2015-MS-85(CE)00000119183

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

MS Computer Engineering

Thesis Supervisor

Dr. Arslan Shaukat

Thesis Co-Supervisor

Dr Usman Akram

Thesis Supervisor's Signature: ____

Thesis Co-Supervisor's Signature: _____ ______________________________

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD

July, 2018

Declaration

I certify that this research work titled "*Deep Learning based Classification Framework to save Right Whales***" is my own research work.** The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Muhammad Ali

Fall 2015-MS-85(CE)00000119183

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the university.

Signature of Student

Muhammad Ali

Fall 2015-MS-85(CE)00000119183

Signature of Supervisor

Dr.Arslan Shaukat

Signature of co-Supervisor

Dr.Usman Akram

Copyright Statement

Acknowledgements

All praise and glory to Almighty Allah (the most glorified, the most high Merciful) who gave me the courage, patience, knowledge and ability to carry out this work and to persevere and complete it satisfactorily. Undoubtedly, HE eased my way and without HIS blessings I can achieve nothing.

I would like to express my sincere gratitude to my advisor Dr.Arslan Shaukat and Dr. Muhammad Usman Akram for boosting my morale and for his continual assistance, motivation, dedication and invaluable guidance in my quest for knowledge. I am blessed to have such a co-operative advisor and kind mentor for my research.

Along with my advisor, I would like to acknowledge my entire thesis committee: Dr. Sajid Gul Khawaja, Dr. Farhan Riaz for their cooperation and prudent suggestions.

My acknowledgement would be incomplete without thanking the biggest source of my strength, my family. I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout in every department of my life and my loving sisters who were with me through my thick and thin.

Finally, I would like to express my gratitude to all my friends and the individuals who have encouraged and supported me through this entire period special thanks to them.

*Dedicated to my exceptional parents: **Habib Ahmed**, my father my best friend,my mentor who's cooperation led me to this accomplishment. He was always there for me in all circumstances. He has made numerous sacrifices for my success. He fulfill all my demands before they leave my tongue. A father like him is a great blessing of Almighty Allah .I would also like to thank my wife and my beloved daughter Hania, for supporting me throughout my research work .*

Abstract

Few decades ago, experienced Marine biologists started a campaign to safeguard Right Whales with a name of National Oceanic Atmospheric Administration. Different aerial campaigns were started by them, with the primary objective of studying their health and counting their population. Photographs were taken by helicopters and then these were compared with an online database. Manual comparing of Right Whales was very time consuming and lot of training and experience was needed to do it.

So to overcome this issue NOAA in collaboration with Kaggle decided to launch a competition, the objective was to launch a system to monitor Right Whales and efforts must be done to free a Right Whale that has been accidentally caught in fishing gear. We only have 4544 training images in our dataset; training deep convolutional neural networks with such a small number is a real challenge. These photographs are taken by helicopters, so they are badly focused. They are taken at different times of day with different quality of cameras. Some images of Right Whales are of really bad contrast with poor exposure. Moreover the dataset was not balanced i.e. the number of pictures per whale varies a lot. We have about 20 whales with just 1 image, some whales have around 40 images, the average number of images per whale was around 10.

With such a sparse distribution, it was really challenging for us to train our deep convolutional neural networks. To minimize the effect of small dataset, we have divided our problem wisely. Instead of training our neural network on whole images, which is no use to us, as most of the images contains ocean waves. We first localize head of the whale, in this way we can focus more on our desired feature, that is the callosity pattern on Whale’s head. After localizing head we find two points Bonnet and Blowhead, callosity pattern lies between these two points. We then aligns whale’s head in such a way, that Bonnet is on right blowhead is on left and whale’s head is pointing towards east. Now images are align, so our classifier can only focus on area of interest. This has improved our results significantly aswe have achieved an accuracy of 78.70%. The results can be improved if we have more images in the dataset or if we have more clearer images with better resolution and area of interest i.e Callosity Pattern more clearly visible.

Key Words: *Kaggle, Deep convolutional neural networks, deep learning, fully convolutional network, Right Whales , Callosity Patterns , Bonnet and Blowhead*

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

Right whales is considered one of the most endangered species now a days. Their expected number has decreased to even less than 500, so a campaign has started to preserve them [1]. Actually since 17th century, right whales have been favorite hunting attraction, due to their countless benefits. These whales have been ideal for hunting due to their behavior of living close to shores. Their economical advantage is that they are rich in whale oil. They are easy target for hunters because their bodies float when killed. Despite being protected species since 19th century, there life is under continuous threat due to numerous reasons like being caught in fishing gear or unwanted collisions by ships.

1.1 Motivation

It is different from the scenario where we have to classify between mountains, animals, trees , like totally different objects. Here we have two challenges first we have to identify that it is a right whale or not and then we have to tell which of the 427 whales, the whale is according to whale ID's. It is difficult even by physically examining the two pictures that these whales are same or not, although by looking at clear perfect pictures. The area of interest is very small the callosity pattern, so it is difficult for classifier to focus on it regardless of image quality.

1.2 Problem Statement

To make our classifier just to focus on areas of interest i.e the callosity pattern, we first localize the head of the whale, water and remaining of whale is of no use to us. After localizing head we have to find callosity pattern, but first we have to find two points 'Bonnet' and 'Blowhead' i.e starting and ending points of callosity pattern[3]. These points are very good features for us as they are different for each whale. Callosity pattern lies between these two points, after detecting these patterns we will train our neural networks on it. After Localizing head and finding blowhead and bonnet tip, we will align our images all in one direction. Bonnet pointing to the right and blow head to the left, whale head towards east[4,5]. These images will be passed through classifier which will calculate the similarity probability distribution of the whale with testing set.

1.3 Aims and Objectives

Each whale has a unique callosity pattern on its head, it's like similar to human finger print as every person has unique finger print pattern, likewise whales have unique callosity pattern . Some whales have continuous and some have broken callosity patterns (this is explained in detail in later chapters). Thanks to the whale lice on their head callosity pattern appear white forming a good contrast with rest of image and becoming an excellent distinguishing feature for us[6-8].

There were lot of challenges that we faced while completing this some of them are mentioned in this paragraph. The most severe challenge was that we only have limited dataset available .There were only 4237 images for 427 right whales. Actually number of images varies a lot some whales have around 40 images most of them have around dozen and some whales have less than 10 images per whale[9]. There were about 20 whales that have only one image in the dataset, if we add them in training dataset it will be over fitting we can't test them and if these single images are added in testing them, it will decrease our accuracy percentage significantly. Images quality varies a lot, because images were taken at different times of the day with different cameras and with different heights. Contrast of some images was not good, some images were overexposed and some are underexposed, light intensity on whale images also varies a lot.

1.4 Structure of Thesis

This work is structured as follows:

Chapter 2 covers all the necessary Details required to know about Right Whales, their characteristics, features of our interest.

Chapter 3 gives review of the literature and the significant work done by researchers and machine learning engineers in deep learning.

Chapter 4 consists of the proposed methodology in detail. Different approaches used, lesson learned, which approaches gave us success and which does not.

Chapter 5 introduces the databases and experimental results of Kaggle Right Whale Challenge.

Chapter 6 concludes the thesis and reveals future scope of this research.

CHAPTER 2: Right Whales

Inspite of international protection from commercial whaling since 1935, north atlantic right whales is still suffering the major threat of getting extinct[10]. This is because Whales love to live near coastal areas and these areas are also used by international shipping, military and fishing industry [11]. Recent Mortality rate is really alarming, in last 14 months 9 whales are reported dead due to collisions with ships [12].

Right whales are getting extinct from Atlantic Ocean. Their survival is under major threat due to various reasons. A solid step must be taken to preserve them, before they only remain in our books. A group of marine biologists started a campaign to preserve them, this organization is called NOAA (North Atlantic Right Whales). The started aerial surveys by helicopters, to study the health behaviors and take aerial photographs of whale. Marine biologists than match these photographs with their register that which of the 437 whales this right whales is that we are identifying[14]. This manual process is not feasible in real time.

Figure 2.1:Callosity Patterns 1

2.1Characteristics of Right whales

In Figure 2.1 we are showing random images from dataset, all the preprocessing is done on them and they are ready for training in recognition model. There are numerous features by which we can classify whales, the best one is by their callosity pattern, some have continuous and some have broken callosity patterns. Thanks to the lice in whale which forms white patches on whales head[14]. Each whale has different callosity pattern like every person on earth has different finger print pattern and different retina scan.

Other two important points on whale head are bonnet and blowhead. Bonnet is at end of whale’s head, from here callosity pattern start’s and other point which is little far away is blow head.

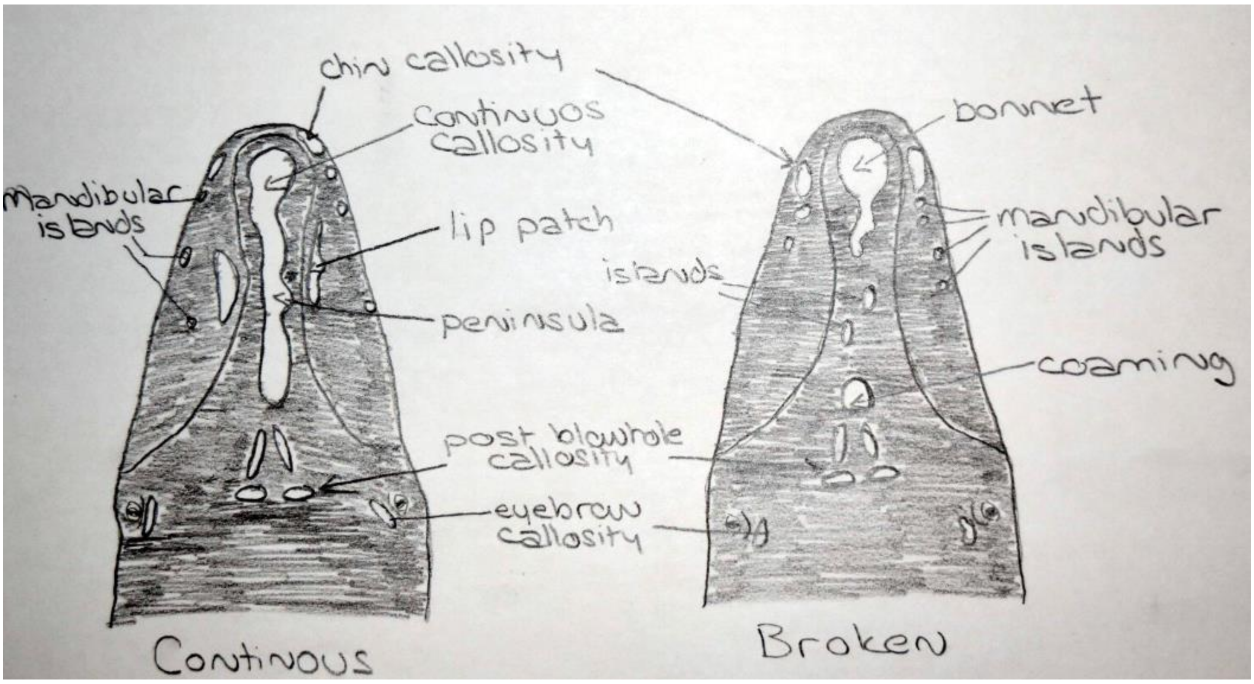

Figure 2.2:Whale's Head Structure [Felixaumon[5]]

The Figure 2.2 above shows vital points on whale’s head. Other features include chin callosity and eye brow callosity but it is not common to all whales and hence does not form a good set of features on which we can train our convolutional neural network.
Right whales grow upto 50 feet and weight about 75 tons.

