LION RE-ID: A NON-INVASIVE METHOD FOR INDIVIDUAL IDENTIFICATION OF ASIATIC LIONS

A thesis submitted to the
University of Petroleum and Energy Studies

For the award of
Doctor of Philosophy
in
Computer Science Engineering

BY
Glen Bennet Hermon

December 2021

SUPERVISOR
Dr. Durgansh Sharma

School of Computer Science
University of Petroleum & Energy Studies,
Dehradun – 248007: Uttarakhand

LION RE-ID: A NON-INVASIVE METHOD FOR INDIVIDUAL IDENTIFICATION OF ASIATIC LIONS

A thesis submitted to the
University of Petroleum and Energy Studies

For the award of
Doctor of Philosophy
in
Computer Science and Engineering

BY
Glen Bennet Hermon
(SAP ID- 500057008)

December 2021

SUPERVISOR
Dr. Durgansh Sharma
Sr. Associate Professor
School of Computer Science
University of Petroleum and Energy Studies
Dehradun, Uttarakhand

School of Computer Science
University of Petroleum &Energy Studies,
Dehradun – 248007: Uttarakhand

DECLARATION

I declare that the thesis entitled "**Lion Re-ID: A Non-Invasive Method for Individual Identification of Asiatic Lions"** has been prepared by me under the guidance of **Dr. Durgansh Sharma**, Sr. Associate Professor of the School of Computer Science, University of Petroleum & Energy Studies. No part of this thesis has formed the basis for the award of any degree or fellowship previously.

Glen Bennet Hermon
School of Computer Science,
University of Petroleum & Energy Studies
Dehradun-248007: Uttarakhand

Date: **27-12-2021**

Great Place To Work. Certified MAR 2021-FEB 2022 INDIA

UPES
UNIVERSITY WITH A PURPOSE

CERTIFICATE

I certify that **Glen Bennet Hermon** has prepared his thesis entitled "**Lion Re-ID: A Non-Invasive Method for Individual Identification of Asiatic Lions**," for the award of a Ph.D. degree from the University of Petroleum & Energy Studies, under my guidance. He has carried out the work at the School of Computer Science, University of Petroleum & Energy Studies, Dehradun.

Supervisor
Dr. Durgansh Sharma
Till July 2021:
Sr. Associate Professor
School of Computer Science
University of Petroleum and Energy Studies
Dehradun, Uttarakhand

Currently:
Associate Professor
School of Business and Management
Christ University
Delhi-NCR Campus

Date: 27-12-2021

Energy Acres: Bidholi Via Prem Nagar, Dehradun - 248 007 (Uttarakhand), India T: +91 135 2770137, 2776053/54/91, 2776201, 9997799474 F: +91 135 2776090/95
Knowledge Acres: Kandoli Via Prem Nagar, Dehradun - 248 007 (Uttarakhand), India T: +91 8171979021/2/3, 7060111775

ABSTRACT

Former techniques for the identification of lion individuals (Panthera Leo) relied on manual methods of recording data. Such processes have various shortcomings due to the manual nature of recording data. This research work aims to automate the process of encoding the uniqueness within the whisker spot patterns for each lion individual by non-invasively using photographs. Towards this research work the main bottleneck was the availability of image data for individual lions. The proposed model embeds the uniqueness within the patterns for a specific individual as a unique cluster within its embedding space. This is achieved by using a triplet loss function which, due to its one-shot learning nature trains a deep inception network with less training data. Photographic images are known to have variations in lighting, pose variation, angle variation and other inconsistencies. Since the nature of these issues are nonlinear, it is preferred to create the target model using deep learning techniques. An inception network is trained to generate 128-dimensional vectors unique to each lion. This research paper elaborates on such deep machine learning techniques and other processes that are used to create this model.

The task of animal identification can be undertaken by a wide variety of techniques that may be invasive or non-invasive in nature. The harmony in lifestyle of animal individuals identified by the invasive processes may suffer due to the intrusive nature of such techniques. There is thus incentive to make progress in the direction of non-invasive identification techniques. Using various image processing techniques to analyse unique natural markings captured non- invasively in photographs is the predominant technique that is able to satisfy such a necessity.

Image processing and pattern recognition have come a long way since its inception and we have seen many useful implementations of the same in various fields and sciences. But there is always scope for more. Constructing a tighter

confidence interval for similarity between each image does solve the problem in most cases. This work deals with such an optimization.

The classification of many animals in perspective of the wildlife heritage that we are blessed with is comparatively easier when compared to that of the Lion. The reason being that there is a lack of peculiar patterns like spots or stripes over the Lion's body.

Visual animal biometrics is non-invasive and cost effective for wildlife monitoring activities. There is a problem of detection of individual species in multiple captured images. We have a large collection of images which is hard to manage and to recognize using manual identification techniques. There is a lack of automation in the detection process while retaining robustness to image ambiguities (like blur, fading, occlusions and pose variation).

To identify individual Lions (Panthera Leo) noninvasively there is a current requirement for developing an automated Lion detection and recognizing system for individuals based on whisker spot patterns that are present on each side of the Lion's face. No two whisker spot patterns are the same and they do not change over time.

By identifying each lion, one can track individuals using advancements through the proposed technique and thus identify pride home ranges and population trends. This allows for effective conservation.

The identification process is done by using various deep machine learning techniques to generate a learned model that is able to embed the uniqueness of individuality within a fixed dimensional embedding space. This model is thus able to create unique identification vectors for each lion that acts as a key while trying to identify a lion, given a new photograph. During the identification process the system then just needs to compute the distance of the current vector with the set of stored vectors of known lions to check for a match with the least distance. To have a higher confidence, a tighter threshold of similarity can be used. This thesis elaborates on the aforementioned techniques that help in uniquely re-identifying Lion Individuals to their photographs.

ACKNOWLEDGMENT

I would like to begin by thanking God for enabling my very existence, eventually leading to the accomplishment of this work. I would like to thank all the people who supported me, undertaking this research. I would first like to thank my PhD supervisor, Dr. Durgansh Sharma, for the past three years at UPES. His passion and persistence in technology has constantly encouraged me during my PhD study. In my first year, he spent enormous efforts guiding me on research and never lost patience when my efforts were not producing desired outcomes. He would seek every chance to train my paper writing and presenting skills that help build my confidence and become a qualified researcher. Without his constant support, I would never be able to accomplish my PhD study. All the knowledge and merit that he has passed on to me, has made me what I am today and will continue benefiting my future life.

In addition to my Supervisor, I would like to express my profuse appreciation and gratefulness to Dr. S. J. Chopra (Chancellor), Dr. Sunil Rai (Vice-Chancellor), Dr. Priyadarshan Patra (Dean - School of Computer Science), Dr. Devesh Kumar Avasthi (Dean - Research) and Dr. Vijendra Singh (Assoc. Dean – Research, School of Computer Science) for supporting this work in all its stages as well as in helping me pursue my research at the University of Petroleum & Energy Studies (UPES).

