Humpback Whale Identification using CNN and CAM to understand Whale population dynamics

Report

Table of Contents

Table of Figures

Abstract

After centuries of intense whaling, recovering whale populations still have a hard time adapting to warming oceans and struggle to compete every day with the industrial fishing industry for food. In some areas of the Northern Pacific Ocean, Whales are also being hunted for their skin and fat. The Humpback Whale is also currently identified as an endangered species by the International Union for Conservation of Nature. It is important to keep a consensus of these species. Since these are extremely large mammals, it is hard to find mechanisms to count. Oceanographers capture pictures of Whales everytime they surface. These pictures are used to estimate their population per square kilometer. In the interest of making conservation efforts for the Whale population, it is important to understand and analyse their population dynamics, This paper aims to uniquely identify whales for this purpose. The above is carried out using Transfer Learning approaches for classification after preprocessing the images for foreground extraction and visualizing the localisation of feature mappings using Class Activation Maps(CAM). The paper also focuses on the comparative analysis of these transfer learning techniques and the impact that foreground extraction and pre-processing techniques have on image classification and feature localisation for identification of the whale. Current existing methods only focus on achieving higher accuracies with deeper network architectures, for classification while this paper first focuses on Whale fluke localisation using CAM and then uses a Deep Learning model. There is currently no such existing technique that uses Foreground Extraction on whale flukes to get the sea and sky background out of the way. Our Method uses CAM to show that there is a clear difference before and after Foreground Extraction.

Keywords

Image Segmentation, Convolutional Neural Networks, Transfer Learning, Class Activation Maps, Image Enhancement

Introduction

Application

The varying patterns on the tail flukes distinguish individual animals. Identification is done by comparing the amount of white vs black and scars on the fluke. The humpback whales are then given a catalogue number. A study using data from 1973 to 1998 on whales in the North Atlantic gave researchers detailed information on gestation times, growth rates and calving periods, as well as allowing more accurate population predictions by simulating the mark-release-recapture technique. A photographic catalogue of all known North Atlantic whales was developed over this period and is maintained by the College of the Atlantic. Similar photographic identification projects operate around the world.

Vessel-based and aerial sighting surveys have been the most used methods to count whales. The information gathered from this fieldwork is used as the basis for population modelling which produces an abundance estimate. Happy Whale is a platform that uses image processing algorithms to let anyone submit their whale photo and have it automatically identified, and this was available as a Competition on Kaggle.

Objective

To aid whale conservation efforts, scientists have been using photo surveillance systems and manually identifying whales based on the shape of their tails and unique markings found in footage. With Image Processing techniques now at our disposal, Identification of Unique Whales has become much easier.

Proposed Technique

Our paper aims to use Image Enhancement and segmentation techniques and train a model that accurately classifies whales into their respective classes. We have tried three different approaches on four different Transfer Learning Models respectively and made a comparative study to conclude which model and which approach works best.

1) Class Activation Maps

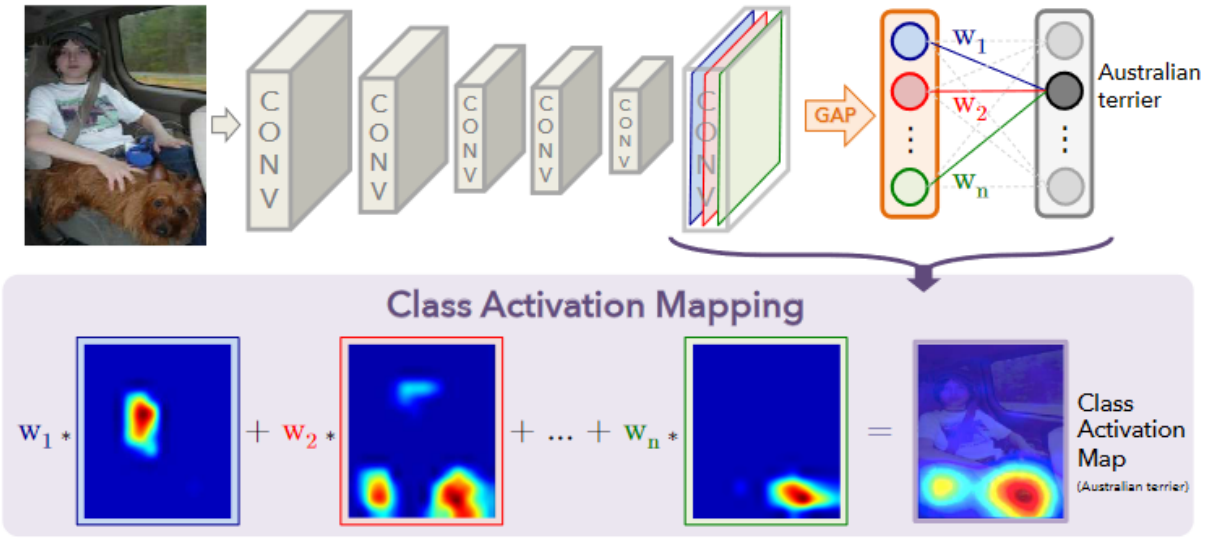


Fig1. CAM highlighting class specific discriminative regions[10].

While performing classification, there are chances that the models recognise the wrong part of the image for classification. The background, in this case, the sea and sky, which is irrelevant

sometimes contributes to the classification. So that is why we found it important to remove the background. But before doing that, we must first analyze what parts of the image are responsible for classifying it into that particular class. Class Activation Maps allow us to do so.
The predicted class score is mapped to the previous convolutional layer. As illustrated in Fig.1, from the last convolutional layer global average pooling outputs a spatial average of the feature map of each unit. A weighted sum of these spatial feature values is used to generate the final activation map. Given an image, $f_x(x, y)$ represents a activation unit k from the last layer of the

network architecture. The global average pooling function is $F_x = \sum_{x,y} f_x(x, y)$.

The final weighted average sum of the activations are $\sum_{x,y} w_c^x F_x$ where w_c^k is the weight.is the weight corresponding to class c for the particular node k. We define as the class activation map for class c, where each spatial activation mapped output is given by

$$\sum_k w_k^c f_k(x, y)$$

2) RESNET50

ResNet is a short name for Residual Network. As the name of the network indicates, the new terminology that this network introduces is residual learning.In residual learning, instead of trying to learn some features, we try to learn some residual. Residual can be simply understood as subtraction of features learned from input of that layer. ResNet does this using shortcut connections (directly connecting input of the nth layer to some (n+x)th layer. It has proved that training this form of networks is easier than training simple deep convolutional neural networks and also the problem of degrading accuracy is resolved. ResNet50 is a 50 layer Residual Network.