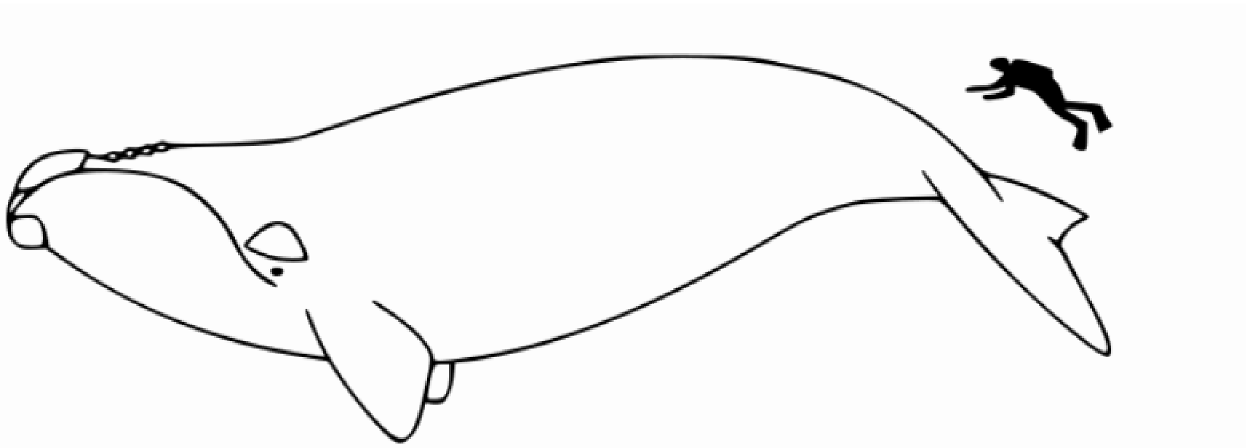

Figure 2.3:Estimating Size of Whale[61]

This Figure 2.3 shows just a rough difference to estimate the size of whale as compared to a human. Their average age is 75 years but few can manage to live a century. Their appearance include they don't have dorsal fins and are dull black in color. Their diet includes eating other small marine species. Their females grow larger in size than males. Their head is enormously large about ¼th of the body. Their birth rate, reproduction is slow, females gives birth to calf after 10 years. Seeing cross sections in their teeth is one way to identify the age of mammals, but surprisingly whales have no teeth. Right Whales just have baleens. In whales, eye lenses and in some cases ear bones are used to identify their age. Right whales feed themselves in spring to fall and in rare cases also in winter. Their favorite place to live is coastal areas near shallow waters.

2.2 Threats to right whales

As we know whale oil was in huge demand and rigorous hunting of Right Whales are causing them to be extinct. Now according to laws hunting right whale is a major crime, now lot of measures are taken to protect them from extinction. There are still threats to their life; some of them are discussed here. As Right whales prefer to live in shallow water near coastal areas there is a constant threat of ship collision to them. Ship collisions are reduced by numerous measures, areas are marked red or no ship zone in the region where the concentration of right whales in more. Some are unexpectedly caught in fishing gear. Other important factor that is affecting the lives of right whales in water pollution. Water contamination is making them sick, creating a very unhealthy environment for their survival. Global warming and other climatic variations are also not suitable for them. Noise of ships and disturbance caused by whale surveillance is also not good for them. It seems that Right whales are not friendly in adapting new things and they are not comfortable in having company with them.

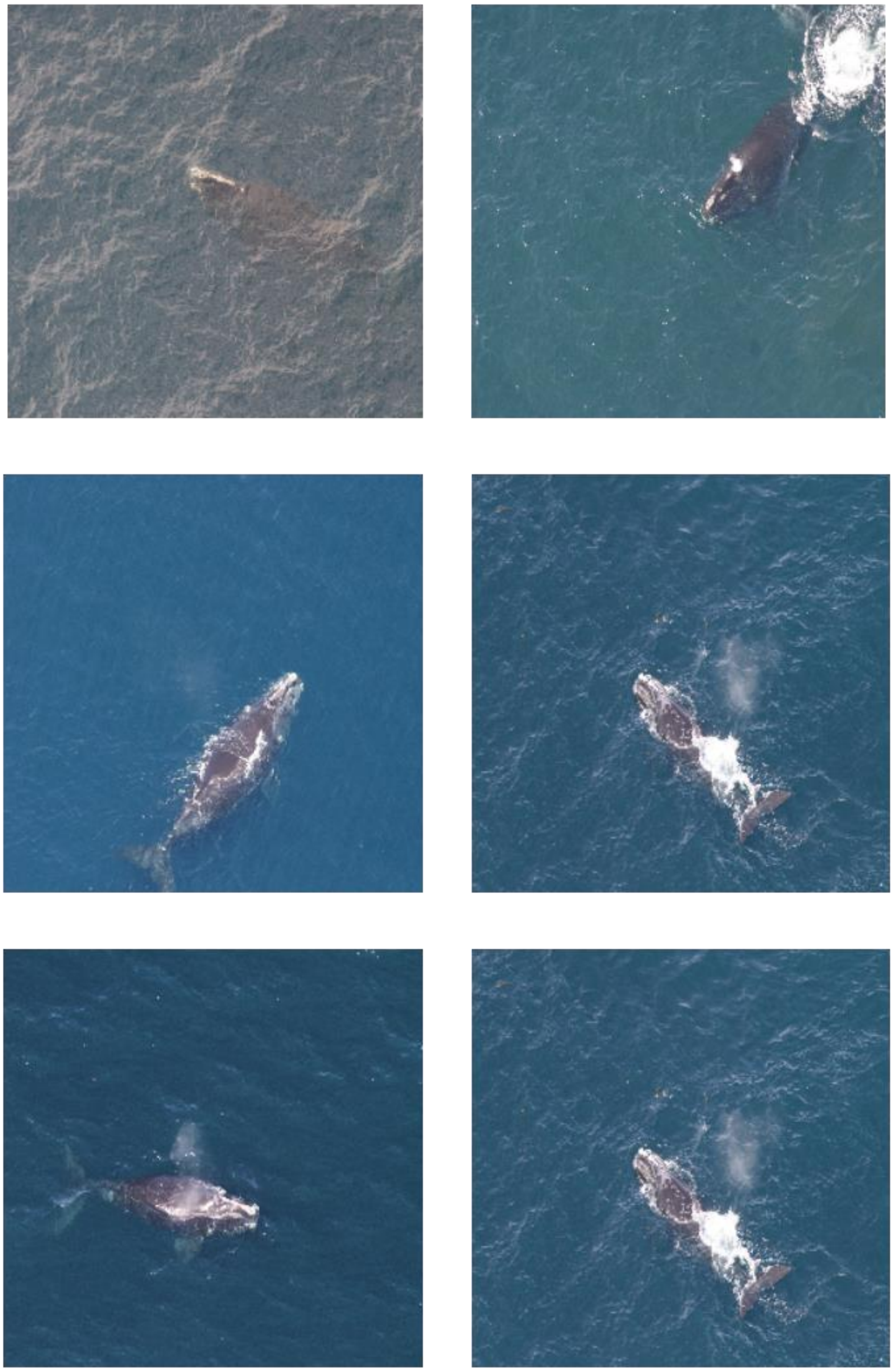

Fig 2.4 Random Images from dataset

Figure 2.4 shows some pictures of whales just for a small challenge and you have guess which pair of whales is same and which is not, believe me it's not that simple by looking without lot of experience.

Have you taken this challenge? Hope it will be a pretty good brain exercise for you! Now time to announce the results. First two pair of images are different whales, 3rd pair of whale is same, as you have seen by results it's very difficult to recognize manually just by looking without any proper training by marine experts.

Actually the difficulty arises because images are not of good quality, this is because some are badly focused, and there contrast is very poor. They are taken at different times of day so some are dark and some are bright. They are taken from different cameras having different frame size and different resolution. Plus whales don't know how to pose for images, or they don't care about their good looks as humans. In some images head is half visible. In Few images splash of water is covering the head.

CHAPTER 3: LITERATURE REVIEW

Deep learning is the emerging era of machine learning research brining remarkable revolutions in Computer Vision, pattern recognition and artificial intelligence. It extracts several types of features namely high level and low level features, It works like brain where neurons carry information in a layer wise fashion[16]. The deeper the layers the more information is processed in them. A lower level layer contains the major details like taking the example of an image, lower layers will identify edges and higher level layer will contain minute details about that image. In a nut shell it's overall purpose is to optimize objective function that is done during training[17].

A common neural network consists of many simple processors called neurons, where each neuron is producing its own real valued activation. First layer of neurons get activated from sensors, subsequent layers use values of previous layers multiplied by weights. Now the process involved of the training weights so that actual output of the model matches our expected output[19]. A general purpose of machine learning is to learn from representations in input data and to use this information in predicting unseen data. Large data sets available with increased computation capacity played a vital role for the interest in deep learning. Extracting good relevant features is an another major challenge which will bring revolution to deep learning. Extracting relevant and those features which will give amazing classification results without human intervention in under research.

Deep learning has shown amazing results in discovering intricate structures in high dimensional data, so it is most widely used in medical science, in making predictions and government. Barak Obama used machine learning on the tweets of people to know their sentiments about him, processing was done across all the states so that he can have a know-how of which state is with him and which is against him. In this way he can successfully focus on only those states that are against him. Deep learning has also achieved amazing results on speech recognition [20-22] and image processing [23-26]. Other astonishing improvements were in reconstructing brain signals [27], predicting the effect of mutations in non-coding DNA on gene expression and disease[28,29] and in analyzing particle accelerator data[30,31]. Superb results are also show on natural language processing [32,33] specially question answering[34] , sentiment analysis and in translation[35].

3.1 Unsupervised learning

Numerous Models are available for deep neural networks. The most commonly used was auto encoder model which is now known as deep auto encoder model. It is the refined form of auto encoder model with multiple hidden layers in it[36]. It is a form of generative models and extracts features with the help of hidden layers in it. Similar to the conventional neural networks, deep auto encoder model performance is also dependent on their weight initialization method. Improper weight initialization results in saturation and network paralysis [37]. Deep auto encoder networks also encounter difficulties while propagating gradients in backward direction.

Unsupervised learning is computationally expensive so they are not good with large datasets. To do so we integrate, convolutional models with generative models [38]. Convolutional layers with pooling make it convenient and feasible with large datasets especially in the field of medical sciences.

3.2 Supervised learning

The most widely used machine learning technique at present is supervised learning. Among supervised learning one of the most protuberant approach is Convolutional Neural Networks (CNNs). They are based on Hubel and Wiesel's Discovery of cell [39]. It is originated by neorcognitron model proposed by Fukushima [39]. Convolutional Neural Networks creates its own invariance by shared weights technique. It was applied to recognition of hand written characters [40]. CNN follows the concept of shared weights and subsampling/pooling. The convolutional neural networks are considered different from conventional neural networks where each neuron is replaced by three dimensional feature space maps. Here patch of neurons is connected to the neuron of next layer in feature map. The size of the patch is governed by the type of example in which it is used. Convolution followed by averaging/subsampling making it robust towards distortion and noise in input.

Advanced applications in pattern recognition system and computer vision has gained a significant amount of attention from researchers all around the word. In LeNet-5[41] advanced network with seven layers was applied to MNIST [42]. The seven layers were further divided into two parts. The first part consists of 4 layers two convolutional layers followed there pooling layers. The first half was responsible for extracting features from raw data. Second half consist of two fully connected layers which was responsible for classifying the features that are extracted in first part based on the assumption that they are linearly separable from each other.

This has gained considerable amount of familiarity in 2012 after winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[43] by AlexNet [44]. The Network [45] followed the same approach as used by LeNet, but the depth of network was more than the hidden layers (five convolutional and three fully connected layers). To suppress over-fitting dropout technique was used [46,47]. Then the network with more hidden layers has will be given foremost priority for feature extraction [48,49].

In 2014, Lin[50] presented a different kind of algorithm called ‘Network in Network(NIN)’ it uses multilayer perceptron networks in convolutional layers. Here we have assumed that features extracted are linearly separable but this is not true for all the cases. We know that multilayer perceptron networks are universal approximator using them can disentangle many nonlinear functions easily. Depth of such approach is still to be explored; a lot of work is needed to be done on it.