I would also like to mention my gratitude to the HOD and the faculty colleagues of the School of Computer Science for motivation and cooperation. I would like to acknowledge the in-house Image processing lab, Cybernetics Lab and the Machine Intelligence Research Centre at UPES, Dehradun for providing facilities to conduct my research experiments. I am also grateful to the lab staff and their support at our Institute (UPES) for facilitating the various machine training procedures towards the creation of the resulting model.

A sincere note of thanks to Dr. Manish Prateek, Dr. Kiran Kumar Ravulakollu, Dr. Monit Kapoor, Dr. Neelu Jyoti Ahuja, Dr. Susheela Dahiya, Mr. Anil Kumar who were asscociated with the School of Computer Science as well as Dr. Y.V. Jhala and Dr. Qamar Qureshi at the Wildlife Institute of India (WII) for their valuable suggestions and guidance at every stage of this work. Special thanks to Mr. Saurabh Shanu for their support and for connecting me with this opportunity and with the folks over at the Wildlife Institute of India. I am deeply thankful to all my colleagues at the University of Petroleum and Energy Studies: Dr. Debajyoti Bose, Dr. Akash Gupta, Dr. Meera C. S, Dr. Caneon Kurien, Dr. Varnita Verma, Dr. S.R.V.S Prasanna, Dr. P. Sairam, Dr. B. Aslesha, Mr. Amar Shukla, Mr. Sudharsan J., M., Mr. Anuraj Malav, Mr. Madam Chakradhar and Ms. Shubhangi Kala for their invaluable comments and suggestions during group meetings, which were instrumental in my graduate research and getting this dissertation completed in time. I would like to extend my gratitude to Ms. Nissie Glory Chegudi, Mr. Kevin Pant and his family, Mr. Ravi Teja Panga and his wife Mrs. Divya Swarupa Bathala, for all the support during the last moments of the thesis documentation process.

Finally, I would like to thank my dear parents Dr. Terrence Robert Hermon and Dr. Jiby Maria Hermon for their immense support and love at every stage of my life. Completing this degree would never be possible without their love and encouragement. I want to thank my little sister as well, Miss Sharon Liz Hermon for her steadfast support. I also want to thank all the members of my family, uncles, aunts, cousins and most of all, my grandparents who were the source of my inspiration. I have become brave enough to face the toughness of life only because of all of their encouragement and motivation.

To my Family

And

My Late Paternal Grandfather “Er. Bennet George Hermon”

And

My Late Maternal Grandfather “Er. Thomas P John”

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Greek symbols

$\boldsymbol{\alpha}$	Margin parameter in embedding space
$\boldsymbol{\lambda}$	Focus hyperparameter

Other Symbols

LI-No²	Lion Identification Non-Invasive Normalization
LMFT	Lion Mystacial Features by Triads
LMId	Lion Mystacial Identity
YOLO	You Only Look Once (Neural Network)
$\boldsymbol{x}$	x-coordinate
$\boldsymbol{y}$	y-coordinate
$\hat{\boldsymbol{x}}$	predicted x-coordinate
$\hat{\boldsymbol{y}}$	predicted y-coordinate
$\boldsymbol{w}$	Anchor box width
$\boldsymbol{h}$	Anchor box height
$\boldsymbol{C}$	Object presence score
p(c)	Classification loss
$\mathbb{1}_{ij}^{obj}$	Object presence parameter
$\mathbb{1}_{ij}^{noobj}$	Object absence parameter
$\boldsymbol{B}$	Number of anchor boxes
$\boldsymbol{S}^2$	Total number of grid cells
$\boldsymbol{A}_{(i)}$	Anchor image vector
$\boldsymbol{P}_{(i)}$	Positive match vector
$\boldsymbol{N}_{(i)}$	Negative match vector
LFlab	Dataset of labelled lion faces
CLWlab	Dataset of cropped lion face with labelled whisker region
CLwin	Dataset of cropped whiskers tagged by individual identity
Y1	First YOLO net
Y2	Second YOLO net
Incep	Inception net
_BackProp	Back Propagation (postfix)
_ForwardProp	Forward Propagation (postfix)
LiDb	Dataset for vectors of known lions
I	Input image
crop_img	Image cropping function
Square_diff	Calculate sum of the squared differences
Threshold_diff	Threshold variability for a match

ABBREVIATIONS

RMS	Root Mean Square
SVM	Support Vector Machine
ROI	Region Of Interest

CHAPTER 1 INTRODUCTION

Image processing and pattern recognition have come a long way since its inception and we have seen many useful implementations of the same in various fields and sciences. But there is always scope for more. Current object detection and pattern classification algorithms classify to a limited degree of granularity which limits us in various ways. Constructing a tighter confidence interval for similarity between each image does solve the problem in most cases. This research intended to be carried out, deals with such an optimization.

As technological advances are made to sustain the wildlife in our environment there is a requirement of robust non-invasive techniques equipped to distinctly identify individual animals within a species. Identification of individuals from species and tracking of the identified individuals have been conducted with mostly invasive techniques in the past. However, recently multiple methods for non-invasive identification of individuals in the species have emerged, using computer vision algorithms [142] [143]. In addition, the time complexity and space complexity of the technological approaches are far better than the previously used manual approach for individual identification.

The classification of many animals in perspective of the wildlife heritage that we are blessed with is comparatively easier when compared to that of the Lion. The reason being that there is a lack of peculiar patterns like spots or stripes over the Lion's body.

The Asiatic lion has been of concern for many years. Previously, reintroduction and translocation efforts had been undertaken to try and establish another population but these efforts were not successful, due to lack of proper planning and methodology.

The Asiatic lion is a large, predatory carnivore which used to range over much of the Indian subcontinent and surrounding area. It is an animal whose size strength and nobility have earned it an identification with emperors and Kings. It is an important cultural and historical symbol for India, having been selected as the emblem of the Government of India. The behavioral and biological characteristics of the animal are such that it requires a large area to permit normal social interaction with its conspecifics as well as containment in a protected area away from human habitation.

The Wildlife Institute of India and the Government of India have supported research to study the ecology of the Gir population and take up the matter of finding an alternative habitat for Asiatic lion.

The thesis illustrates the procedure and the design of the deep machine learning technique used to generate a learned model that is able to embed the uniqueness of individuality within a fixed dimensional embedding space. The thesis also illustrates how the model is thus able to create unique identification vectors for each lion that acts as a key while trying to identify a lion, given a new photograph. The performance estimation for different embeddings and epoch schemes for the re-identification is also a part of this thesis. The chapter introduces the aspects related to this topic, the outlines to this thesis and the organization of this dissertation.

1.1 MOTIVATION

Q. Why are the Asiatic lions the center of focus for this research?

Ans:

1. Proper planning and methodologies are required for carrying out census of Asiatic Lions found in India [119] [182] [191].
2. It is an important cultural and historical symbol for India, having been selected as the emblem of the Government of India.

3. By identifying each lion, one can track individuals and thus identify Pride home ranges and population trends. This allows for effective conservation technique [88] [91] [94] [120] [184].

Q. How do you identify individual lions noninvasively?

Ans:

Identifying individual lions noninvasively [163] can be done by developing an automatic Lion detection and recognizing system for individuals based on whiskers spot patterns that are present on each side of the Lion's face. This is possivle as, no two whisker spot patterns are the same and they do not change over time, which is proven by its usage as the current convention (Figure 1.1).