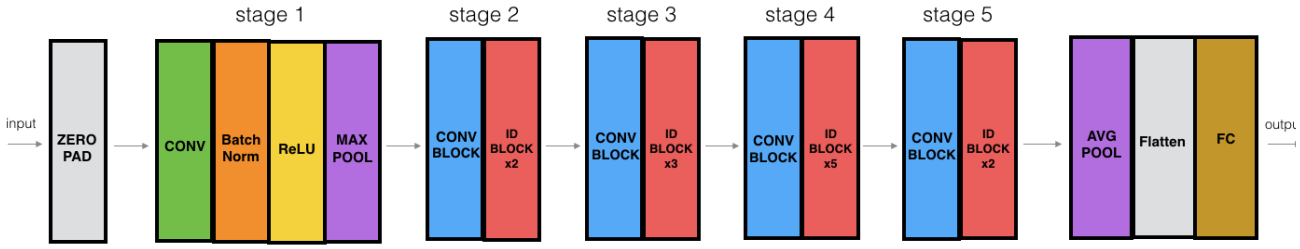


Fig2. ResNet Schematic diagram.

3) VGG16.

VGG16 (also called OxfordNet) is a convolutional neural network architecture named after the Visual Geometry Group from Oxford.VGG-16 is a convolutional neural network consisting of 16

layers. The model loads a set of weights pre-trained on ImageNet. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes.The default input size for VGG16 model is 224 x 224 pixels with 3 channels for RGB image. It has convolution layers of 3x3 filter with a stride 1 and maxpool layer of 2x2 filter of stride 2.

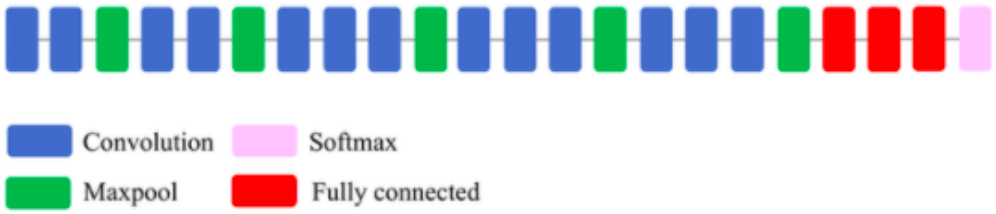


Fig3. VGG16 Schematic diagram[11].

4) InceptionResNetV2

Inception-ResNet-v2 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 164 layers deep and can classify images into 1000 object categories, such as the keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 299-by-299, and the output is a list of estimated class probabilities.

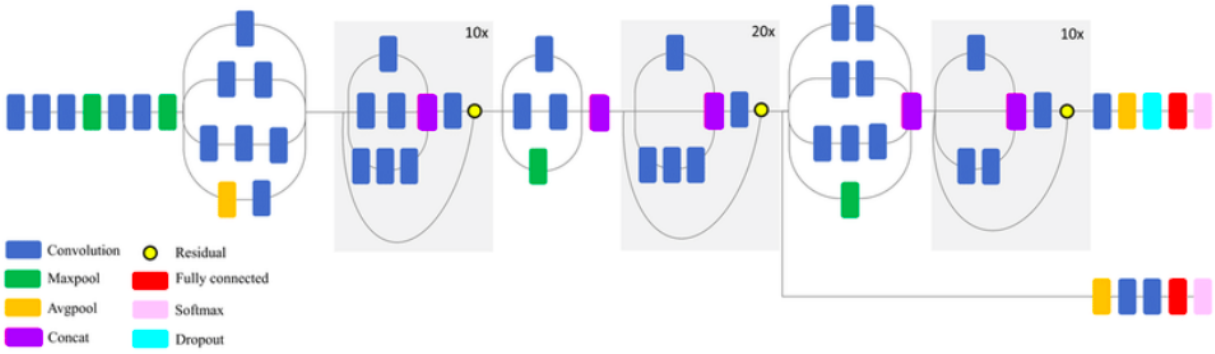


Fig4. InceptionResNet Schematic diagram

5) VGG19

VGG-19 is a trained Convolutional Neural Network, from Visual Geometry Group, Department of Engineering Science, University of Oxford. To reduce the number of parameters in such deep networks, it uses small 3×3 filters in all convolutional layers and is best utilized with its 7.3% error rate.

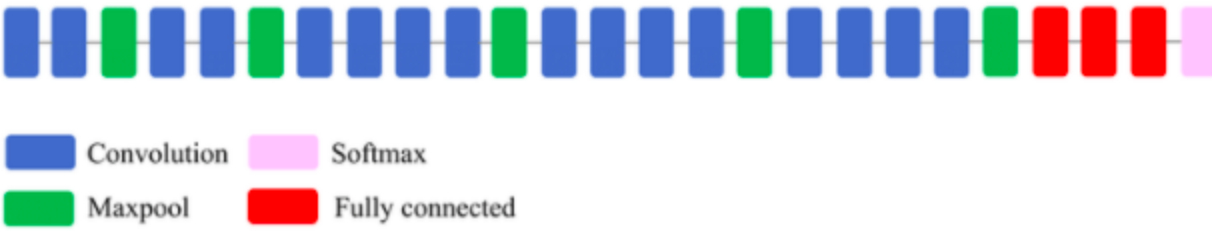


Fig5. VGG19 Schematic diagram

Literature Review

The algorithm proposed in **[1]** includes two new methods of clustering DHFCM and DHKM; they use simple and easy techniques, such as DWT and reducing the color palette that provides the robustness needed in the segmentation and classification of the dorsal fin to the photo-id. The photo-identification system on mobile devices is an alternative portability for researchers in the field to obtain a quick way to identify blue whales in their habitat; this system represents a double tool to assist the process of photo-identification as it can be run from a computer or of a mobile device. The new proposals DHFCM and DHKM due to its easy operation and preprocessing of the images obtained in the habitat of the blue whale are a feasible application for mobile devices, where some processes that run on mobile devices come to be limited by the equity of these, such as the battery and memory. Finally, the proposal App offers a real-time solution to the blue whale photo-identification using a mobile device such as a portable computer.The results presented here demonstrate that the present proposal is 31% higher for results in and 15% for results in.The results obtained by other methods' precision photo-identification ranging from 55% accuracy in identifying pink dolphin and 97% accuracy in identifying sperm are discussed. It does not capture complex features of the whale although it might be efficient in terms of time and is not as efficient in the task. Due to the low computation cost for mobile devices they did not perform segmentation at a pixel level as DL algorithms might be able to do it.