Activation function of hidden units plays a vital rule in the performance of our networks. The LeNet-5[51] network works on hyperbolic tangent functions. The activation functions like hyperbolic tangent and sigmoidal works amazing with shallow neural networks [52]. Their performance is degraded in Deep Neural Networks because error is propagated in backward direction so it gets saturated. Moreover there weight matrices are very dense which creates extra overhead for the networks. Moreover weight matrices are also saturated which increased computational complexity for neural networks. The rectified linear units (ReLu) are performing much better than biological neurons[53], that’s why they have become the best choice of researchers for activation function .

	Tensorflow	**Caffe**	**Keras**	**Theano**	**MXNet**	**Deeplearning4j**	**Torch**	**Matlab + NeuralNetworks**	**PyTorch**	**OpenNN**	**Dlib**	**neural designer**
Developed By	Google Brain Team	Berkeley Vision and learning center	Francois Chollet	Universite' de Montreal	Apache Software Foundation	Skymind engineering team	Ronan Collobert, Koray Kavukcuoglu, Clement Farabet	MathWorks	Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan	Artelnics	Davis King	Artelnics
Platform	Linucx, MacOS, Windows, Android	Linux, MacOS, Windows	Linux, MacOS, Windows	Cross platform	Linux, MacOS, Windows, AWS, Android, IOS, Javascript	Linux, MacOS, Windows, Andoird	Linux, MacOS, Windows, Andoird	Linux, MacOS, Windows	Linux, MacOS	Cross platform	Cross platform	Linux, MacOS, Windows
Language used	C++, Python, CUDA	C++	Python	Python	C++	C++, Java	C, Lua	C, C++, Java, MATLAB	Python, C, CUDA	C++	C++	C++
GPU support	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
CUDA support	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Parallel Execution	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Release date	2015, Nov 9	2017, April 18	2015, March 27	2017, Nov 15	2017	2017, August 13	2017, Feb 27	2017	2016, October	2014	2002	2014

Table 3.1: Detailed Comparison

In Table 3.1 we have done an extensive research on different deep neural network findings and have gathered all the information in a tabular form. It is the summary generated of important findings.

3.3 Logistic Regression

There are a lot of machine learning algorithms, like naïve Bayes, Support vector machines, decision trees etc. Logistic regression is mostly the first algorithm that a machine learning engineer learns. There a number of reasons behind its popularity, it's scalable, it's fast, it can be easily interpreted and it's outputs are well defined predicted probabilities. For example take the example of car, bus and apple. Suppose our neural network is completely trained and we have a test image of a car, so the expected results will be 85% similarity with car, 14.5% similarity with bus and 0.5% similarity with apple.

3.4 Batch Normalization

Actually in batch normalization, we normalize the values (outputs) of hidden layers in a neural network. This reduces the uncertain impact of earlier layers on later layers. Training of neural network is much faster with batch normalization. It makes the algorithm really powerful. Dropout is also applied in our right whale challenge. Dropout is the technique where randomly selected neurons are cut off ; there input and output connection is over. As some neurons are dropped randomly, other neuron takes their place in learning the representation. It depends on scenarios, sometimes it improves results for us, sometimes it make things worse.

3.5 Max pooling

Max pooling is done usually to reduce computational cost, where parameters are large enough to train and it is difficult to handle such a large volume of data. The primary purpose of max pooling is to reduce the size of image while maintaining the maximum information of image. Working of max pooling is really simple to understand, suppose starting with basic, in 2x2 max pooling, a 2x2 max filter slides through whole image with a stride of 2. This filter selects the maximum value among the 4 values that are in filter, hoping important information is retained. Stride of 2 meaning that we will move our filter 2 steps ahead, for example the four boxes that a filter is covering will not traverse any of these values again. Similarly in 3x3 max pooling we have to use stride of 3.

3.6 Back propagation

Back propagation is the basic and most commonly used technique to train neural networks. It is mostly used by gradient decent algorithm to optimize the weights of neurons by calculating the values of gradient and loss function. In this error is calculated in the output and is back propagated backwards for the purpose of reducing it.

3.7 Fine tuning

For some applications, small training data is accessible. Convolutional neural systems more often than not require a lot of training data so as to abstain from over fitting. A typical procedure is to prepare the system on a bigger informational collection from a related area. Once the system parameters have met an extra preparing venture is performed utilizing the in-area information to tweak the system weights. This permits convolutional systems to be effectively connected to issues with small training data.

3.8 Convolutional Neural Networks

Convolutional Neural Networks CNNs are made up of neural networks that is used to process large amounts of data sets. The name convolutional neural networks shows that it contains mathematical operation convolution. It is actually a linear operation. Basic definition of convolution is the overlap of one function f with another function g. In this a function is rotated and traversed over another function. It is nearly similar to cross correlation. In fully connected neural networks each neuron input is connected to the other neuron in the output. To understand convolution more take the example that we are tracking our enemy with laser light. Our sensor is showing the value of x(t), that is the position of our enemy at time t. Both x and t are real valued which shows that we can get different values at different time instants. Supposing that our sensor is noisy then we have to take several measurements and then their average value.

$$y(t) = (x * v)(t) \qquad \{3.1\}$$

This asterisk "*" is used to denote convolution. In the above equation 'x' is usually called input, 'v' is called kernel and output is referred to as feature map.

In our scenario of right whale recognition input is usually a multi-dimensional array of pixel values and kernel is filter that is adapted by learning values to give best possible output. We called this multi-dimensional array as tensors.

3.9 Literature review

Here we discuss and review some latest research done on whale classification. Kabani knows data set was very small and it's really challenging to train neural network on this dataset . This problem was solved by dividing the problem into smaller tasks by reducing view point variance in our Right whale dataset . First of all he localizing the body and head of the whale using deep learning . Then he align the whale by finding two points Bonnet and Blowhead along which callosity pattern exists . Finally a recognition network is used to classify whales in their respective classes [17] .

Kostiantyn uses PCA algorithm for preprocessing raw images of dataset . Applying PCA shows that reducing features have significantly increase performance of classifier and reduces overfitting . He proposed a classifier with three layer neural networks of 40 , 100 and 400 input neurons . He achieves F1 score of 90% [18] .

Yin [54] find out that Whale vocalizations can be studied as polynomial-phase signals , which are mostly used nowadays in radar and sonar applications . These types of signals lie on a nonlinear manifold characterized by polynomial phase coefficients . Yin applied manifold learning methods particularly ISOMAP and Laplacian Eigenma . He find out that Laplacian eigenmap performs well than linear dimension reduction methods such as Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) in our scenario .

Dezhi [55] inspired by rapid advancements in the field of deep learning techniques and seeing successful applications and remarkable achievements in the different domains decided to apply it on whale call classification . A complete performance study of different state of the art CNN architectures is done on whale call classification dataset . Accuracy and computational efficiency varies with different CNN architectures . He find out that Xception [56] provides the best results in comparison to all four CNNs .

Ho Chun [57] used a unique approach for detecting and as well as classifying foraging calls of two mysticete species from the dataset of passive acoustic recordings . This detector/classifier uses a computer-vision based technique in combination with pattern recognition method, to achieve better results , for detecting the foraging calls and also for removing ambient noise effects . This data was collected from Southern California Bight . Ho Chun achieved an accuracy 92.03 % in pattern recognition algorithm while an accuracy of 95.20 % was obtained by machine leaning algorithm .

Marwa [58] introduced a feature selection technique that applies to whale optimization alogorithm(WOA) . She used wrapper-based method to find the best subset of features . These subset of features performs better than particle swarm optimization (PSO) and genetic algorithm (GA) algorithms .

G.Roberts [59] designed a non-stationary signal classification algorithm that uses time-frequency representation and a multiple hypothesis test . His time-dependent quadratic discriminant function shows that time frequency performs better than frequency domain method . At selected points in time he evaluates discriminant function and forms a set of statistics .

Mahdi [60] proposed in this paper that a study is carried out for detecting North Atlantic Right Whale upcalls with measurements taken from passive acoustic monitoring devices . Preprocessed spectrograms of upcalls are subjected to two different techniques , one is based on getting of time-frequency features from upcall contours and the other employs a Local Binary Pattern operator to find salient texture features of the upcalls .

Chenchen[62] knows that fish classification is not that easy because convolutional neural networks always need large data sets for training . So he uses transfer learning with Bilinear Convolutional Neural Networks(BCNNs) . In comparison to popular CNNs VGG net or ResNet , he achieves and accuracy of 78.25 % on Croatian fish dataset .

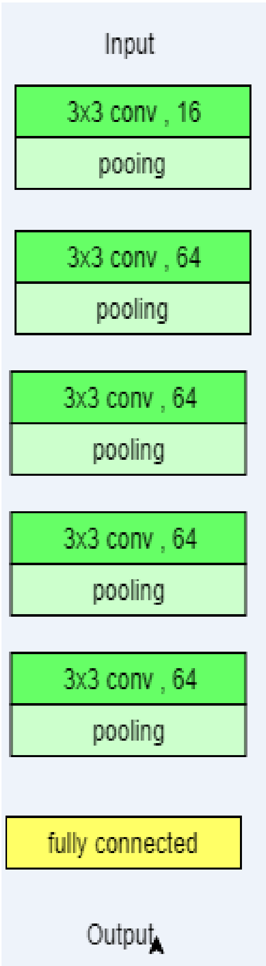

Fig : 3.7Head Localizer

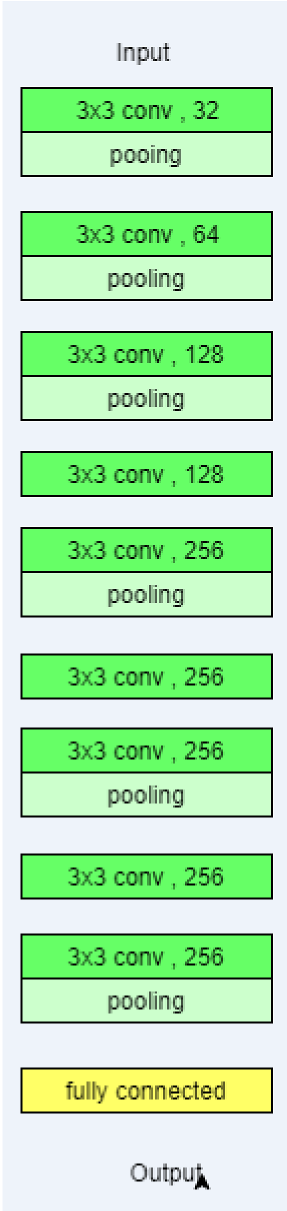

Fig : 3.8 Identify bonnet and blowhead

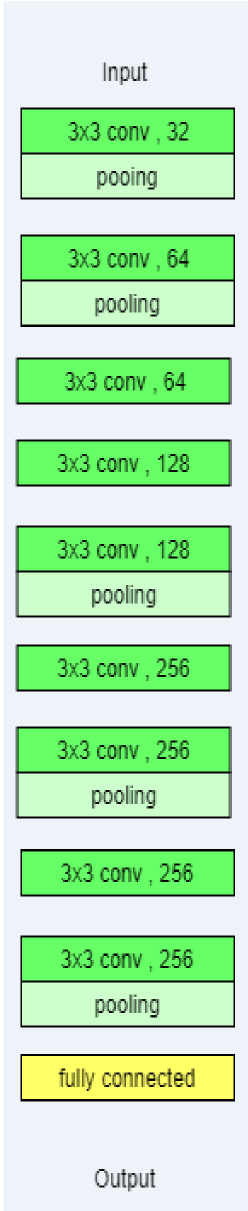

Fig : 3.9 Final Classifier

In Figure 3.7 Figure 3.8 and Figure 3.9, We have used three neural networks in our right whale recognition problem. One for localizing the head, then for identifying the callosity pattern identifying bonnet and blowhead and 3rd for classification of the whales. Our third neural network have two major tasks to do, one is to identify that it is a right whale or not and then to figure out from which of the 437 whales this whale is.

CHAPTER 4: EXPERIMENTAL METHODOLOGY

As it is completely clear by now that purpose of right whale recognition, was to save it from getting extinct. Our Focus was to design such a system that will help marine biologists to classify whale quickly and efficiently, because manual classification is extremely time consuming and required lot of efforts. We have used Deep Convolutional Neural Networks due to its exceptional accuracy on images from the last few years.