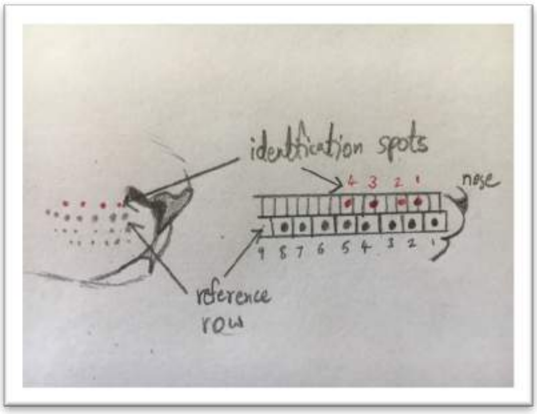


Figure 1.1: The conventional representation for lion identification

Q. What is Re-Id as occurring in the title of this thesis document?

Ans:

The target of this research work is to create models for identifying lion individuals such that, these models [86] [90] encode representations to individuality and thereby re-identify [84] [87] these representations by their respective real world lion identity, i.e., the name or identification number of the lion. This may be done only for known or previously identified lions [78] [79] [80] [83]. This process therefore, enables a system to re-identify known lions within images that contain unidentified lion individuals therefore the term re-identification or Re-Id (Figure 1.2).

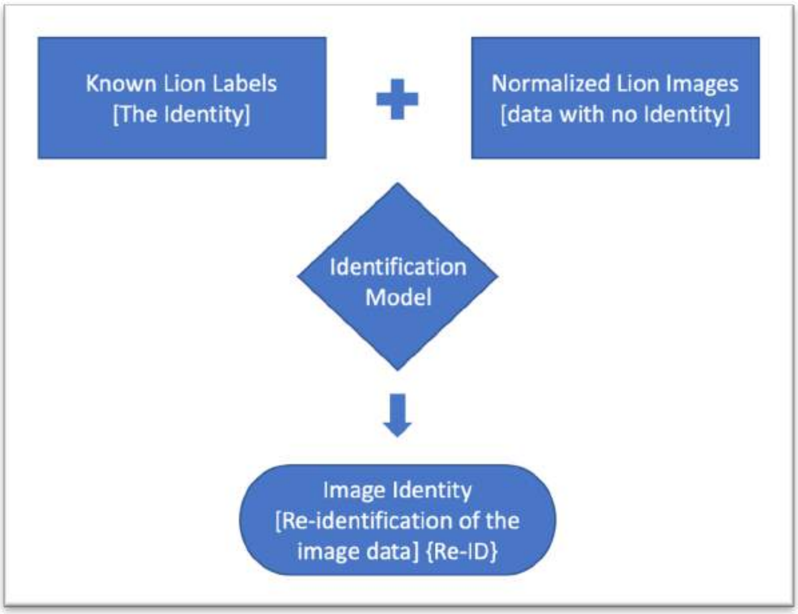


Figure 1.2: Re-Identification

1.2 PROBLEM STATEMENT

"To design and develop a Non-Invasive method of individual re-identification using computer vision techniques for the Asiatic Lion species."

Visual animal biometrics (for Asiatic lions) is non-invasive and cost effective for wildlife monitoring activities. There is a problem of detection of individual species in multiple captured images. We have a large collection of images which is hard to manage and to recognize using manual identification techniques. There is a lack of automation in the detection process while retaining robustness to image ambiguities (like blur, fading, occlusions and pose variation).

1.3 RESEARCH GAPS

The aim of this work is to uniquely identify Asiatic Lions out of the photographs provided. The classification of the photographs will be done by grouping the images of the same Lion as a single group by constructing a confidence interval for similarity between the lions in each image.

- Currently, individual identification of Lion is done manually by recording the spot patterns from the photographs, which is tedious, erroneous and time consuming.
- The accuracy of the current manual system has a lot of scope for improvement.
- Tools for individual identification of Lions are not available.
- Current object detection and pattern classification algorithms classify to a limited degree of granularity.

In addition, we will explore and implement different pattern recognition algorithms that are currently used for pattern extraction, classification and recognition purposes. Then we will focus on the extraction of the Region of Interest (ROI) out of the images that need to be uniquely identified, classified and recognized.

1.3.1 SOME SPECULATIONS ASSOCIATED WITH RESEARCH QUESTIONS:

Investigation has been carried out and the following research questions were addressed that will helps in understanding the current research gaps and thereby provide guidance to fill the same:

- What are the different methods used by wildlife agencies to conduct a census?

- How many of the different wildlife census methods have been automated?
- What is the accuracy and ease of the automated systems that are currently developed?
- Are there other species like the Lion that suffer the difficulty to make such an automated system?
- Why it is difficult to identify lions and what is the current convention used for the record keeping of unique lions and the ones that are already identified.
- What results can be further achieved, like age or gender.
- What may the ambiguities in detection be? Special cases such as scars on the area of the whisker spots.

1.4 OBJECTIVES OF RESEARCH

1. Design and development of a tool to extract the region of interest for patterns of the Lion.
2. Implement image grouping methods for uniquely identifying lions.
3. Quantitative analysis through pattern recognition algorithms.

1.5 OUTLINE OF THE PRESENT WORK

Starting out by exploring the various techniques and processes used to identify animal individuals within a single species, Chapter 2 records a detailed account of the various techniques used within the individual identification processes implementerd for various animal species. The wide range of techniques are categorised by the nature of the various target patterns. After the exploration of the animal identification technique landscape, Chapter 3 talks about all the efforts undertaken towards the goal that this research work targets. The efforts undertaken prior to this research work as well as all the shortcomings that various approaches might have due to the constraints faced by the nature of this work, are thereby addressed in this chapter.

Chapter 4 mentiones in detail the current manual mode technique for Lion individual identification, followed by an implementation of a semi-automated Lion individual identification technique. The developed tool for the semi-automated system has also been described, while providing a peek into its interface. This semi-automated identification technique is implemented by structuring the problem as a classification problem, solved by various machine learning approaches using the data collected by manual input, using various normalization and ambiguity rectification techniques. This paves the way for the research work to further into the Deep Learning domain, the intricacies of which are elaborated in Chapter 5. The Chapter 5 covers in detail the LI-No^2 normalization algorithm, which is a novel method for normalizing the data for this domain. Details of the building blocks of the Lion identity extimation system using deep learning is also focused upon. This coves the implementation of the Siamese network archetceture with its triplet loss function, to facilitate the training of the embedding space of the fearure extractor.

Moving onto the various comparative and qualitative analyses, Chapter 6 records the various reports of performance with respect to accuracy as the performance parameter for the different identity estimation models created by the multiple approaches and configurations. This chapter also elaborates on the dataset creation process, thereby providing the information for the model creation. And lastly, Chapter 7 discusses about the various aspects of this research, and concludes with a look into the future, paving the way for the furtherance of work within this domain.