Since deep convolutional neural networks (CNNs) are achieving great performance in several computer vision tasks, Guirado**[2]** et al. propose a robust and generalizable CNN-based system for automatically detecting and counting whales in satellite and aerial images based on open data and tools. In particular, we designed a two-step whale counting approach, where the first CNN finds the input images with whale presence, and the second CNN locates and counts each whale in those images.

Deep learning methods, particularly Convolutional Neural Networks (CNNs), could help in this sense since CNNs are already outperforming humans in visual tasks such as image classification and object detection. CNN models have the capacity to automatically learn the distinctive features of different object classes from a large number of annotated images to later make correct predictions on new images.

The analysis of the first step CNN-based model on ten marine mammal hotspots for whale watching confirmed the presence of whales in six of the ten assessed whale watching hotspots. Then these hotspots are further analysed based on a CNN model which attains a high accuracy.A test of the system on Google Earth images in ten global whale-watching hotspots achieved a performance (F1-measure) of 81% in detecting and 94% in counting whales. Combining these two steps increased accuracy by 36% compared to a baseline detection model alone.Satellite and aerial based assessment can complement and be compared with other aerial, marine, and land observations. The coastal images of Google Earth at zoom 18 that we used correspond to a visual altitude of ~254 m, similar to the aerial surveys for grey whales, and up to ~4 km offshore the

coast, the maximum distance for whale visual surveys from land.20.58% of test grid cells containing whales were misclassified as water (19.11%) or ships (1.47%). A small number of water + submerged rocks and ship images were misclassified as whales.Whales behaviour affects the performance of the first step CNN-based model for detecting the presence of whales. Higher detectability (greater than 90% of true positives) was obtained for the following whale postures: blowing, breaching, peduncle, and logging.

Paper **[3]** presents a study and implementation of a convolutional neural network to identify and recognize humpback whale specimens by processing their tails patterns .Humpback whales have patterns of black and white pigmentation and scars on the underside of their tails that are unique to each whale, just as fingerprints are to humans. Each whale has a unique recognition pattern on their tail. Thus an approximate measure can be found , per area This work collects datasets of composed images of whale tails, then trains a neural network by analyzing and pre-processing images using TensorFlow and Keras frameworks. This paper focuses on an identification problem, that is, since it is an identification challenge, each whale is a separate class and whales were photographed multiple times and one attempts to identify a whale class in the testing set.In the image processing phase, the image is converted to a grey-scale image, ranging from three different channel colors to a single one. There is a double reason behind this conversion—the existence of original images in the black and white dataset and the absence of color characteristics and information. Each pixel has a value ranging between 0 and 255. This amplitude of range, due to the operation of the convolutional networks, allows the incorrect identification of the characteristics for each vector. To correct this disadvantage, it is convenient to normalize the image previously, in a process known as zero mean and unit variance normalization. This paper reports about a network that is not necessarily the best one in terms of accuracy, but this work tries to minimize resources using an image downsampling and a small architecture, interesting for embedded systems. The disadvantage of this paper is that a maximum accuracy of only 75% was obtained, and the triplet loss has not been implemented because of memory constraints.

[4] addresses the problem of identifying individual animals in images based on extracting and matching contours, focusing in particular on the trailing edges of humpback whale flukes and the outline of the ears of African savanna elephants.“Photo-identification” of animals using patterns of stripes, spots or texture, appearance of faces, and body outlines is gaining lots of importance as a potential replacement for capture-mark-recapture techniques, which are extremely costly, labor intensive, and very often dangerous. A coarse-grained FCNN is learned to isolate the contour in an image, and a fine-grained FCNN is learned to provide more precise boundary information. The latter is trained by generating synthetic boundaries from coarse, easily-extracted training data, avoiding cumbersome manual effort. An A* algorithm extracts the final contour, which is converted to a set of digital curvature descriptors and matched against a database of descriptors using local-naive Bayes nearest neighbors. It was found that using the

learned fine-grained FCNN produces more accurate contours than using image gradients for fine localization, especially for elephant ears where the boundaries are primarily texture. Matching using contours extracted using the fine-grained FCNN improves top-1 accuracy from 80% to 85% for flukes and 78% to 84% for ears.The proposed algorithm for contour extraction was then evaluated in the context of its ability to accurately represent the contour and of its effect on the accuracy of the rankings produced by a matching algorithm.
The Advantages of this paper was that the model captures boundary information from transitions in color and texture as well as intensity. The model has been integrated into a complete contour extraction algorithm that also includes a coarse - grained contour model and an A* search algorithm.Approximately 5% improvement rate was found.

[5] proposes a SAR image segmentation method with the optimal level sets using chaotic whale optimization algorithm (CWOA). First, various image segmentation results are obtained by the multi- texture-based model with several random sets of weighting parameters, then, samples are then automatically generated by comparing these segmentation results. In the second step, search agents ,humpback whales in this case are defined with respect to the weighting parameters that need to be optimized, and the fitness function (prey) is associated with the samples.In this model, edge feature is obtained by the modified ratio of exponentially weighted averages (ROEWA) operator, and region information is obtained by the IESE. IESE is a highly parametric model that can model statistical properties of data sets from various types of regions in a SAR image. In addition to these two operators, a boundary regularization operator that ensures the smoothness of the segmentation boundary and a penalty operator that preserves level sets as signed distance functions are both integrated to the framework. The optimal level sets are then established by integrating the multi-texture-based model and CWOA. Finally, the experimental result achieved from a SAR image shows the effectiveness of the proposed method.The experiment on a SAR image of two-region is carried out. In comparison to the model with the default and the empirical values of the weighting parameters, it is visually observed that the model with the optimal level sets attained by CWOA exhibits the best performance, which qualitatively verifies the flexibility and the effectiveness of the proposed method. It is noted that the proposed method also can be implemented on the segmentation of a SAR image composed of multiple regions.