4.1 Data Set

Dataset was another major challenge in our classification of Right Whales. Actually it is the characteristics of Deep Neural Networks that huge amount of dataset are needed for training of neurons. The larger the dataset the precise the training is and hence better results. The dataset provided by NOAA to kaggle contains just 4237 images for 427 right whales

. So our primary objective to use this limited dataset widely to achieve maximum benefits. It was not so simple to split the Right Whale dataset randomly into testing and training data.

This was because number of pictures per whale varies a lot. Astonishingly twenty four whales are with only one image. This is a big challenge for us. These "celebrates" don't like fresh air a lot. Now it was difficult for us to decide, where to add these images. If these single images are added in testing set, we can't test them so our accuracy dropped significantly. If we add them in training set it will be our fitting as we can never test them, they will be of no use to us. Plus training our neural network on them is another significant challenge. Seeing the bright side, some right whales have around 40 images, most of them are around 15, few right whales are less than five. A data scientist can well image with such a diverse dataset, how hard it will be for us to proceed and achieve remarkable results.

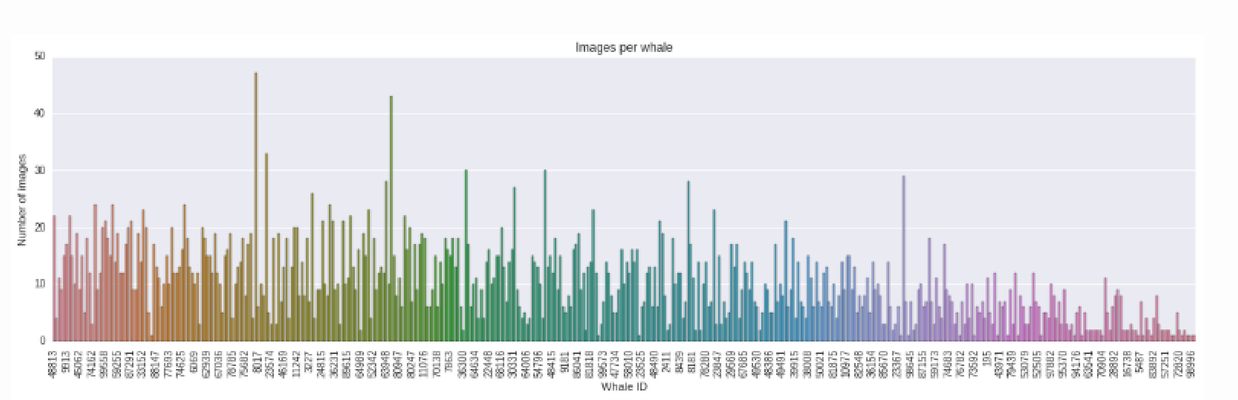

Figure 4.4: Dataset Distribution

In figure 4.7 the graph above shows dataset distribution according to Whale ID's. One can well image by seeing the bars how roughly distributed the data set it. A pretty good validation set was required to achieve remarkable results.

4.2 Dataset distribution

As shown by above bar chart and number of whales discussed with respect to images it is now clear that dataset distribution was not that simple. A stratified result will have boosted the accuracy a lot, initially I did random split of Right Whale dataset just for testing. It is a saying if you have done a mistake in start it will be with you and will keep on affecting other things. Things get worse when different classifiers were trained on different splits. I observed this issue when I start to notice results. So I decided to fix, so it's never late to correct things or to restart things in life in the correct manner, until they proceed in wrong manner all the way.

Classification between goat, bird, car, apple and trees is a pretty easy task as the similarities between them are very rare. If we have trained our convolutional neural network and suppose we are deducting trees so we will get a clear probability distribution, 90 % similarity with trees and remaining 10% will be distributed among goat, bird car and apple. This example was taken to clear our concept about the working of convolutional neural networks.

In our case the scenario is different. The area of interest only occupies small portion of the image, that is the Right whale head and in head specially the callosity pattern. We must not forget about data set distribution the images are of various quality and position and angel of whale in each image is different.

4.3 A glance at Dataset

We all know that dataset was not balanced and was in sufficient for training of neural networks. The quality of images was very random somewhere overexposed somewhere underexposed. In some images whale was occupying an extremely small portion of image. I have selected some images just to study them , they are discussed in this section below.

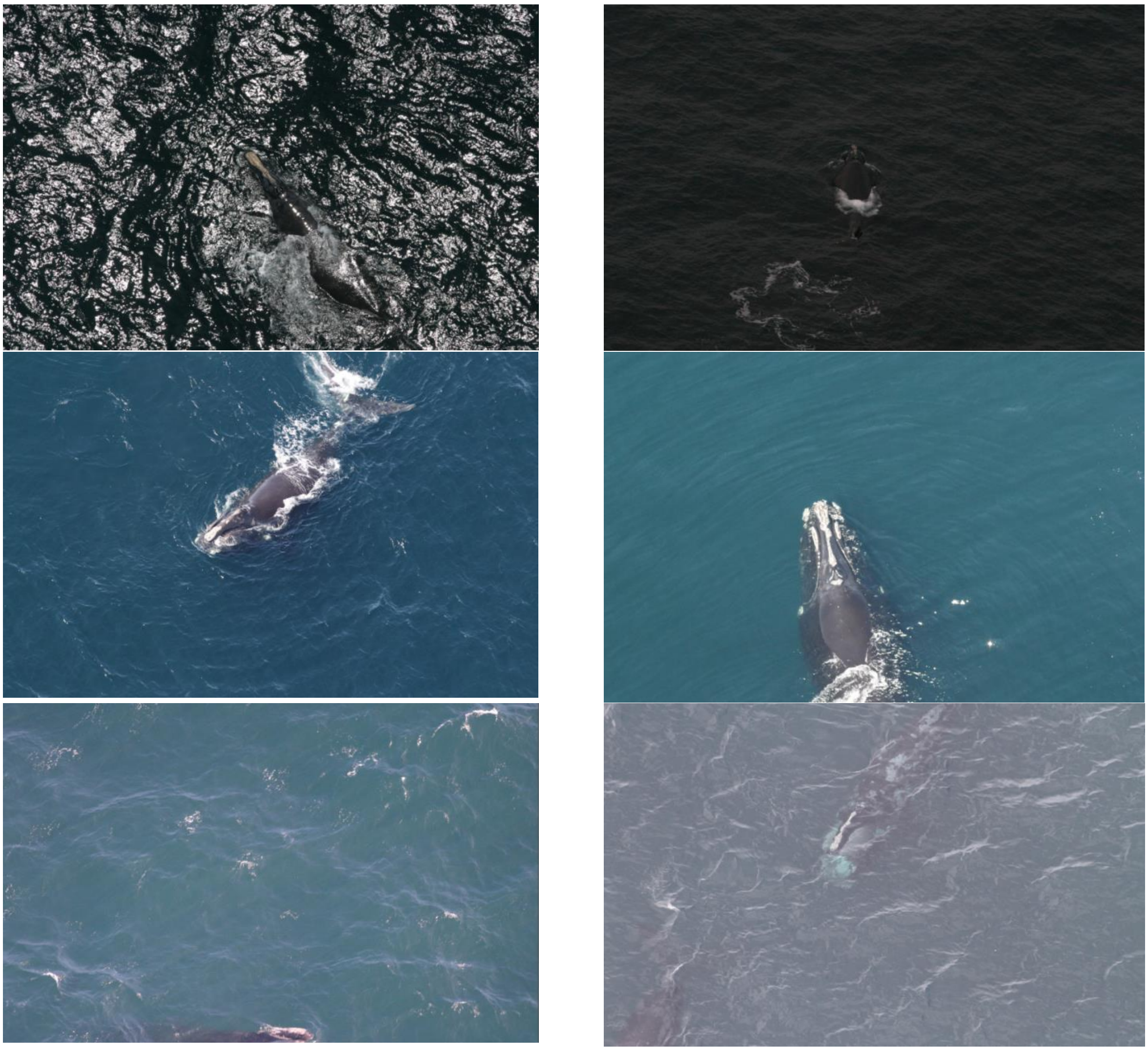

Figure 4.8 Showing variations in data set

Figure 4.8 shows that these images are taken at different times of day with different camera resolutions. See in A1 and B1 both are dark and low contrast images. It is nearly impossible by naked eye to see the callosity pattern on Right Whales head on these pair of images. A2 and B2 are pretty good whales head is focused and is more easily visible callosity pattern. The contrast, exposure location and size of Right Whale in these images are perfect. Now have a glance at A3 barely half of the whale is visible. We will add this image in our training set, but I don’t think it will give us some benefits. In B3 most of the whale is under water, actually whales don’t know that their photograph is being taken and they have to pose for it. Bonnet and blowhead points will be slightly difficult to locate here but callosity pattern will be extracted from this image, it will be difficult to do so because contrast, lightness level, exposure ,brightness scale and image quality is below threshold of a good image.

4.4 Software used

Almost all the code is written in anaconda python. Theano and cuDNN are used to implement neural networks. All the image processing rotation, shearing is done in scikit-image. Data processing of Right Whales is done in Pandas. All other operations are done in scikit-learn . Almost all the charts are used in Matplotlib .

4.4.1 Anaconda Python

Anaconda is a open source and free framework for python and R programming. it is widely loved by data scientists and machine learning engineers throughout the world as it is complete package of all the required libraries. It is suitable and works well for both windows and linux users. According to a report anaconda have roughly 7 million users worldwide. It's keep features include conda virtual environments though which all libraries are installed with a single click , no need to read and learn documentation of each library, it was created ease for programmers and have saved plenty of time. These commands 'pip install' and 'conda install' are doing the job for us. It has both command line interface and graphical user interface (GUI), so it is up to choice of developer in which he/she is comfortable. It also supports iPython, interactive python.

4.4.2 Theano

Theano is the open source library in python, it is designed to run on both GPU and CPU implementations. Its calculations and most of the work is done by numpy. It performs well on multi dimensional arrays means on images.

4.4.3 cuDNN

This is specially a GPU python library primarily designed for NVIDIA. cuDNN stands for Cuda Deep Neural Networks. It provides built in functionalities for forward and backword propagation, pooling, convolutions and activation functions. Currently it is giving amazing results for RNNs. It is supported on MacOs, Windows and linux.

4.4.4 Scikit-Image

Scikit is a complete and one of the best collections of algorithms to be used in image processing. It is available free of cost to developers. It covers almost everything like segmentation, color space manipulation, analysis and filtering and all important geometric transformations.

4.4.5 Pandas

It is freely available and is written for python for data manipulation and it's analysis. Data aligning reshaping, merging are its key features.

4.4.6 Scikit-Learn

It's primary features included regression, classification and clustering algorithms. It's mostly used when we wanted to use SVMs support vector machines. It was originated by a campaign called Google summer of code. It also supports numpy and scipy.

4.4.7 Matplotlib

It is plotting library (bar charts, graphs etc) designed to be used in python, it is an extension of numpy. All most all the graphs made here in Right whale challenge used matplotlib. It was originated by John D. Hunter . Of course free to use and easily available. It is also used to draw a variety of 3 D plots.

4.4.8 Numpy

It is mainly used to add support to deal with large multi-dimensional arrays, like large high resolution images. It is the mostly used library of python. Cross platform easy to use was one of the main factors we have used it in our Right Whale problem. Initially python was not designed for image processing, development of amazing libraries like numpy, scikit learn, pandas have attracted data scientists specially the one that are dealing with large datasets.

4.5 Hardware used

Most of the models are trained on GTX 670. During training GPU was extremely heated up. Air conditioner was turned on in the lab during training, but that was not affecting GPU Temperature. If it was winter I can assure that heater was not needed for sure.

Figure 4.5: GPU Temperature 1

This figure 4.8 above shows the heat generated by GPU, during training of Right Whales Data Set.