CHAPTER 2 LITERATURE REVIEW

2.1 INTRODUCTION

This chapter reviews the various research conducted based on the region of interest used in the identification processes of multiple species [139]. There exist a wide range of differences among these techniques, based on the patterns in consideration and the approaches used. This amounts to a large collection of techniques that have value and scope in the subject of pattern recognition and thus demand for a comparative study of these techniques, with a discussion on their accuracy, ease of use and their adaptability in various scenarios. This work reviews the different pattern matching techniques among a plethora of algorithms used to identify animal individuals, within various animal species along with their complexity, categorizing them based on the type of pattern used for recognition.

The emphasis is towards the insights of computational techniques for various image- based animal biometrics with the intention to automate these processes. It is observed that mathematically modelled identification techniques with algorithmically specialized normalization techniques are the most efficient identification techniques in a broad range of scenarios.

Unique Individual Identification within a species creates a foundation for the study of species demography, behavioural patterns, lifespan and other ecological parameters. These parameters help in the creation of various statistical models based on multiple individuals of the same species. The task of individual identification within a species has been handled utilizing various methodologies for various species with shifting degrees of accomplishment in the last few decades. Each species has certain highlights that end up being one of a kind among the specimens of that species [125] [182]. Such highlights like spot designs, bristle spots, coat stripes, and so forth, are used to comprehend the uniqueness of those examples with the goal that productive procedures and calculations can be performed for non-obtrusive identifications.

2.2 THE LANDSCAPE OF INDIVIDUAL ANIMAL IDENTIFICATION

For unique identification of individuals within a population, various invasive techniques such as tattooing, attaching radio tags, etc. have been adopted. However, these practices involve physical contact with the animal, causing disturbance in the regularity of the animal's life-routine [196]. In addition, the invasive methods have the potential to cause injury, disrupt the food chain and sometimes lead to separation of individuals from the community [63]. On the other hand, non-invasively, capturing photographs makes it possible to record identification patterns unique to the individuals within the species. Researchers in the early times have also used manual drawing and techniques of highlighting the unique features, assigning identification cards to each individual within the focal species. With the advancement of technology, the older techniques proved to be slow and tedious even though they tend to have a higher amount of accuracy. Thus, it created a necessity for adoption of technology and computer systems to capture, store and retrieve features and information of individuals using various identification strategies [111] [112] [113] [177] [178] [179].

The best and most used image-based animal identification models are those based on spot patterns, fluke or fin Patterns, coat patterns, whisker spots, face recognition and body part deformation patterns. Individuals of tigers are identified by the unique stripe patterns over the skin. The skin texture information has been used for individual identification [110] [24]. For Polar Bears, stating some of the key findings and the relevance of the work, a short analysis of areas in respect to the above discussions shows us that the identification system developed uses their whisker spot patterns. Three reference points are manually needed to be selected, to extract the region of the whiskers from the photographs [44]. Secondly, as for the non-invasive methods of identifying elephants, the nicks and cuts on the elephant's ears pose to be unique.

A context-free approach of the Generalized Hough Transform, with the ability of handling non-connected curves is used to extract these patterns from the edges of an elephant's ears [17]. Further, in Humpback Whales and Sperm Whales, the fluke patterns such as notches, holes, spots, lines etc. are observed and recorded manually onto computer systems, with an observer based coding scheme [59][66]. Next, African Penguins [64] have black spots on their chests that remain same during their adult life. The prominent black stripe on the chest that lies between the white areas of the neck and the chest provides as an initial factor to pick out adult penguins [5]. Progressively, Pinnipeds are carnivorous aquatic mammals like seals or walruses that can be identified through photographs of the flippers - fore or hind that display damage or abnormalities, large scars on the body and size of the lower canine [58]. Similarly, the individual identification of the yellow bellied toad species is done by capturing photographs of the belly patterns and generating an identification code. The code comprised of eight numbers that the first digit that denotes sex or the life stage and the remaining digits denoting the pattern and the presence of spots on different regions of the belly [61]. Chimpanzees being primates, have many similar structural features to humans, the most common face recognition algorithms used for humans tested on chimpanzees to produced good results. Multiple modifications are applied to the basic algorithms used for humans, in which similarity scores computed over global features are combined with similarity scores computed over local features to generate a more accurate score [8].

Table 2.1: Outlook into the various approaches to animal individual identification

Author year of publication	**Observed species**	**Spotting area finding**	**Identified computer vision and pattern recognition techniques**			**Additional findings**
			Modelling and learning algorithms	**Model representation**	**Matching and localization technique**	
Kshitij Bakliwal and Sai Ravela[67] (2020)	Salamanders and Geckos	Coat and body patch patterns	• CNNs, SVM	• Mean-shift, and segmentation	• Graph cut, pairwise comparison	Only shallow CNNs required
Rotimi-Williams BELLO , Abdullah Zawawi Hj TALIB , Ahmad Sufril Azlan Bin MOHAMED[68] [16](2020)	Cow	Nose Image Pattern	• Deep Belief Network, Stacked denoising autoencoder	• Difference of gaussian (DOG) features	• Image texture features – fixed window image captured	Gradients used efficiently
Sagonas, C., Panagakis, Y., Zafeiriou, S., et al.[36] (2016)	Cat, tiger, lion, Panda, fox, cheetah	Head counting based feature of animal (deformable decomposition of head shape and texture)	• boosting • SVM classification	• box bounding features • census features • HOOG features	• dual approach • sliding windows	Head and facial features
Kiapour, M.H., Jagadeesh, W., Di, V., et al.[28] [41](2015)	African Penguins Spheniscus [64] demersus South Polar Skuas and Adélie penguins	Chest patterns (body) natural markings in the chest plumage	• Computer vision models • Features Procrustes + mean square error • Regression models	• Chest pattern Features	• Matching software • cascaded rules	• Training the labeled key points on chest body as annotation
Kiapour, M.H., Jagadeesh, W., Di, V., et al.[28] (2015) Atanbori, J., Duan, W., Murray, J., et al.[18] (2016)	Oriole bird (Baltimore oriole (Icterus galbula))	Colour feature	• Deep learning • CNN feature baselines	• LDA • Entropy-rank curve	• Grab-cut mask • Fine-tuned CNN features	• Training the model with annotation colour-based features using bounding box

Ernst, A., Ku blbeck, C. [22] -2011	Chimpanzee, gorilla, monkey	Facial images (rigid spatial decomposition) and (representation of facial features)	• AdaBoost classifiers • indexing • lookup tables	• bounded box-based features• census features • histogram of oriented gradient (HOG) features	• cascaded rules• sliding windows based algorithms	Facial region using bounding box
Chen, J., Wen, Q., Qu, W., et al. [21] (2012)	Pet animal (cat, dog)	Shape and texture features, facial, body and nose print (for cat) (deformable decomposition of face features and uniform body texture)	• Sparse Representation • latent SVM • Colour histograms, Gabor • Descriptor model	• Geometric model • Active appearance models • Deformable model • HOG features • Pixel colour	• Descriptors of part localization • Matching • Sliding windows • Grab-cut • Segment compactness	Facial region, annotation using bounding box
Jill, M.L., Buler, J.J., Frank, R.M. [26] (2016) Piczak, K. [35] (2016) Van Horn, G., Branson, S., Farrell, R., et al. [40] (2015)	Birds (NABirds, dataset containing 48,562 images)	Colour feature	• Machine learning• MTurkers	• Pixel intensity of bird images	• Machine learning (SVM) classification	• Body colour annotation using bounding boxes