In **[6]**, several techniques that use machine learning and pattern recognition methods are described to recognize wild animal images, which has gained less attention from the community. The concept of recognition of objects based on variations in image content has gained attention over several decades now, and has lately received an increased interest due to the advance of deep learning techniques.
These learning techniques have been successfully applied to many applications such as human face recognition, object recognition, handwritten character recognition, and medical image recognition. The use of deep learning to learn from large datasets has led to the evolution of deep architectures like AlexNet, GoogleNet and Residual Networks (ResNets).

Final feature vectors from both BOW and HOG-BOW are fed into the regularized linear L2-SVM classifier which predicts the class labels of the Wild-Animal images. In a linear multi-class SVM, the output zk(x) of the k-th class is computed.
The pre-trained GoogleNet and AlexNet architectures perform exceptionally well, but being trained on ImageNet that contains the same classes, but different images.The paper has also been able to demonstrate that the reduction in the number of neurons in the last inception layer of the GoogleNet and fully connected layers in AlexNet have shown to be competitive when compared to the original GoogleNet and AlexNet architectures.The paper also reports that the effect of color on BOW with the max-pooling strategy is relatively competitive compared to the AlexNet architecture when trained from scratch. Finally, the BOW technique outperforms the HOG-BOW method. Future work should involve the application of segmentation and data augmentation techniques on our dataset. We also want to study the effect of different color spaces using deep learning architectures.

Dolphin identification techniques proposed in **[7]** are very similar and are easily extrapolated to the domain of whale identification. It discusses many well documented photo identification methods for fin pattern extraction (ICP), Pigmentations of fin analysis (DFP) and matching algorithm (SVM). The paper proposes a deep learning based dolphin identification algorithm using an advanced technique called hybrid saliency method, for feature extraction and to efficiently integrate several well-known techniques to make dolphins distinguishable. The proposed techniques are used to separate the region of interest from the background (e.g. the sea water) to improve the accuracy of the identification results, which is usually a problem of most convolutional neural network based methods.
The team performed an analysis of the pixels in the raw image set by calculating the gradient of ground truth score per pixel and found the waves in the sea surface had a very high impact on the ground truth score. To overcome interference from the sea surface the image was multiplied with a mask on the dolphin to remove pixels of the sea surface. The mask is calculated using an average of the outputs calculated by 4 algorithms: SODM, NLDF, GBVS and Saliency-HDCT. Clear improvement in classification accuracy over training classifiers over unmasked images was observed.The calculated mask is also able to outperform the mask generated by any one of the contributing algorithms. More work can be done on working out what algorithms go into computing the saliency map used to generate the mask.

Identification of freshwater fish is a similarly complex task comparable to whale identification, Challenges in the image include damage of the freshwater fish body and the complex processing environment such as the illumination, hardware equipment, visibility and noise, the images for target detection appear blurred, damaged, and degraded, and the difficulty of target detection has increased significantly. The traditional method of target detection includes several steps such as selecting regions, extracting features, classifier classification, and positioning targets. The main feature extraction methods based on learning methods include edge detection and threshold

segmentation, such as the global threshold segmentation method of maximizing the variance between classes and the global edge adaptive smoothing filter algorithm of Canny operator in edge extraction The features identified and used for freshwater fish included texture, color, shape **[8]** discusses how to deal with the problems of poor image quality, inadequate data samples, model overfitting, continuous network down sampling, low recognition and segmentation accuracy. To achieve this, a freshwater fish body semantic segmentation model of the confrontation network is proposed. Using the generation confrontation network to generate heterogeneous target images, enrich the available image information, and expand the training dataset. Advantages are that it can extract more information from limited available training data thanks to use of Generative Adversarial Network.Because there is more information available and more targets to identify, segmentation performance also improves.

Risso's Dolphins are a relatively unknown species of dolphins. Photo-identification studies based on the recognition of single individuals through specific markers on their body can help to evaluate the abundance of cetaceans, providing relevant biological information usable for marine environment protection and also prove to be useful in better studying the species.
Commonly, adult Risso's dolphins display extensive white scarring on their bodies, solid grey at birth . These scars are thought to be caused by intraspecific interactions. These appear as scratches, stains, or circular marks, and in some animals can cover most of their body surface. As a result, the unique markings on their dorsal fin can be successfully analyzed to identify single individuals.
In **[9]**, the algorithm used for the automated photo-identification of Risso's dolphins is SPIR (Smart Photo-identification of Risso's dolphin). Scale Invariant Feature Transform (SIFT) is applied to detect the key-points over the scars on the dorsal fin of Risso's dolphins. The photo ID tool then uses these SIFT features to classify these dolphins as the dolphin with the most similar features that has already been recorded.
The identification strategy can be broadly classified into 3 steps: fin segmentation (fin mask creation), feature extraction inside the mask, and RUSBoost classification (to deal with imbalanced classes and lesser samples). The authors propose a novel method named NNPool to recognize known vs unknown dolphins without explicitly identifying the name of the known and previously labeled dolphin. An important advantage of NNPool is that it analyzes the full fin images, thus its performance is independent from the accuracy of the segmentation of the fin.RUSPool is able to recognize the unknown dolphins with an accuracy = 78%, sensitivity =58% and specificity = 81%. The only Disadvantage is that, to work at full potential NNPool needs to gather large amounts of dolphin images.

[12] presents a framework for discriminative localization, which helps shed some light into the decision-making of Convolutional Neural Networks. The refined CAMs were produced by combining the original, low-res maps with the ones produced by a network called the expansion network. The latter was trained to generate high resolution maps, which can provide more

fine-grained detail than their lower-resolution counterparts. By combining both, the refined CAMs were found to be much more robust and can work even in cases where either one of their components fail. Future work will be directed towards jointly training the classification and expansion networks, further refining the post processing procedure and investigating the reasons for failure of each of the two types of CAMs.
The aim in **[13]** is to estimate foreground pixel colours without colours bleeding in from the background of the source image. Such bleeding can occur with Bayes matting because of the probabilistic algorithm used which aims to strip the background component from mixed pixels but cannot do so precisely. The residue of the stripping process can show up as colour bleeding. Here we avoid this by stealing pixels from the foreground TF itself. First the Bayesmatte algorithm is applied to obtain an estimate of foreground colour on a pixel . Then, from the neighbourhood as defined above, the pixel colour that is most similar to the foreground color is stolen to form the foreground colour. Finally, the combined results of border matting, using both regularised alpha computation and foreground pixel stealing are obtained.