4.6 Methods Adopted

All our approaches are based on deep neural networks CNNs. Machines perform better than humans when there is lot of images, huge data set but in accuracy there is no match to humans. When we applied CNN on the raw images we were getting very poor results as most of the image contains ocean waves. We were getting accuracy of 5 to 10 %. Once we tested on passport size pictures of whales we have achieved an accuracy of about 50 %. This was much better than previous result but still not suitable for industrial use. Thanks to lice on whales head which forms callosity pattern. Localizing callosity patterns and applying CNNs on whales head improved our results remarkably providing Log Loss value of 1.65.

4.6.1 Baseline Naïve Approach on Whales

After started working on this competition and doing literature review the first thing I do was to establish a baseline classifier in CNNs.

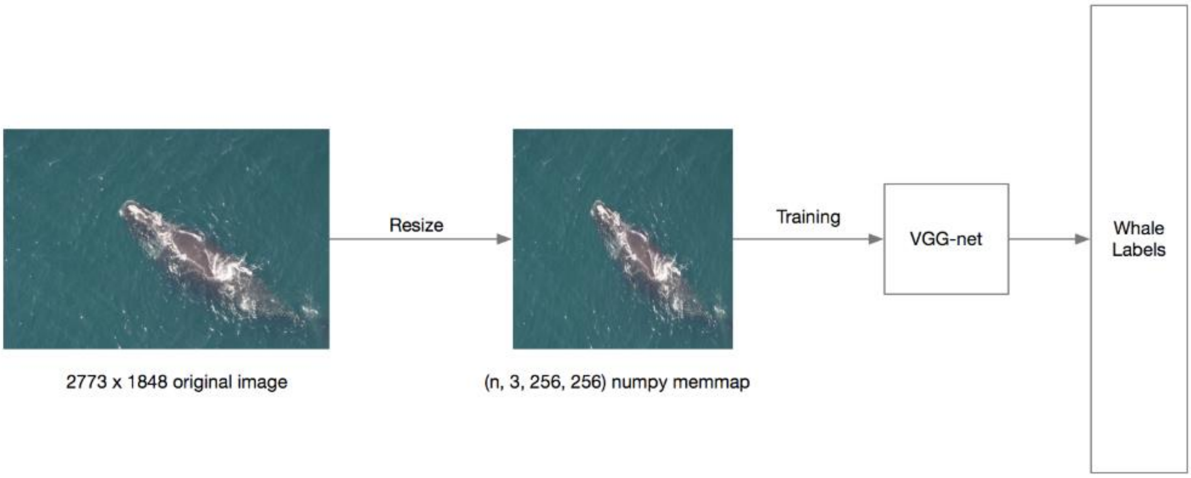

Figure 4.6: Initial Approach

In figure 4.9 actually no special image processing is performed. We have an original of 2773 x 1848, we can't load such a big image into our neural network. I will be millions of features and nearly impossible to handle with the current state of the art hardware. To reduce computational difficulty we have reduced size of the image and store in 256 x 256 numpy array named memmap. We have tried to protect aspect ratio during resize. The architecture of network is based on OxfordNet (VGG-net).

4.6.1.1 Oxford Net

OxfordNet or VGG-net is convolutional neural network architecture. It is one of the most widely used convolutional neural networks. VGG has won Imagenet ILSVR competition in 2014 by a significant margin. It is constructed by adding 3x3 convolution filters followed by a layer of max pooling. 3x3 convolution and 2x2 pooling is used throughout the network. As other neural networks, increasing depth of VGG increasing accuracy but more computational power is needed.

Heavy data augmentation was done for Right Whales training including translation, rotation and shearing and scaling. As we have seen images are of different contrast brightness augmentation was added to take care for underexposed and overexposed images. I have also tried color permutation, but it did not affect my results because it can be seen from dataset color does not differ a lot. Whales almost are of same color. Color permutation effects if we are doing classification of different species like cats, dogs, bus, train etc.

Very leaky rectified linear unit (VLeakyReLU) is used for all the architectures.

This approach yielded a validation score of 5.6 logloss. The lower the value of logloss is means the better our results are. It is actually a measure of uncertainty in our architecture. logloss or logarithm loss is used to measure classification accuracy of a data set. It is used when the classifier is outputting probabilities as the likely of class match. To sum up it is actually a cross entropy between predictions and distribution of true labels. If we minimize cross entropy accuracy of our convolutional neural network is increased.

To check why my results are poor at start, I used saliency map, I traverse the map thoroughly on the image and check the differences, then I find out deep convolutional learning was concentrating more on ocean waves than whales. Just for checking I also trained neural network with 512 x 512 numpy array and results remain the almost same just the slight variations.

4.6.2 Applying bounding box localization

Figure 4.7: Bounding Box 1

In Figure 4.10, I have also checked through saliency map, my neural network is not only focusing on whales, so I have decided to localize the whale head. As we know whale head consists of ¼ of whale's body approximately.

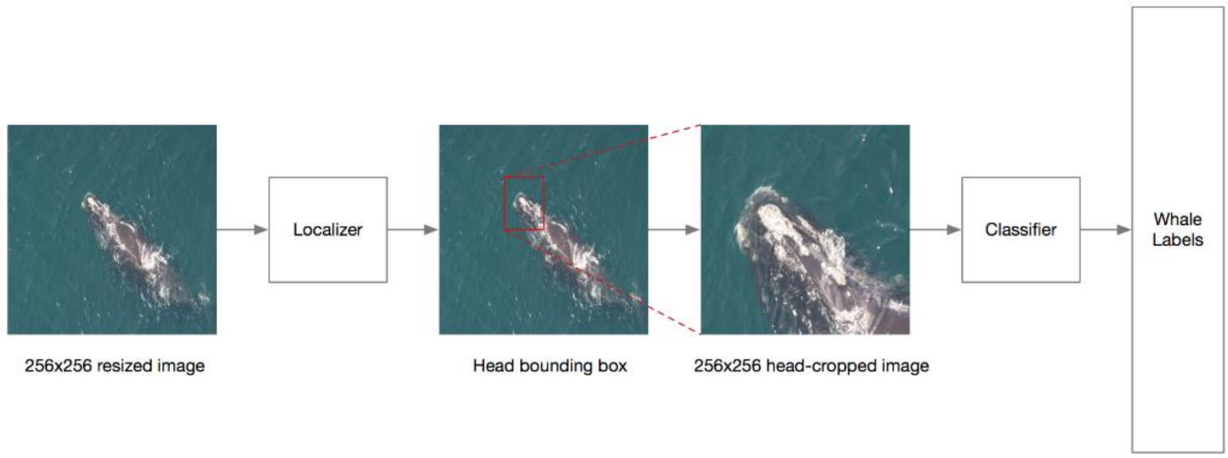

Figure 4.8: Head Localization Approach

In figure 4.11 now Resized image is feed into localization CNN, here bounding boxes will be drawn across the whale's head to separate the area of interest, that is the callosity pattern, so that in this way CNN will only focus the desired parameters. Now are results will be improved and logloss(less the better) will be reduced. Now our Classifier will be trained on these cropped images of head.

4.6.3 Head Localizer

For our ease, we have treated this classification problem as regression problem, with the objective of reducing the mean square error between actual and predicted bounding boxes .Just like baseline model the network Localizer architecture is trained at Oxfordnet(VGG.net). To improve our accuracy data augmentation was applied to the dataset as well.

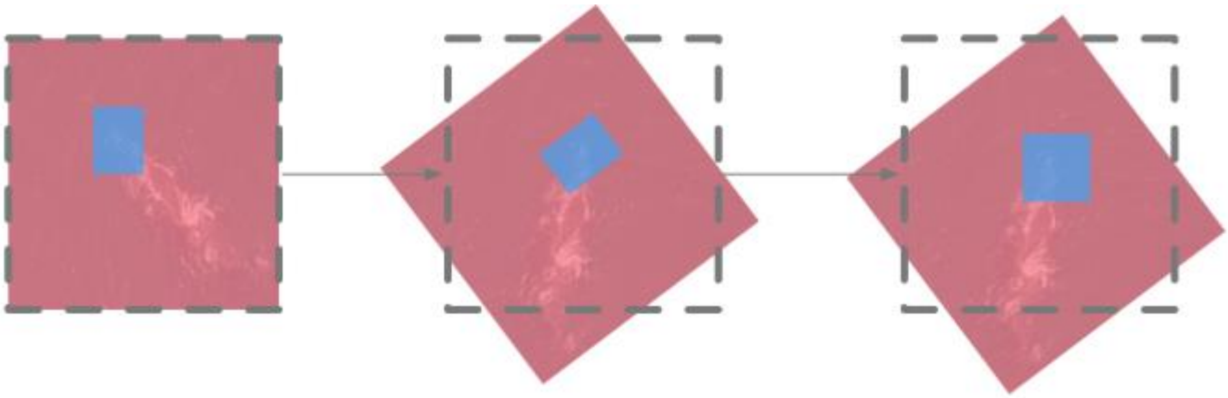

Figure 4.9: Bounding box mask 1

In Figure first square shows bounding box mask, then the whale is rotated so that we can localize head in best possible manner. The training took a long time, mean square error convergence was slow, I think using another loss function will be much better.

Figure 4.10: Difficulties in Bounding Box

In Figure 4.13 here are some more scenarios of ground truth Bounding Box and predicted bounding box. The first case leads to overestimation that is predicted box is greater than bounding box. The 2nd figure shows the example of underestimation, this is also not good for our results, here predicted box is less than ground truth bounding box. Similarly in the third figure, we have partial overlap, here prediction is also not up to the mark.

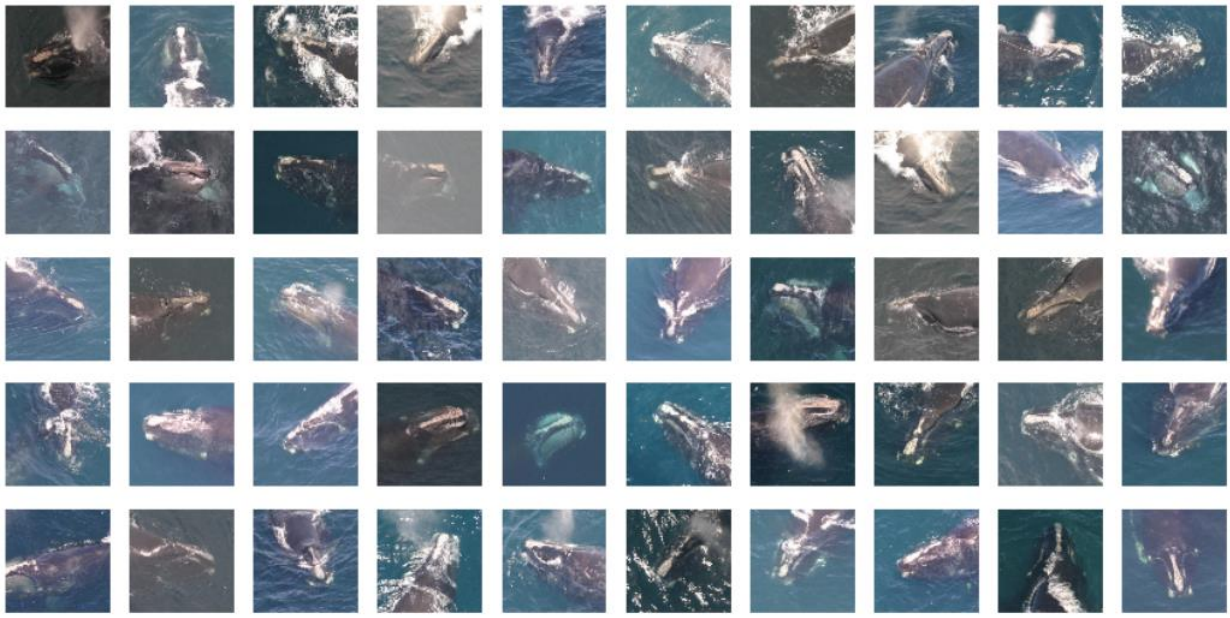

Figure 4.11: Head Localization Result

In Figure 4.15 these are some results of head localization; the results are surprisingly more than expected. As you can see we can clearly identifies whale's head in majority of the images. These cropped images of Right Whales are now feed into classifier.