For Chimpanzees further techniques like Deep convolution Neural Networks (CNN), using Gabor magnitude pictures (GMPs), achieving dimensionality reduction by locality preserving projections (LPP) with classification techniques such as sparse representation classification (SRC), are used. The Stochastic Gradient Technique is also used along with Backpropagation to compute gradients of intermediate layers. The Main focus is to achieve fine-grained recognition, and for such tasks of animal identification the matrix logarithm method proved to increase accuracy, while in the domain of second order statistics [52]. Lemur identification is also done by face recognition after normalizing the images to extract a local binary Multiscale pattern by a patch-wise method. Various techniques are used to normalize for facial hair and ambient lighting with Linear Discriminant Analysis (LDA) being the dimensionality reduction technique used [49] [116]. The Great White Sharks are identified by their fins with a multi-scale approach that gives the ability to have multiple labels for each image. The process collects contour information rather than colour or texture information, this is then used to generate an ultrametric contour map used for the identification process [55]. For Australian sea lions whisker spots are selected manually and stored in a semi-automatic system, with pairwise comparisons of the Chamfer distance transform under a minimal variation in angle [9]. For Bewick Swans, the distinct identification feature is the yellow and black patterns found on the upper surface and either side of each bird's bill. These patterns are recorded on the identification card of each bird, along with other information such as the location, sex, and other features like gape etc [62]. For leopards, the whisker spot patterns are taken for the identification of individuals. The algorithm used is the Oblique Principal component cluster analysis technique and to minimize correlation, unique character sets are generated [2]. When identifying cheetahs, their coat spot pattern is unique to every individual. A model of a 3D plane is made in such a way that it fits on the photograph after a few points are identified such as the position of the shoulder. An Identifier array which is a sample of grey-scale intensities stored as a matrix of numbers, serves as an identifier for individuals. Individuals are uniquely identified by calculating the correlation coefficient and similarity coefficients between the models of known individuals and the patterns in the identifier arrays of the image input in consideration [27]. Grey seals display spot patterns on their pelage and major patterns of the head and neck are taken into consideration to identify individuals. The matching is done manually by at least two independent observers [65]. Multiple Marine Vertebrates like the Delphinid (dolphin like) species have nicks and notches on their dorsal fin’s trailing edge, that provides for the matching process involved in identifying individuals. The methods that take advantage of these patterns are the curve matching and string-matching techniques. Curve matching plots the Euclidean distances of the depth of each notch in sequence and saves the pattern. The string-matching technique involves the assignment of different string characters for respective crests and troughs in the notch pattern and thus saves this string as the pattern. For a multispecies approach, a live wire algorithm to locate the fin boundary along with various noise reduction and smoothing functions are applied before storing the curves or strings for the respective methods [44][54]. Badgers are small carnivorous animals very similar to rats and squirrels that have peculiar tail patterns that encapsulate the uniqueness of individuals [50]. Whale Sharks are a species of fish whose individuals are identified by the spot pattern on the body of the fish. After extracting the region of interest and subsequently the blobs, an algorithm used in astronomy to find the position of stars, the Groth’s algorithm, is used to generate a similarity score [6]. The identification procedure used for Zebras is like the fingerprint identification techniques used for human fingerprints [51]. For Marbled Salamanders, Long-Tailed Salamanders and Spotted Salamanders, recording the dorsal patterns enabled individual identification[32][33]. Multi scale PCA, the ANOVA method and region based encoding techniques are used to extract and store features [53][56][48].

The shift in identification techniques of individuals from the manual mode to the technological mode has improved both the time as well as the space complexity involved with the purpose but has also shown variation in the level of accuracy, adaptability and ease of use. This creates a demand for a comparative study between the various techniques.

We present a comparative study of various techniques used for individual identification within the species. A computational comparison of the different techniques provides a structure for categorizing these models based on complexity, similarity and repeatability of algorithms used. One of the primary goals of the paper is to analyse different techniques and emphasise on the various issues faced by the stakeholders while implementing these techniques, which may be both physical and technical.

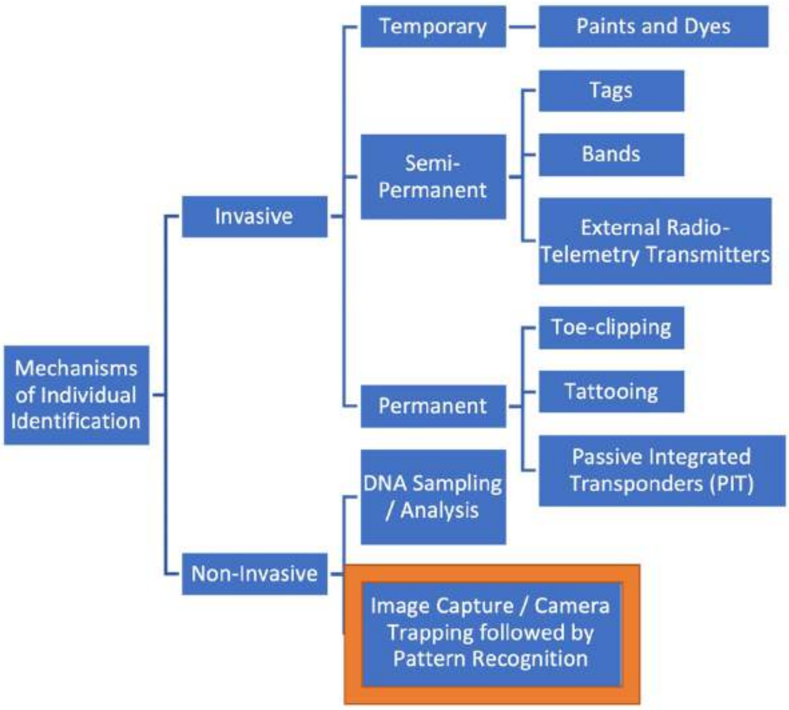


Figure 2.1: Mechanisms of Individual Identification

The techniques, when compared with each other, distinctly show that the ones that rely on mathematical modelling are chosen more often, hinting to the accuracy that it provides and proving that such techniques based on mathematical modelling are the better techniques. Such a mathematical model will be able to quantify the uniquely identifiable aspects of appearances by reducing noise and ambiguity and by implementing various normalization techniques to quantify observable properties.

Unique individual identification amongst the species have been performed either by marking them manually or by recognizing the uniqueness of the texture of their body, which may include patterns, colour, etc. as demonstrated in Figure 2.1. This unique identity and feature must be sufficient to prove individuality over the whole size of the population of that particular species in consideration.