Paper **[14]** Comparison of Two Computer-Assisted PhotoIdentification Methods Applied to Sperm Whales (Physeter macrocephalus) namely the Highlight method and the Europhlukes method were compared. Performance was measured in terms of speed and accuracy. A test set was constructed containing two photographs of each of 296 individuals. The test set was divided into three classes of photographic quality and three classes of pattern distinctiveness.Limitations of the Study face mainly two problems.Firstly, there was no ground-truth regarding the real number of matches in the test set. Known matched pairs were found by using either the Highlight or the Europhlukes methods.The whole test was executed by one user. We therefore neglected differences in user dependence during the testing. Secondly, the two methods were compared with respect to the matching of photographs.

Proposed Architecture

To highlight the importance of image enhancement techniques, The designed model has been fed with three types of data. In the first scenario, raw data from the Humpback Whale Identification dataset was read as the model's input.

Model Name:

1. Foreground Extraction +Keras Preprocessing [FE + KP]
2. Foreground Extracted [FE]
3. With Keras Preprocessing [KP]

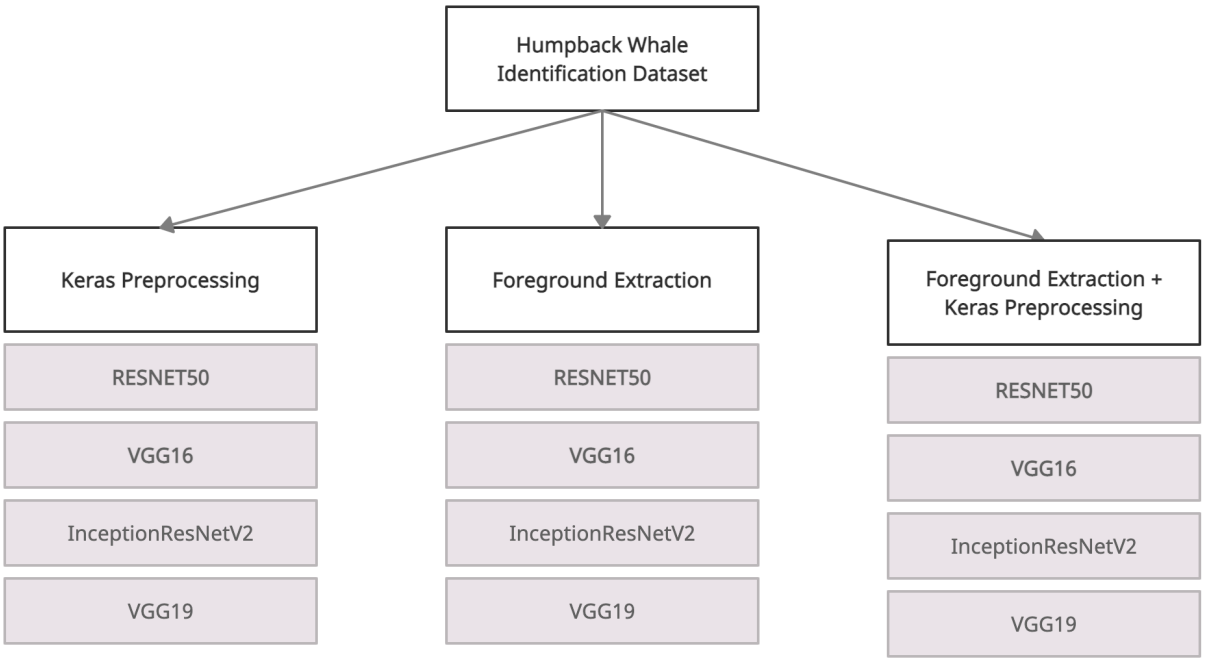


Fig6. Proposed Architecture diagram

Methodology

1) Initial EDA

Like all datasets that require cleaning, The Humpback Whale Identification dataset also included some outliers that could cause the model to have a decreased efficiency. The Dataset included whales which had very few images of a whale fluke. The data was cleaned in order to include only whales which consisted of more than 20 Whale fluke images. The original dataset had 25,361 images.After filtering out images below the 20 image threshold, there were 1,977 leftover images which belonged to 59 different whales.

Fig7. *Dataset sample*

Fixed size

Throughout the rest of the experiment, the size of the images considered for classification was set to 100x100. This had to be done to optimize the training time and the classification accuracy for all models.

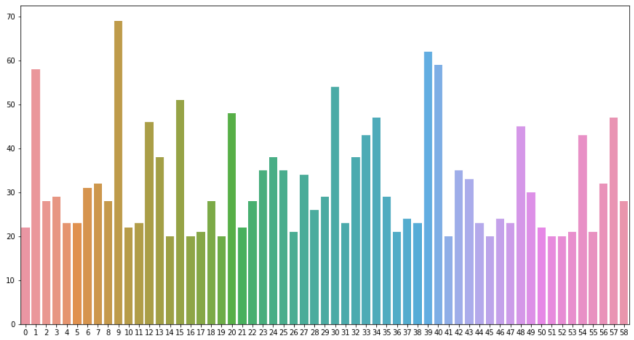
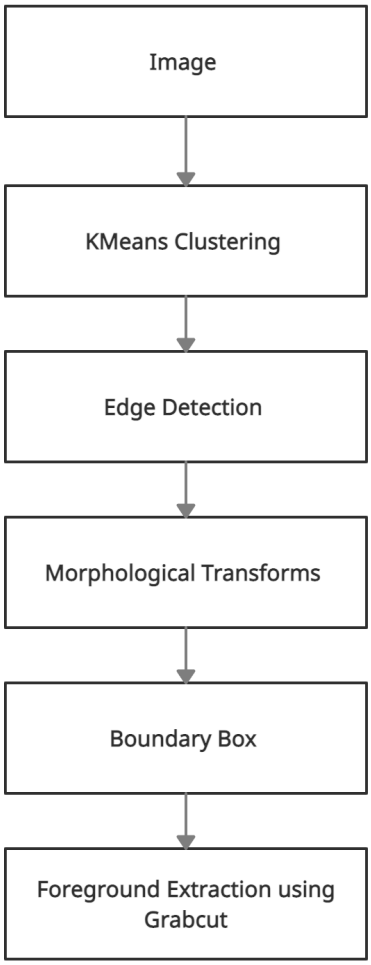


Fig8. *Distribution of images in different classes*

2) Image Processing

a) Foreground Extraction



It is important for the image to be enhanced so that the classification will be much easier for the model.

Foreground extraction is a technique which allows an image's foreground to be extracted for further processing like object recognition, tracking etc. The algorithm used for foreground extraction here is GrabCut Algorithm.