4.6.4 Classifier

Classifier for Right Whales recognition is also OxfordNet architecture. It is trained on 256 x 256 cropped images of head, which are shown above and are extracted from head localization architecture. Hence our results were improved but not a significant or remarkable improvement. We have achieved a validation score of approximately 5 which was much better than previous naïve approach. I have also tried by adding padding across predicted and bounding boxes for the purpose of increasing accuracy but no effect deflected on results.

4.7 Whale head Alignment

After experimenting lot of techniques, at this point was clear that, the more we intent our classifier to focus on correct features of whale, the more accurate results we are expected to get. First we just cropped the images and feed them into classifier, that was barely good than a random guess off course it is not expectable. Then we proceed closer to our solution we draw bounding boxes around the whale's head and helping our classifier to focus more on whale's head and our region of interest the callosity pattern's on Right Whales head. The approach brings us to closer to the solution. Our accuracy was increased but not to a remarkable state. So we decided to align the callosity pattern. Our approach was to align the coordinates of two vital points Bonnet and Blowhead(callosity pattern lies between this region).

We have Rotated the whale in such a way, that Bonnet is on right and Blowhead is on Left. The head cropped image was obtained by applying affine transform, now in all the images Bonnet point of Right Whales is pointing to the east.

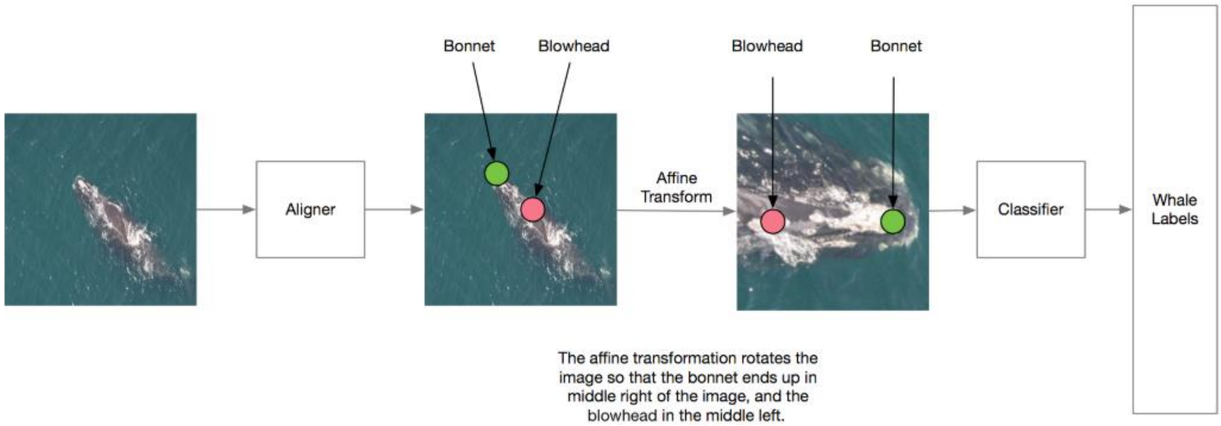

Figure 4.12: Whale Alignment

In Figure 4.16 here classifier will get accurately head cropped and perfectly align images so it can now completely focus on the desired features. Here classifier is at ease, it no longer need to learn features that are invariant to extreme translation and rotation. All other hurdles was crossed only one thing is left, we can't forget that these normalized images were not normalized by exposure, contrast and camera prospective.

This approach is slightly similar to the approach of Facebook's Deepface paper. Deep Face was human facial recognition system and it uses nonlinear approach of 3D formalization to the input image that is the face in this system before feeding it to convolutional neural network. We can't apply nonlinear approach of 3D formalization to our Right Whales data set because here we have only two points Bonnet and Blowhead here. If we have more points we can go for nonlinear approach, that will certainly effect accuracy by good margin.

4.7.1 Aligner

The aligner was again our all-time favorite OxfordNet architecture. It's primary objective was to align the x, y coordinates of Bonnet and Blowhead.Batch Normalization was applied to refine the accuracy. Two rows of fully connected layers was replaced by Global Averaging layer similar max pooling is replaced by stride= 2 convolution. A large scale of data augmentation was applied to increase accuracy and to prevent over fitting. Data

Augmentation includes like rotation, translation and brightness adjustment. To do these we simply used affline transformation to the matrix. Similarly applying test time augmentation also help in improving our Right Whale classification by a significant factor. Mean Square error objective function was used here as training.

4.7.2 Final Classifier

The primary classifier used in this approach is a similar 19 layers OxfordNet(VGG-net) architecture. Global averaging layer was used with stride=2 as max pooling. Other significant difference was that minimal augmentation was applied to both Testing and Training Images. So heavy augmentation prevents the network to converge before time and to cater over fitting lighter augmentation is used. Considering the same factors we did not apply test time augmentation to the Right Whale final classifier.

Remember the mistake I have mentioned in the start of my approach about making a local validation dataset. Actually I have selected 25 % of the images for Aligner and 20 % of images for our final Classifier. This explained that some testing images on validation was part of aligner training and vice versa. Actually it was good for our results and they have shown a remarkable improvement, so it was included it is good practice to split the dataset at start and store them save separately.

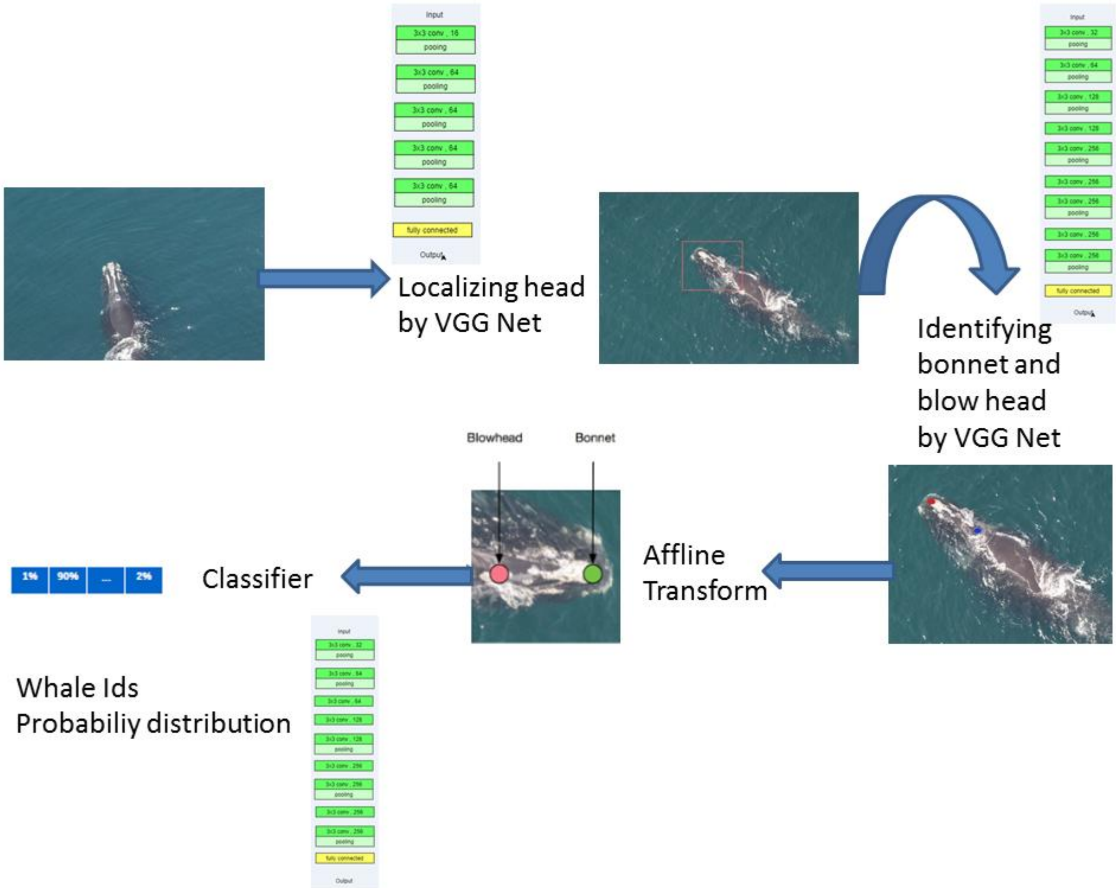

Figure 4.13: Complete System Diagram 1

Figure 4.17 above shows complete system diagram explaining all the processes in details .. In the first step we localizing head by VGG.net then Bonnet and Blowhead was diagnosed, after that we applied affline transform to rotate image in our desired direction. At this stage all the pre-processing of our dataset is done. Image is ready to be fed into classifier. The final setup is yielding us Whale Ids Probability Distribution, the results clear demonstrates, the whale having highest probability is the true whale.

4.8 Some Random Experiments

A lot of random experiments are done, some for testing and some for improving results. . When we first applied CNN on raw images, our accuracy was about 8%, it was focusing on non-discriminative features i.e the ocean waves. Applying CNN on the area of interest i.e callosity pattern yield an accuracy of 78.7%. Then we started trying different state of the art CNN architectures.

4.8.1 Deep Residual Networks

The success of deep learning is highly depending on the depth of Convolutional Neural Network and due to its highly nonlinear nature. However ResNet does not follow conventional phenomena of adding more layers, adding more layers in ResNet sometimes increases training error. In ILSVRC competition last year, a 200 layer ResNet has won first prize in competition. According to my perception ResNet will be a great boom in the field of Machine learning after AlexNet in 2012.

The First ResNet Architecture I tried on my Right Whale Challenge is CIFAR's 10 architecture. There are several factors for using CIFAR's 10 architecture, as our images was already aligned and callosity patterns were localized so a highly non-linear deep architecture was not needed. I choose a network with n=3, with 19 layers and 9 shortcut layers. My ResNet approach does not yield results as expected. Just for experimenting, I also tried 50 layer ResNet on my Right Whale challenge, this network was much more deeper nonlinear and complex so it resulted in over fitting. Then I used dropout technique and it shows over fitting have been reduced significantly. Using a dropout value greater than 0.5 for example like 0.7 was giving good results.

4.8.2 Inception V3

After having some Success with ResNet on my Right Whale Dataset, I have decided to test on Inception V3. Nothing was changed in original architecture, I introduced pooling. After start of training it quickly do over fitting. Over fitting was not decreased by even increasing dropout. I figured out Inception was not good at this type of dataset.

CHAPTER 5: EXPERIMENTAL RESULTS

In the previous section I have described in detail the methodology I have used and other techniques that were testing and not used for your final results. Different Techniques was tested, I have learned a lot form it, specially learned which of the methods are less useful and not to be used again. As we all know the challenge was hosted by Kaggle and data set was less and improper images and all other difficulties. There were 4544 images of Right whales in training set and whale IDs and images to these respective whale IDs was provided. Similarly 6925 images were given in testing set. In order to test our approaches, we separated 10 % of images from testing set. 10% of 4544 is about 452 so now 4092 images are remained in testing set , training set count is further reduced.

We also knows that number of whales varies a lot. There is a whale with 46 training images and 24 whales with just a single image and 29 whales with just 2 images. This uneven distribution was one of the major challenges for us. The average number of images in each training class was 12 Right whales. Training each of localization networks took about 9 hours each and final classifier took around 24 hours.