In the past few decades, researchers have tried to identify several techniques to identify the unique patterns for individual tagging within the species with a considerable amount of accuracy. Initially, researchers used to manually draw and assign identification cards to each individual that stored values of various other features, which could not be drawn. With the advancement of technology, the traditional techniques appeared to be slow and complex even though they held a good amount of accuracy. This led to the adoption of upcoming technologies and computer systems, which captured stored and retrieved features and information of individuals from focal species using various identification strategies.

According to the analysis presented in [60], larger the sample size, lower is the level of accuracy for identification of individuals. However, the accuracy happens to be directly proportional to information content, i.e. if the information content of a pattern is substantial to identify an individual out of a hypothetical sample size, then such a pattern is very reliable and tends for a higher level of accuracy in individual identification of the focal species. A proper literature review in the matters of advancements of Artificial Intelligence and computational aspects suggests that reliability assessments are difficult and provide better results by assuring the information content of a pattern.

The above discussion suggests that a good practice would be to assess the likelihood of a pattern replication. If the actual size of a population is large, then the complexity of the pattern used for identification should be considerably better in order to obtain the required information content. Hence, it suggests that the pattern complexity also holds a proportionate relationship with the [25] population size of the focal species. Reliability hence can be expressed as the probability of a pattern being duplicated within a selected size of a population.

The authors [60] suggest that a starting iteration would be, odds of 100:1 chosen and the addition of a few bits (1.7 bits) to make the odds 1000:1 to be preferred, and adding a similar step-up of more information in bits (1.7 bits) in order to increase reliability to the odds of 10000:1 chosen as a preferred reliability for most identification procedures. As most identification procedures also record locations of detection, even in cases of migratory avian species and Cetaceans, the reliability scores and the level of accuracy can be made higher. A similar argument had also been in [62] which states that, while considering all animals and birds, the questions that can be asked depends on the number of individuals that can be represented and recognized [62].

Considering all the various species that the research community has worked with, a broad categorization of the methods used for identification based on the similarity of the type of patterns that are used for unique identification is possible, as shown in Table 2.2.

Table 2.2: Features used for Non-Invasive Identification of Animals

Feature	**Animals Identifiable by said feature**
Unique Spot Patterns	African Penguins, Whale sharks, Spotted Salamanders, Long tailed salamanders, Grey Seals
Fluke / Fin Deformation Patterns	Great White sharks, Dusky dolphins, Spinner dolphin, Bottlenose dolphin, Long finned pilot whale, Sperm whale, Humpback whale, other cetaceans

Coat Patterns	Tigers, Cheetahs, Amphibians (yellow bellied toad), Zebras, Marbled salamanders
Whisker Spots	Leopards, Polar Bears, Lions, Sea Lions
Deformation of body part	Elephant, New Zealand Sea Lions,
Face Recognition	Chimpanzees, Lemurs.
Unique body features	Swans, Badgers

Described below are detailed accounts on these approaches mentioned in Table 2.2, that are used to uniquely identify individuals of different birds and animals. The time complexity and space complexity of the documented techniques have been generated based on the computation model described for each process, giving a reasonable insight into each technique.

2.3 SPOT PATTERNS

The animals that are identified by quantifiable Spot Patterns that cover only specific regions of the body are African Penguins, Whale sharks, Spotted Salamanders, Long tailed salamanders and Grey Seals.

Grey seals display spot patterns on their pelage. To identify individuals, major patterns of the head and neck are taken into consideration. The patterns include black patterns over white. The whole coat in general becomes darker as the seals get older and due to sexual dimorphism, it becomes difficult to identify male individuals as the coat usually becomes totally black deeming it impossible to identify many males individually [65]. The patterns on females get more prominent with time, making it easier. The matching is done manually by observers and the quality of the pattern is classified into good, medium and bad.

Whale Sharks are a species of fish that are vulnerable to extinction due to excess targeted fishing due to its international trade value [6]. A unique technique using numerical pattern matching is used to identify individual whale sharks. Previous to this usage of the numerical algorithm it is majorly used in the field of astronomy to compare patterns of stars[46]. This algorithm is majorly based on the Groth's algorithm that helps to identify the positions of stars. To extract the region of interest the image is rotated so that the vertebral column is horizontal. After a generic method of Blob extraction, the algorithm basically finds out all possible triangles between any two coordinate points, these triangles should also be of a specific shape, having specified the location of the shortest, intermediate and longest edges between respective vertices. Many unwanted triangles are also filtered out, such as a threshold of triangles too small to be useful. The matching procedure after the generation of these triangles are based on the length ratios and the internal angle cosine for each triangle. A magnification factor is computed for each triangle compared, which leads to the computation of the similarity score.

For African Penguins, to extract the black spots on their chest right below the prominent black stripe of the neck [5]. During the acquisition of images, constraints such as weather conditions of no rain and neither images of dusk nor dawn are chosen. The chest is outlined by first starting with an initial seed outline and then expanded to the actual borders of the chest. The spot patterns are extracted by a series of image processing techniques that includes dilation, morphological differencing, spot kernel convolution, thresholding to generate a unique identifier. While querying with an identifier for a new image with the identifiers stored in the system, the weighted sum of matching every spot in the identifiers under comparison generates the similarity score for that pair. The database entries are also sorted based on similarity so that comparisons are fast. The natural markings on African penguins are the spots over their chest plumage that prove to be unique, and in a later work [37], the AnimalID software [47] is used for the enrollment of the penguins into the system, by selecting penguin images that the camera could capture with near-frontal positions, being orthogonal to the camera within 20◦ to 30◦. The artificial intelligence algorithms in the AnimalID software handles the identification and matching operations. Three different matching techniques are executed to observe how they perform. A basic benchmark of performing a binary task to classify authenticatable images (images of individuals enrolled in the database) apart from those that aren't, creates an error statistic that is called a receiver operating characteristic (ROC). The first step involves a technique that normalizes scale, shift and rotation before calculating the mean square error between landmarks that are closest. This is a Procrustes analysis method that results in a rigid alignment technique. The second technique is by creating shape contexts of the spot neighborhoods by their polar histograms, sets of such histograms are used for comparison between individuals as a matching technique. The third method is an extension to shape contexts by creating a distribution context that is distortion-specific, which is produced by generating a model for the landmark location uncertainty. To further quantify the system's performance statistical values like Genuine Acceptance Rate (GAR) and Failure to Enroll (FTE) rate are also formulated. Methods like attentional cascades and fast rectangular pixel sum is used so that the process can be sped up to be used in video as well.

Long-Tailed Salamanders have unique spot patterns on the dorsal side of their body [33]. A spot recognition method can be performed to recognize individuals with the help of the overall pattern along with the spot count [56]. It is based on the use of just the number of spots present on the head region, which are the dorsal spots that are on the region that starts from the nose area to the area just at the beginning of the forelimbs, not considering the spots that occur laterally on the sides. The ANOVA method is used to determine if there exists a relationship of the number of spots with the length of the salamander. The Least Significant Differences mean separation test revealed that the number of spots increased with the increase in length [34]. Even though it could be inferred that spot patterns increased as the salamander grew, the general patterns of the spots remained unchanged and unique. It is also observed that there is a higher tendency for spots to appear than disappear. The pattern recognition of spots hence proved to be a more useful technique.