In this algorithm, the region is drawn in accordance with the foreground, a rectangle is drawn over it. This is the rectangle that encases our main object. The region coordinates are decided over understanding the foreground mask. But this segmentation is not perfect, as it may have marked some foreground regions as background and vice versa. The main objective was to get the sea and the sky out of the picture. Still so, the process of foreground extraction increased the accuracy of the classification model.

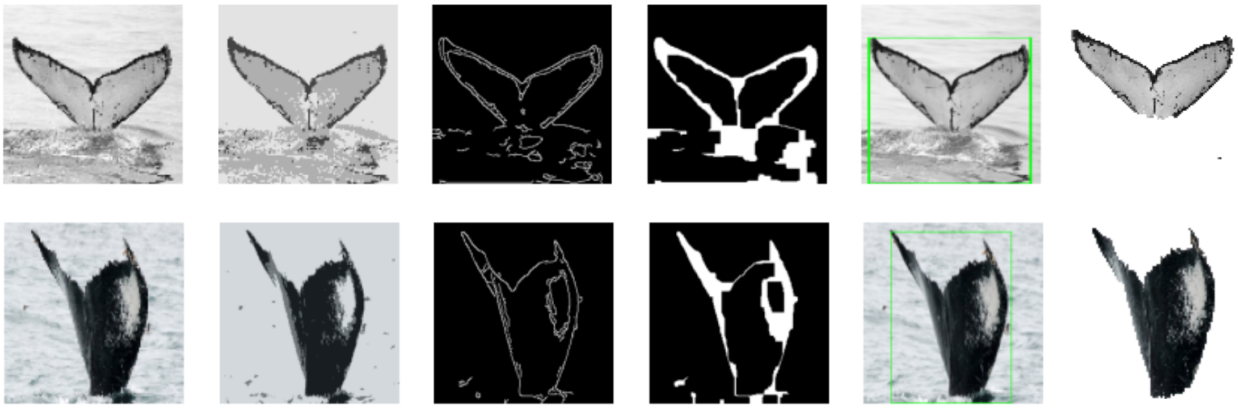
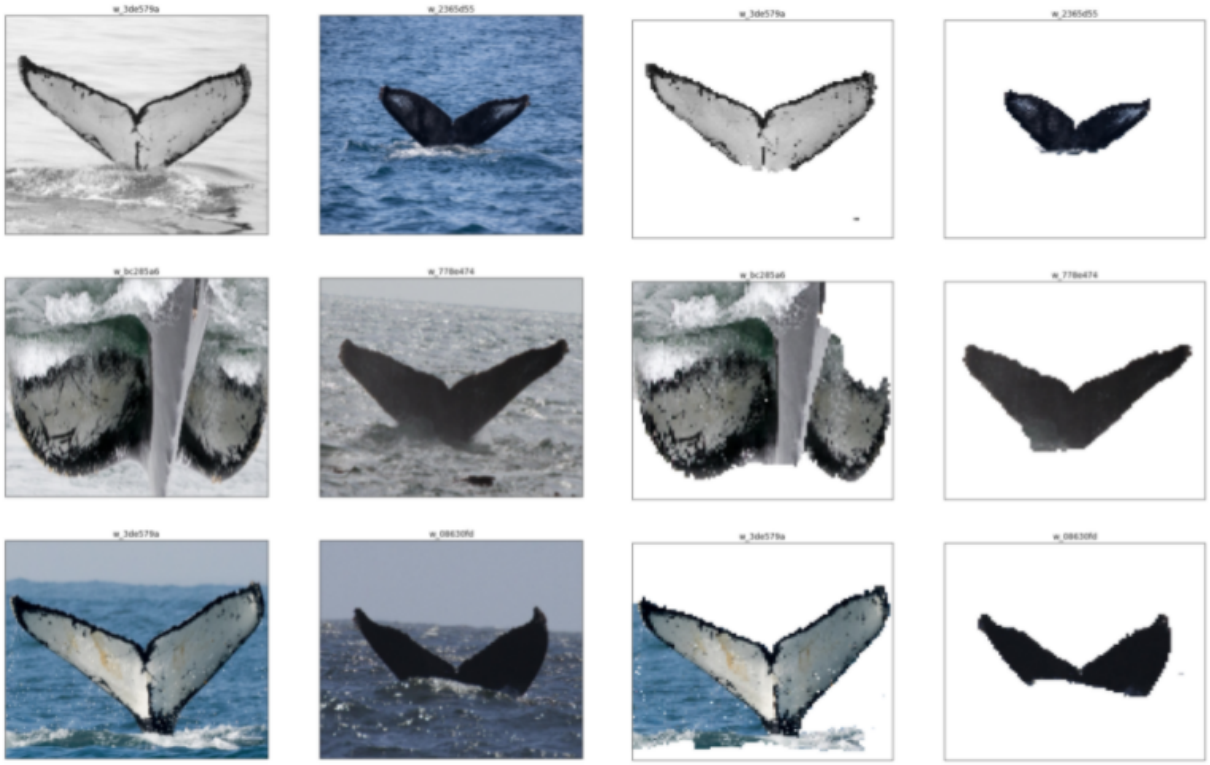