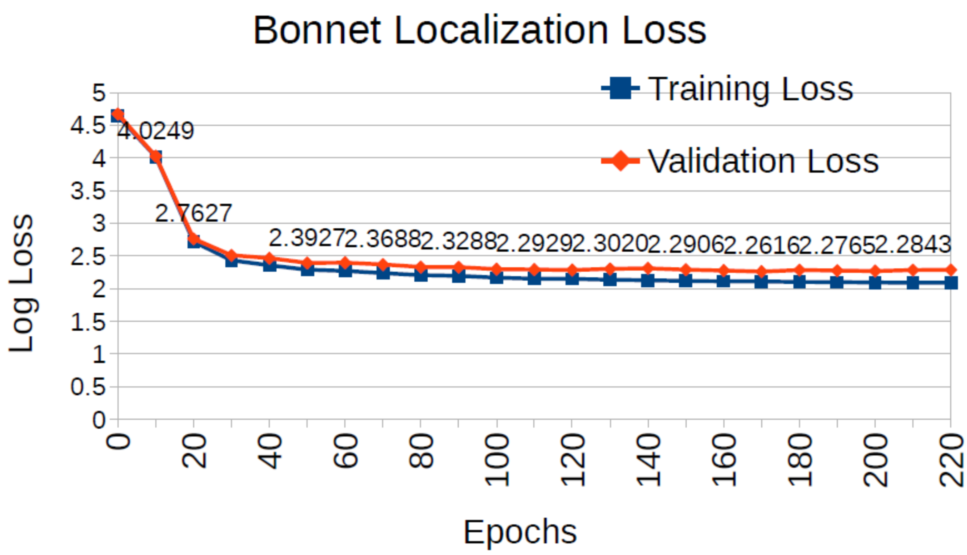

Figure 5.1: Bonnet Localization

Figure 5.1 shows Blue Lines represents the training curve, how Log Loss decreases as the number of Epochs increasing. The graph shows training and testing curve for training of bonnet. Bonnet is the point along which the callosity pattern lies. The best result that is achieved after 220 epochs in 2.83. The lower the value of log loss the better our results are.

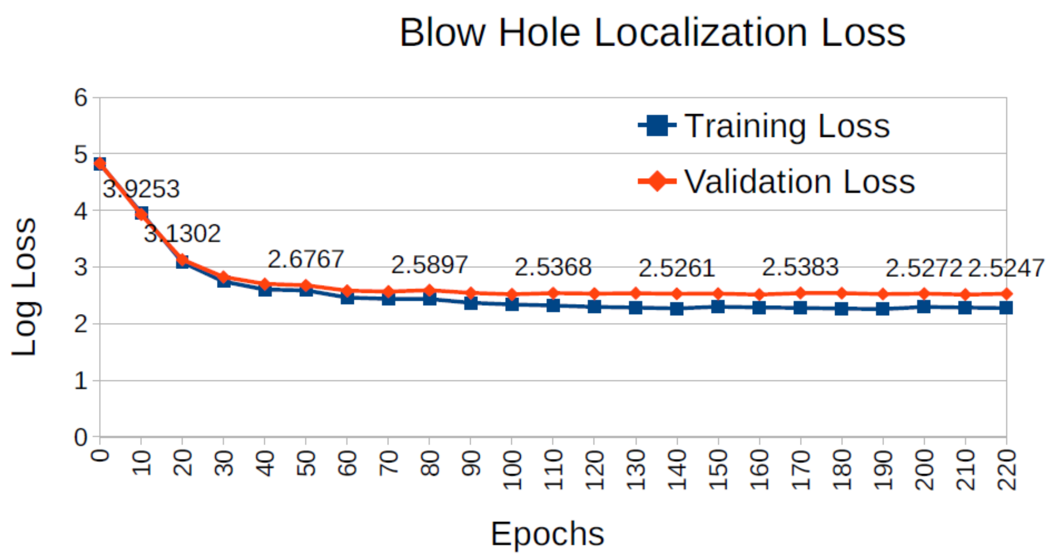

Figure 5.2: Blow Hole Localization

Figure 5.2 shows training and testing graphs for another important feature blowhole . Here we have achieved better result than bonnet , after 220 epochs result achieved is 2.52 . This is the 2nd point on which callosity patterns ends .

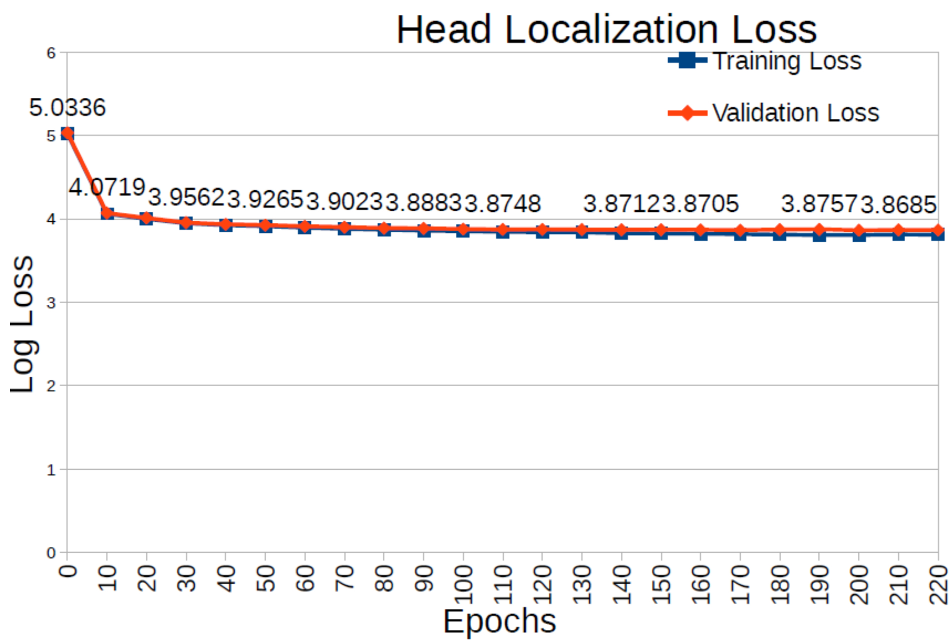

Figure 5.3: Head Localization

In Figure 5.3 we have trained our classifier on passport size photos of whales. That is the callosity pattern that lies between two points bonnet and blowhead. The best training accuracy achieved is 3.8 for this one after 220 epochs. The result remains almost constant after 100 epochs but we increased epochs to check that our results are over fitting or not.

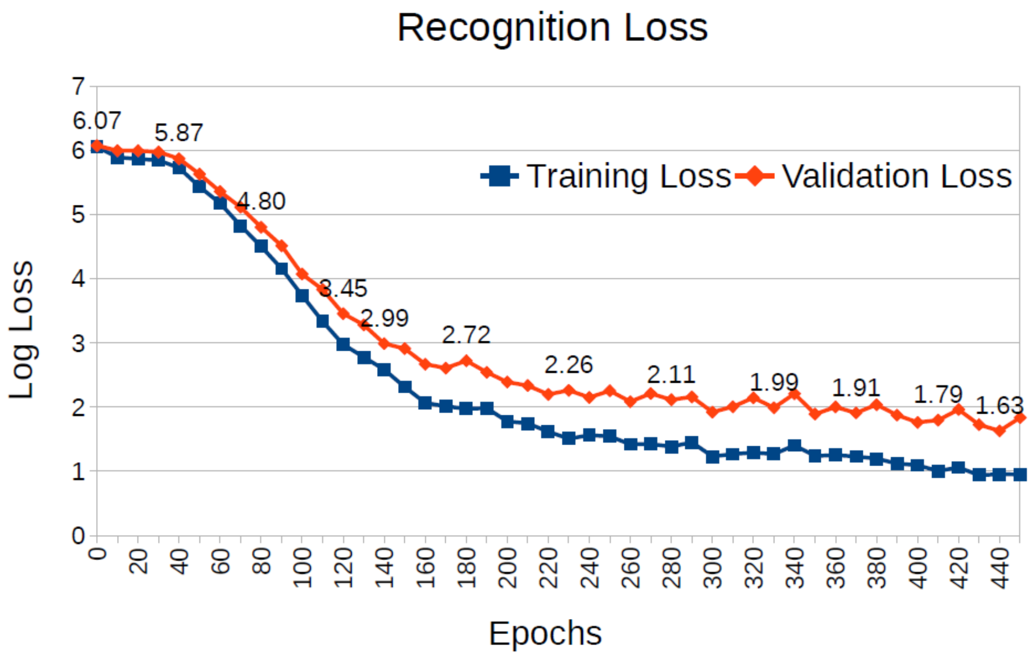

Figure 5.4: Recognition Loss

Figure 5.4 shows the overall accuracy that we have achieved. Log loss value of 1.63 is achieved after 440 epochs. Training took approximately 50 hours of GPU GTX-565. Increasing Epochs was not effecting training loss that remains same.

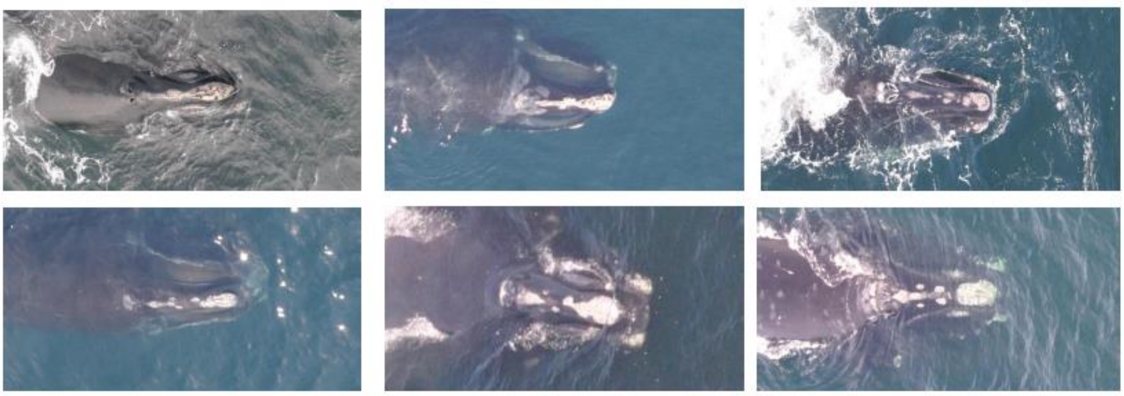

Figure 5.5: Head Alignment

These are some samples of localized whale bodies and heads aligned towards east.

Figure 5.5: Head Alignment

A sample of localized callosity patterns that is the most important feature for us. All the pre-processing of image is done until this point.

Figure 5.6 Flaws in our Dataset

These samples shows where whale head is not localized correctly and hence callosity pattern was not possible to deduct. This is due to poor resolution and splashing effect created by the moment of whales.

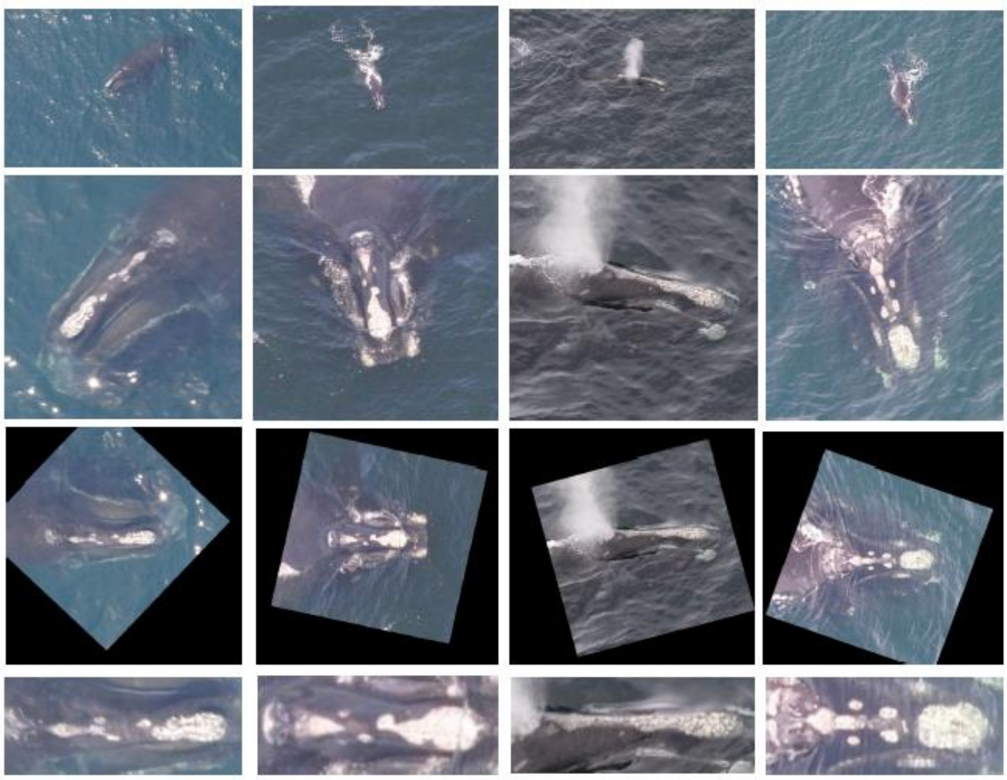

Figure 5.7 All the Inputs

In Figure 5.7 this sample is showing inputs to all the stages. As it can be clearly seen, the last set of images are for final classifier. In second last stage whale head is aligned as desired bonnet on the right and blowhead on the left.