For spotted salamanders a method to generate a code based on the spot patterns on the individuals is used [48]. These codes are generated by recording the number of spots present on six regions on the dorsal part of these salamanders. Since there exists an issue of pattern codes being repeated among few individuals, there is an extra bit of information recorded for such individuals. To determine the accuracy of the identification technique, logistic regression is performed in the software R. The Binary Response variable in this case is defined over how likely the system is able to identify an individual correctly, a yes or no response and the likelihood of this outcome as well.

Table 2.3: Algorithms and techniques used for Animals with Spot Patterns

Algorithms and techniques	**Remarks**
Haar Transform	Haar-like feature extraction methods are used to train classifiers to detect the area of interest along with the patterns in them by learning simple local luminance features, in African Penguins.
AdaBoost	Used to take advantage of the combined performance of all the strong classifiers by taking classification and regression trees remodeled from the feature extraction methods as the input to the algorithm, for African Penguins
Sobel feature extraction	To make sure that there exists a penguin chest in the image the Sobel-like kernal detection techniques are used so that a detailed chest outline could be generated.

Procrustes Algorithm	To compare for a match, a form of the Procrustes algorithm is used, in African Penguins.
Polar Transform	A polar transform of the chest pattern, is used to compare for a match for African Penguins
Groth's Algorithm	The triangles and their internal angles, generated from this algorithm is used for the comparison for a match, in whale sharks.
Global Thresholding	To reduce noise, used for both Whale sharks and African penguins.
ANOVA	Used to determine the relationship between the spot pattern and the length of the tail in long-tailed salamanders.
Logistic Regression	Used to determine the accuracy of identification in spotted salamanders.
Manual observation	Done for the spot patterns in Grey Seals.

2.4 FLUKE OR FIN PATTERNS

The animals that are identified by fluke or fin patterns are Great White sharks, Dusky dolphins, Spinner dolphins, Bottlenose dolphins, Long finned pilot whales, Sperm whales, Humpback whales, and other cetaceans [102].

The identification for dolphins is done by creating a curvature description of its dorsal fins, rather than measuring and comparing the curvature itself [45], by assigning a string of characters that represent significant and insignificant primitives. This provides as a kind of feature reduction approach to dimensionality reduction. Every fin is thus assigned character strings, which are matched to new fins to provide similarity scores. Further calculations provide us with a family of semantic / syntactic distances between two strings, which yields a composite distance to measure dissimilarity. For a multispecies approach, a live wire algorithm to locate the fin boundary along with the help of noise reduction techniques and various smoothing functions are applied. Even though the data entry required experts to handle the inputs the search mechanism proved that computers are more effective in terms of time saved [54]. The number of incorrect matches served as a criterion to measure the performance of the system. The issues faced by this system are mostly due to noise in the images posed by environmental artefacts like glare, splashes etc. Hence quantification of image quality is a required parameter that has a direct effect on the performance of the system.

For Humpback Whales the fluke patterns such as notches, holes, spots, lines etc. are observed and recorded manually onto computer systems as the storage search and retrieval media [59]. The coding scheme is also totally observer based like the techniques used for identifying swans [62]. Independent SCAN and MATCH functions are created for an efficient retrieval of individual information from the search space with the MATCH function searching for the exact match given the encoded identification scheme and the SCAN function that returns groups of results that stand true to the partial encoded information provided (e.g., all whales in the database with striped flukes). The main issue is the quality of the image captured and displayed for the observer who makes the categorization codes for each of the whales. This introduces many inconsistencies like observation and observer ambiguities. The system developed [59] records the recognition quality for each observation into four categories of excellent, moderate, poor and ‘not coded’ – acting like a new observable measure. This method gives us a quantifiable approach to confidence in the system. The other valuable insight is that a single photograph of a unique individual is not enough to make a new entry in the

database, there should be at least two images as a minimum to describe an individual. The measurement of quality of the data collected is an important aspect when it comes to identifying individuals. An interesting observation duly noted (Friday et al., 2000) is that, highly distinctive individuals are more easily identifiable from photographs of lesser quality, but this isn't really a strong point in support of photographs of lesser quality due to the fact that less distinctive individuals are totally left out. Thus, the quality of photographs also depends on how distinctive each individual is from the rest. To quantify both the aspects of quality of photographs and the distinctiveness of an individual, an evaluation criterion involving people as judges to rate each aspect is proposed. The issue of how each judge is supposed to differentiate between both aspects is resolved by breaking each aspect into various variables that are rated individually that contributing to the overall score towards each aspect. The process evaluated the agreement between the judges by evaluating the average kappa statistic score for each judge derived from the agreement statistics of every possible pair of judges – an inter-rater agreement.
For Sperm whales, an operator enters values to the computer system in a unified and standardized structure and the computer is programed to process the identification based on the formulas and techniques specified. A technique involving the calculation of value of match and accuracy of match to derive the match coefficient proves to be an efficient method [66].These values are possible only after a process that involves an operator to look at the images, in this case the fluke images of sperm whales, and then enter a coded representation, as identifiers of the various marks and shapes on the edge of the fluke by their proportional distance, generating a unique code of fixed length for each individual. The measures of orientation, tilt and resolution are used as parameters that indicate the quality of the image.
For Great White Sharks the paradigm of non-invasive identification of individuals is approached by a non-linear model, that incorporates multi-scale segmentation with the uniqueness selected over the space created over scale variations [55]. This approach gives the ability to have multiple labels for each

image. Contour information is collected rather than color or texture information, which is then used to generate an ultrametric contour map. The contour information helps to generate multiple boundary descriptors that make up a Bag of Boundaries (BoB), which is used to detect the presence of the shark fin by classification techniques. After this a Bag of Normals (BoN) is generated, which encompass the shape information along with the regional information. While encoding the contour information the Difference of Gaussian (DoG) is used as a filter to normalize the curves. The training is done with the ground truth contour labelled by a human. The quality score is generated by thresholding of the L2 normalized values of opponent-SIFT(scale invariant Feature Transform) for local appearance which is independent of direction and the thresholded values of contour shape by the histogram of boundary normals which are dependent on direction. The final matching process is handled by a random forest structure, that ranks the various contours. This ensures that the search time is lesser than a linear search time. The standard DoG norm, which is rotation invariant, is combined with LNBNN (Local na¨ıve Bayes nearest neighbor), so that a recognition baseline is generated, after which a scoring pattern can be generated. To achieve scale invariance, classification is done separately for each scale and then combined, also affine-covariant sift descriptors are used. The overall model automates the process of obtaining the identity of the animal from the natural image provided to the system. The results of the system are tested on the basis of Average Precision (AP) over a Precision-recall Curve (PR). They are also able to create a population wide fin-space, which is like a single model for the whole species, with unique patterns for individual distinctiveness. This is done so that the system doesn’t learn patterns of uniqueness separately for each individual and that the learning is not overfitted over individuals. To achieve this, a standard for partitioning each fin contour is devised. Thus, a separate score is generated for each fin-space as well as for each class.