Fig9. Foreground extraction process



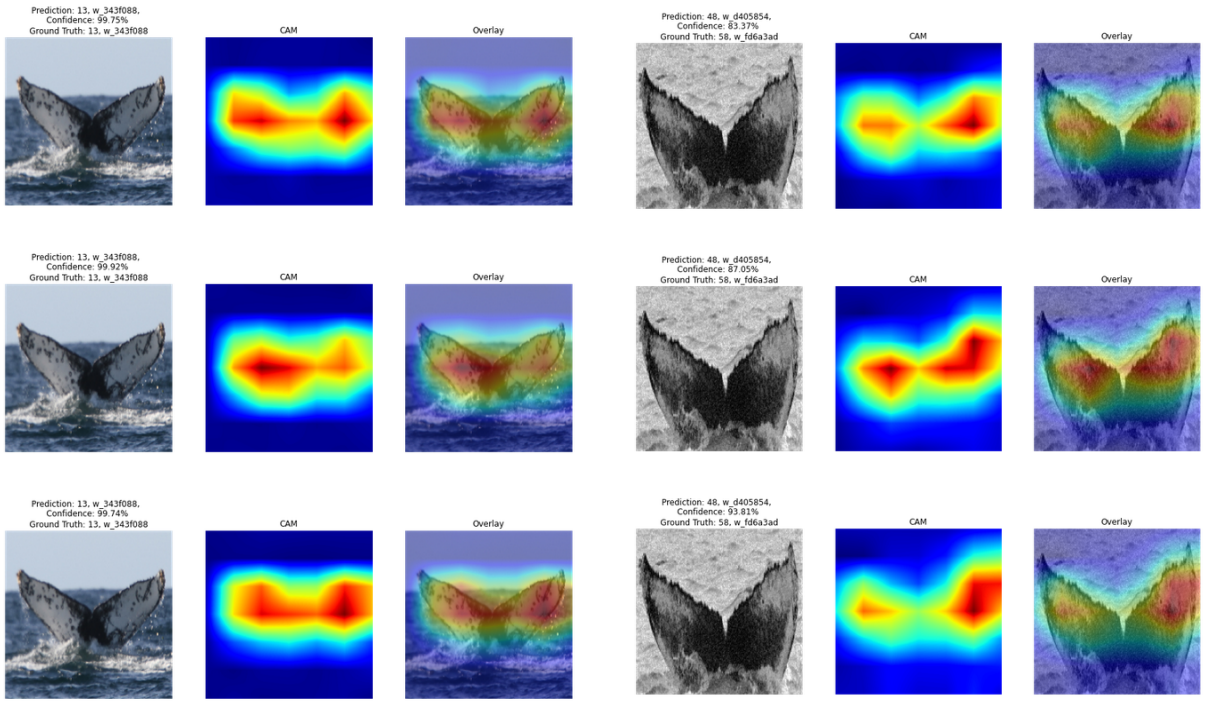
(a) (b)

Fig10. (a) Original image dataset resized to 224x224 (b) Foreground extracted image

As seen in the figures above, The Foreground Extraction Process was carried out on all the images. The objective of the process , that was , to clear the sky and the sea out of the picture seems to have been satisfied.

3) Class Activation Maps

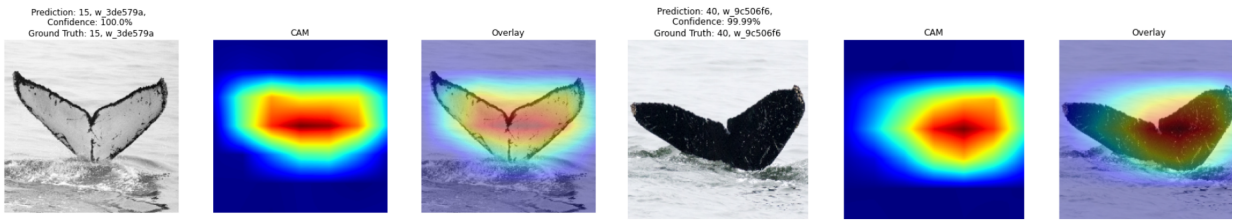
For each of the models (VGG16, ResNet, InceptionResnet) , Images along with their respective class activation maps are plotted to see which parts of the whale fluke activate the neurons. The maximum and a heatmap is generated to overlay on the image to visualise the features which activate the neurons pertaining to that class.



(a) (b)

Fig11. Class Activation maps for each model trained classification of the whale fluke in the test set (a) Correct classification (b) Incorrect classification

When the model is trained on the original image set and the same model used to predict the foreground extracted image the confidence slightly decreases but localisation on the whale tail is higher while predicting on the foreground extracted image while in the second case Fig 11. (b) the classification output was incorrect as it classified as 48 whereas the ground truth value was 58 and although from the class activation map we can clearly see that the localisation is better while using foreground extracted images.



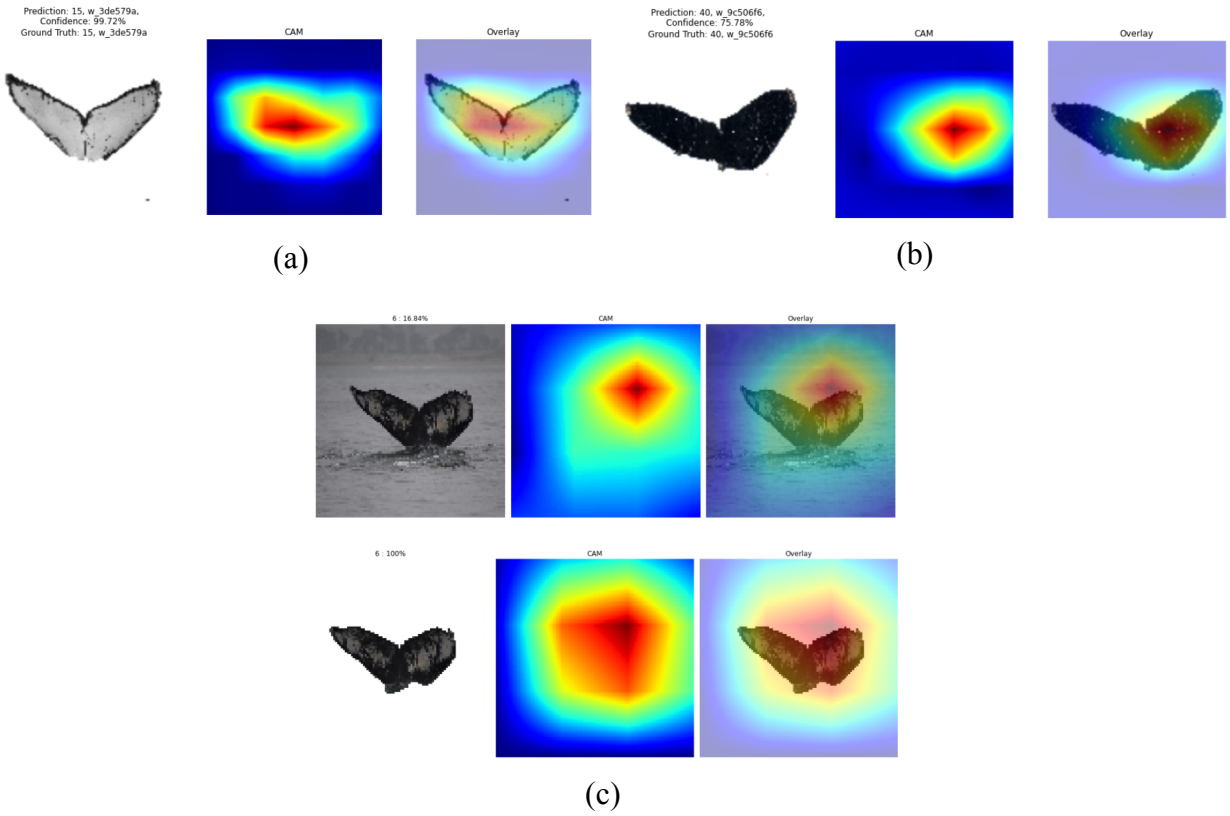


Fig12. *Class Activation maps for each model trained classification of the whale fluke in the test set also with foreground extraction*