Ranking	Team name	Score
1	Deepsense.io [2]	0.59600
2	Felixaumon [3]	1.07585
3(after competition)	Proposed System	1.63100
4	SKE	1.73145
5	Abdul Wahab [5]	1.95413
6	Tsakalis Kostas	2.14000

Table No 5.8 Team Rankings

The Table above shows rankings of the competition. The table is displaying score of top 6 positions while the number of teams that took part in the competition were 364. This is logloss score the lower the value the better the result is. Our score might be increased if we have more training data, the worst case scenario was for some whales we just have one training image.

CHAPTER 6: CONCLUSION & FUTURE WORK

6.1 Conclusion

This thesis findings will help preserve Right Whales that have been mistakenly caught in fishing gear. As we all know right whales are endangered species as fewer than 500 are left in Atlantic ocean. Data set was small, over fitting was difficult to avoid so we break our problem into different sections and apply convolutional neural networks on each part separately. We also tried some new techniques like dropout and applied heavy data augmentation to boost the accuracy of our results.

With such a sparse distribution, it was really challenging for us to train our deep convolutional neural networks. To minimize the effect of small dataset, we have divided our problem wisely. Instead of training our neural network on whole images, which was of no use to us, because most of the images contains ocean waves. So we first localize head of the whale, in this way we can focus on our desired feature more, that is the callosity pattern on Whale's head. After localizing head we find two points Bonnet and Blowhead, callosity pattern lies between these two points. We then align whale's head in such a way, Bonnet is on right and blowhead is on left, whale's head pointing towards east. Now images are align in this way, our classifier can only focus on area of interest. This has improved our results significantly.

We have also testing how results varies by changing number of layers of a CNN architecture. We have achieved accuracy of 78.7%, I believe that this is remarkable progress even with such a small dataset. Applying PCA also help us to increase accuracy .

6.2 Future Work

We strongly believe that, if we have large data set, things will be much easier and results will be much improved. Increasing the depth of neural network will increase accuracy but that will come at the cost of very powerful computing power. We must try by using different activation functions and monitor how results are varying and error is decreasing or not. Doing any pre-processing on the callosity pattern between Bonnet and Blow head to make it more prominent can also make training process quicker.

REFERENCES

[1] R. Bogucki, "Which whale is it, anyway? face recognition for right whales using deep learning," http://deepsense.io/deep-learning-rightwhale-recognition-kaggle/, [Online; accessed 04-August-2016].

[2] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2):157–166, 1994.

[3] S. D. Kraus et al., "North atlantic right whales in crisis," Science, Vol. 309(5734), pp. 561-562, 2005.

[4] C. M. Bishop. Neural networks for pattern recognition. Oxford university press, 1995.

[5] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In AISTATS, 2010.

[6] A.Krizhevsky. Learning multiple layers of features from tiny images. Tech Report, 2009.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[8] J. N. Matthews, S. Brown, D. Gillespie, M. Johnson, R. McLanaghan, A. Moscrop, D. Nowacek, R. Leaper, T. Lewis, and P. Tyack, "Vocalization rates of the North Atlantic right whale (Eubalaena glacialis)," J. Cet. Res. Man., Vol. 3, pp. 271-282.

[9] S. E. Parks and P. L. Tyack, "Sound production by North Atlantic right whales (Eubalaena glacialis) in surface active groups," J. Acoust. Soc. Amer., Vol. 117, pp. 3297-3306, 2005.

[10]M. Esfahanian, H. Zhuang, N. Erdol and E. Gerstein, "Comparison of two methods for detection of North Atlantic Right Whale upcalls," 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, 2015, pp. 559-563.

[11] C. Khan, P. Duley, A. Henry, J. Gatzke2, and T. Cole1, "North atlantic right whale sighting survey (narwss) and right whale sighting advisory system (rwsas) 2013 results summary," US Dept Commer, Northeast Fisheries Science Center Reference Document, pp. 14–11, 2014

[12] Kaggle, "Right whale recognition," https://www.kaggle.com/c/noaaright-whale-recognition, [Online; accessed 19-January-2016].

[14] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Cognitive modeling, vol. 5, no. 3, p. 1, 1988.

[15]Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[16] M. El-Sakka and A. Kabani, "North atlantic right whale localization and recognition using very deep and leaky neural network," Mathematics for Applications, vol. 5, no. 2, pp. 155–170, 2016.
[17] F. Lau, "Recognizing and localizing endangered right whales with extremely deep neural networks," http://felixlaumon.github.io/2015/01/08/kaggle-right-whale.html, [Online; accessed 04-August-2016].

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv preprint arXiv:1512.03385, 2015

[19] Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 1090–1098 (2012).

This report was very detailed and amazing , that used convolutional neural networks to almost halve the error rate for object recognition, and precipitated the rapid adoption of deep learning by the computer vision and artificial intellegence community.

[20] Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1915–1929 (2013).

[21] Tompson, J., Jain, A., LeCun, Y. & Bregler, C. Joint training of a convolutional network and a graphical model for human pose estimation. In Proc. Advances in Neural Information Processing Systems 27 1799–1807 (2014).

[22] Szegedy, C. et al. Going deeper with convolutions. Preprint at http://arxiv.org/ abs/1409.4842 (2014).

[23] Mikolov, T., Deoras, A., Povey, D., Burget, L. & Cernocky, J. Strategies for training large scale neural network language models. In Proc. Automatic Speech Recognition and Understanding 196–201 (2011).

[24] Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29, 82–97 (2012).

This joint paper from the major speech recognition laboratories, summarizing the breakthrough achieved with deep learning on the task of phonetic classification for automatic speech recognition, was the first major industrial application of deep learning.

[25] Sainath, T., Mohamed, A.-R., Kingsbury, B. & Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proc. Acoustics, Speech and Signal Processing 8614–8618 (2013).

[26] Ciodaro, T., Deva, D., de Seixas, J. & Damazio, D. Online particle detection with neural networks based on topological calorimetry information. J. Phys. Conf. Series 368, 012030 (2012).

[27] Kaggle. Higgs boson machine learning challenge. Kaggle https://www.kaggle. com/c/higgs-boson (2014).

[28] Helmstaedter, M. et al. Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature 500, 168–174 (2013).

[29] Leung, M. K., Xiong, H. Y., Lee, L. J. & Frey, B. J. Deep learning of the tissue-regulated splicing code. Bioinformatics 30, i121–i129 (2014).

[30] Collobert, R., et al. Natural language processing (almost) from scratch. J. Mach. Learn. Res. 12, 2493–2537 (2011).
[31] Bordes, A., Chopra, S. & Weston, J. Question answering with subgraph embeddings. In Proc. Empirical Methods in Natural Language Processing http://arxiv.org/abs/1406.3676v3 (2014).
[32] Jean, S., Cho, K., Memisevic, R. & Bengio, Y. On using very large target vocabulary for neural machine translation. In Proc. ACL-IJCNLP http://arxiv.org/abs/1412.2007 (2015).
[33] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313, pp. 504–507, July 2006.
[34] Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29, 82–97 (2012).
[35] H. Lee, R. B. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," Commun. of ACM, vol. 54, no. 10, pp. 95–103, 2011.
[36]D. Hubel and T. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," Journal of Physiology, vol.160,pp.106–154,1962.
[37] K. Fukushima and S. Miyake, "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position," Pattern Recognition, vol. 15, no. 6, pp. 455–469, 1982.

[38] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropogation applied to handwritten zip code recognition," Neural Computation, vol. 1, no. 4, pp. 541–551, 1989.
[39] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in Intelligent Signal Processing (S. Haykin and B. Kosko, eds.), vol. 86, pp. 306–351, IEEE Press, 2001.

[40] Y. LeCun, C. Cortes, and C. J. C. Burges, "The mnist database for handwrittendigits."http://yann.lecun.com/exdb/mnist/.
[41] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma,

Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “Imagenet large scale visual recognition challenge,” International Journal of Computer Vision (IJCV), pp. 1–42, April 2015.
[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held on December 3-6, 2012, (Lake Tahoe, Nevada, United States), pp. 1106–1114,2012.
[43] G. E. Dahl, T. N. Sainath, and G. E. Hinton, “Improving deep neural networks for LVCSR using rectified linear units and dropout,” in IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP), (Vancouver, British Columbia, Canada), pp. 8609–8613, May 26-31 2013.
[44] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Droupout: A simple way to prevent neural networks from overfitting,” Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014.
[45] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in Computer Vision - ECCV 2014 - 13 European Conference, (Zurich, Switzerland), pp. 818–833, September 6-12 2014.
[46] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-sacle image recognition,” Computing Research Repository (CoRR), vol. abs/1409.1556, 2014.

[47] I. J. Goodfellow, Y. Bulatov, J. Ibraz, S. Arnoud, and V. Shet, “Multidigit number recognition from street view imagery using deep convolutional neural networks,” Computing Research Repository (CoRR), vol. abs/1312.6082, 2013.
[48] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in Intelligent Signal Processing (S. Haykin and B. Kosko, eds.), vol. 86, pp. 306–351, IEEE Press, 2001.
[49] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural

networks," in Proceeding of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS - 11), pp. 315–323, 2011

[50] H. Lee, R. B. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," Commun. of ACM, vol. 54, no. 10, pp. 95–103, 2011.

[51] K. Pylypenko, "Right whale detection using artificial neural network and principal component analysis," 2015 IEEE 35th International Conference on Electronics and Nanotechnology (ELNANO), Kiev, 2015, pp. 370-373. doi: 10.1109/ELNANO.2015.7146912

[52] A. Kabani and M. R. El-Sakka, "Improving Right Whale recognition by fine-tuning alignment and using wide localization network," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, 2017, pp. 1-6. doi: 10.1109/CCECE.2017.7946736

[53] A. Kabani and M. R. El-Sakka, "Improving Right Whale recognition by fine-tuning alignment and using wide localization network," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, 2017, pp. 1-6. doi: 10.1109/CCECE.2017.7946736

[54] D. Wang, L. Zhang, Z. Lu and K. Xu, "Large-Scale Whale Call Classification Using Deep Convolutional Neural Network Architectures," 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Qingdao, 2018, pp. 1-5. doi: 10.1109/ICSPCC.2018.8567758

[55] G. Roberts, A. M. Zoubir and B. Boashash, "Time-frequency classification using a multiple hypotheses test: an application to the classification of humpback whale signals," 1997 IEEE International Conference on Acoustics, Speech, and

Signal Processing, Munich, 1997, pp. 563-566 vol.1. doi: 10.1109/ICASSP.1997.599700

[56] M. Esfahanian, H. Zhuang, N. Erdol and E. Gerstein, "Comparison of two methods for detection of North Atlantic Right Whale upcalls," 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, 2015, pp. 559-563. doi: 10.1109/EUSIPCO.2015.7362445

[57] C. Qiu et al., "Transfer Learning for Small-Scale Fish Image Classification," 2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO), Kobe, 2018, pp. 1-5. doi: 10.1109/OCEANSKOBE.2018.8559473

[58] G. Ding et al., "Fish recognition using convolutional neural network," OCEANS 2017 - Anchorage, Anchorage, AK, 2017, pp. 1-4.