Results

Model accuracies

As expected , a below Average accuracy was obtained. In the second scenario, the images from the dataset were as follows. The first case scenario uses the standard preprocessing available from the Keras API. In the second case, recalling that k-means clustering, Morphological operations, Edge Detection are performed while extracting the Foreground.The accuracy after performing this entire Foreground Extraction Process increased by an average of 11%, InceptionResNetV2 and VGG19 being the models showing the most significant difference.

In the third testcase, we performed Foreground Extraction on the Keras Preprocessed data. This is the case with the best results.This showed an average of 6% increase in comparison to the model on which only Foreground Extraction was performed. Overall, this test case showed better results than the models where only Keras Preprocessing was done, by 16%. To conclude, we found that Keras Preprocessing followed by Foreground Extraction showed the best results, and is the best technique for the above Humpback Whale Dataset.

Model Name	[KP]	[FE]	[FE + KP]
ResNet50	0.5556	0.6263	0.6364
VGG16	0.61616	0.6767	0.7879
InceptionResNetV2	0.5353	0.6868	0.7374
VGG19	0.62626	0.74747	0.7894

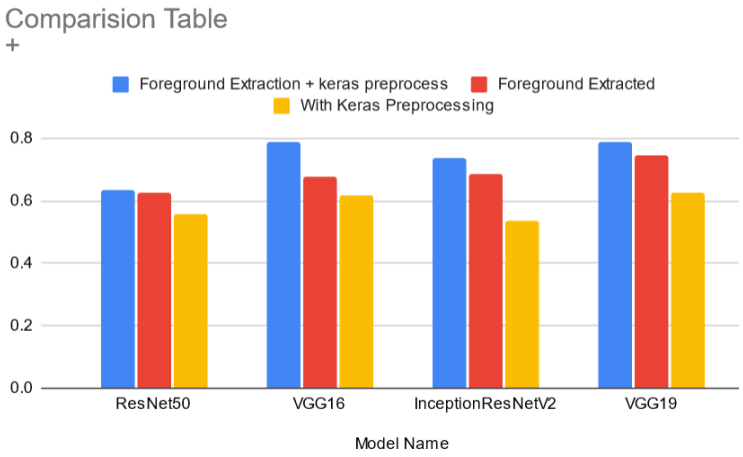


Fig13. Comparison graph

References

[1] Carvajal-Gámez, B.E., Trejo-Salazar, D.B., Gendron, D. *et al.* Photo-id of blue whale by means of the dorsal fin using clustering algorithms and color local complexity estimation for mobile devices. *J Image Video Proc.* **2017,** 6 (2017). [Link]

[2] Guirado, E., Tabik, S., Rivas, M.L. et al. Whale counting in satellite and aerial images with deep learning. Sci Rep 9, 14259 (2019). [Link]

[3] Gómez Blas, N.; de Mingo López, L.F.; Arteta Albert, A.; Martínez Llamas, J. Image Classification with Convolutional Neural Networks Using Gulf of Maine Humpback Whale Catalog. Electronics 2020, 9, 731. [Link]

[4] Hendrik Weideman, Chuck Stewart, Jason Parham, Jason Holmberg, Kiirsten Flynn, John Calambokidis, D. Barry Paul, Anka Bedetti, Michelle Henley, Frank Pope, Jerenimo Lepirei; Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2020, pp. 1276-1285 [Link]

[5] H. J. Weideman et al., "Extracting identifying contours for African elephants and humpback whales using a learned appearance model," 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 2020, pp. 1265-1274, doi: 10.1109/WACV45572.2020.9093266.[Link]

[6] E. Okafor et al., "Comparative study between deep learning and bag of visual words for wild-animal recognition," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850111.[Link]

[7] H. Hsu, Y. Lee, J. Ding and R. Y. Chang, "Dolphin Recognition with Adaptive Hybrid Saliency Detection for Deep Learning Based on DenseNet Recognition," 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Chengdu, China, 2018, pp. 455-458, doi: 10.1109/APCCAS.2018.8605718. [Link]

[8] H. Wang, X. Ji, H. Zhao and Y. Yue, "Semantic Segmentation of Freshwater Fish Body Based on Generative Adversarial Network," 2020 IEEE International Conference on Mechatronics and Automation (ICMA), Beijing, China, 2020, pp. 249-254, doi: 10.1109/ICMA49215.2020.9233767.[Link]

[9] R. Maglietta et al., "Convolutional Neural Networks for Risso's Dolphins Identification," in IEEE Access, vol. 8, pp. 80195-80206, 2020, doi: 10.1109/ACCESS.2020.2990427.[Link]

[10] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2921-2929, doi: 10.1109/CVPR.2016.319.[Link]

[11] Mahdianpari, Masoud & Salehi, Bahram & Rezaee, Mohammad & Mohammadimanesh, Fariba & Zhang, Yun. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. Remote Sensing. 10. 1119. 10.3390/rs10071119. [Link]

[12] T. Tagaris, M. Sdraka and A. Stafylopatis, "High-Resolution Class Activation Mapping," 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 4514-4518, doi: 10.1109/ICIP.2019.8803474.[Link]

[13] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. "GrabCut": interactive foreground extraction using iterated graph cuts.Association for Computing Machinery, New York, NY, USA, 309–314.[Link]

[14] Beekmans, Bas & Whitehead, Hal & Huele, Ruben & Steiner, Lisa & Steenbeek, Adri. (2005). Comparison of Two Computer-Assisted PhotoIdentification Methods Applied to Sperm Whales (Physeter macrocephalus). Aquatic Mammals. 31. 243-247.10.1578/AM.31.2.2005.243. [Link]

[15] Mahdianpari, Masoud & Salehi, Bahram & Rezaee, Mohammad & Mohammadimanesh, Fariba & Zhang, Yun. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. Remote Sensing. 10. 1119. 10.3390/rs10071119.[Link]

[16] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.[Link]