Table 2.4: Algorithms and techniques used for Animals with Fluke or Fin Patterns

Algorithms and techniques	Remarks
String Matching or Encoding	Arbitrary Encoding used for sperm whales, humpback whales and Dolphins
Kappa Statistic Score	Derived from the statistic of multiple judges for the quality of the photographs.
Live Wire Algorithm	For a multispecies fin identification approach
Affine Transform	An additional method for scale invariance used in the identification of great white sharks.
Difference of Gaussian (DoG)	To. Achieve rotational invariance for the fins of the great white sharks.
SIFT (Scale Invariant Feature Transform)	To normalize over scale of the fins in great white sharks.
Random forest algorithm	For the final matching in great white sharks.
Local Na¨ıve Bayes nearest neighbor (LBNN)	To create a baseline for the fin recognition in great white sharks.

2.5 COAT PATTERNS

The animals that are identified by their coat patterns are Tigers, Cheetahs, Amphibians (yellow bellied toad), Zebras and Marbled salamanders [152].

Tigers can be identified by differences in their stripe patterns [57]. Since tigers and leopards occur syntopically, the estimation of their densities, can be performed by capture-recapture theory [19]. The tiger expresses surface patterns on its body coat which prove to be unique [24]. A 3-D surface model is proposed to solve issues related to the angle of the photograph and the posture of the tiger. The patterns are matched to produce a similarity score with the help of two algorithms that complement each other. It is basically a semi-automated system to extract the pattern, and the software goes by the name 'Extract Compare'. Due to the method relying on user inputs there is a risk of subjectivity in the case of model fitting, which is being looked into for further automation.

For Cheetahs the technique of having a 3-D mathematical model [27] of the coat or body part in consideration is a direct solution to resolve multiple issues that exist in the available 2D images such as, position of the animal, angle of the camera, tilt, scaling, orientation, and other issues like partial capture. The model is made in such a way that it fits on the photograph after a few points are identified such as the position of the shoulder.

For Zebras, six location points are used to define the region of interest [51] [135]. After the selection of region of interest, the stripes [114] [115] within them are skeletonized to single pixel lines. Comparing the complexity of the patterns occurring on human fingerprints, the patterns on the zebra coat [30] contain lesser complexity, granting the fact that human fingerprints are way more feasible for individual identification procedures.

Moving towards marbled salamanders, with the widespread reduction of pond basins in the ecological habitat of marbled salamanders, there is a risk of this species to be endangered [53]. The traditional technique of capture-mark-recapture proved to be less accurate and very time consuming with significant intrusion on the regular life of these salamanders. Therefore, the method of photographically recording the dorsal patterns while individuals are captured is found to be more feasible. These dorsal patterns act as unique fingerprints. The process is semi-automated as one has to digitally mark the dorsal midline after which the system pre-processes the image to straighten the midline in effect straightening the whole image warping it from its original form. But the pattern is preserved, and this process makes all the comparisons to have a standard form. Multi scale PCA is used to extract and store features [20], also the whole database of images is ranked on the similarity with each other. The outcome of this research provides new information on migratory patterns and total time spent within basins, due to the information recorded based on individual identification.

Table 2.5: Algorithms and techniques used for Animals with Coat Patterns

Algorithms and techniques	**Remarks**
PCA /Multi Scale PCA	Extracting and storing features in marbled salamanders.
String Matching or Encoding	Encoding the presence or absence of patterns by the region of the body (yellow bellied toad).
Edge Detection	Used in the too that extracts the features in the stripes of tigers.
3-D mathematical model superimposition	Solves issues related to the angle of the photograph and the posture of tigers. A 3D plane is used in the case of cheetahs.
Correlation	Created from the models of known individuals of

coefficient	cheetahs.
Fingerprint Identification Technique	Patterns on zebras are similar to human fingerprints.

2.6 WHISKER SPOT PATTERNS

The animals that are identified by their whisker spots are Leopards, Polar Bears, Lions and Sea Lions [146].

The study on leopards is done on captive leopards making the setting quite invasive [2], the readings being taken while the animals are under anesthesia. The algorithm used here is the Oblique Principal component cluster analysis. To minimize correlation, unique character sets are generated. The Binomial Theorem is used to check these character set spot clusters for the sum of probabilities of zero occurrence and that of a single occurrence to be at a maximum (more than 0.95), to be feasible to be used as a characteristic of identification used in the process.

For Australian sea lions the whisker spots are selected manually and stored in the system with only the right side of the sea-lion's face chosen for the identification process. Scars are not used as the sealions undergo molting. To standardize the images 3 manual reference points are taken. To calculate similarity scores, the Euclidian distance between points in the 1st image in comparison and the nearest points in the 2nd image and then the reverse distance from the 2nd image to the 1st is calculated, also to improve the similarity score the image is shifted manually in steps. An adapted Chamfer distance transform is also used. To improve the system, the use of Groth algorithm is suggested so that a geometric relationship is made between the spots. The encoding and comparison is achieved by superimposing a grid over the manually marked spots with the selection of grid-cell size by accounting for the angle of the photographs. The grid is filled based on the presence or absence of spots. The accuracy of the system decreased when the photograph is not at the 90◦ of the right face-profile of the sealion.

For Polar Bears to validate the reliability of the patterns, the complexity and information content is measured [44]. To compute the similarity scores between images, the Chamfer distance transform is used. Images of only one side of the bear's muzzle is used for analysis, due to the notion of a possible correlation of the patterns of both sides that might result in a statistical bias. The whole procedure of analysis including, selection of photographs, locating the spots, information content and reliability, mutual exclusiveness are repeated across 3 judges over the same number of photographs to account for consistency. The variations in accuracy of identification is analysed based on the angle and quality of the photographs. Affine transformation, global thresholding, adaptive thresholding, histogram specification, logarithmic transformation, are the algorithms used for pre-processing. Based on the similarity scores, and the errors of matching, a tolerance graph of false positives is generated from which the similarity threshold is derived. Even though three reference points are manually selected, the time taken for the processing on one image is under a minute.

Table 2.6: Algorithms and techniques used for Animals with Whisker Spots

Algorithms and techniques	**Remarks**
PCA / Multi Scale PCA	Cluster analysis for leopards.
Chamfer distance transform	Pairwise comparison of these distances are possible only under minimal variation in angle in the photographs of Australian sea lions. For Polar Bears Chamfer distance transform is used to generate similarity scores between

	images.
Groth's Algorithm	To improve the system for Australian sea lions by creating a geometric relationship between the spots.
Affine Transform	These are the various processes used to extract the whisker spots from the images of polar bears.
Global Thresholding	
Adaptive Thresholding	
Histogram Specification	
Logarithmic Transform	

2.7 DEFORMATION PATTERNS

The animals that are identified by the unique deformations in various body parts are Elephants and New Zealand Sea Lions.

For the New Zealand Sea Lion, while trying to capture photographs of the abnormalities in the flippers and other features like large scars and the size of the lower canine, it is found that the animal may not comply, may be aggressive or the animal may depart before the photograph is taken [58]. Also the recorded features of body and facial scars may change by the next moult (shedding feathers).

For Elephants, to extract the patterns of the nicks and cuts on the elephant's ears that pose to be unique [105], a common edge detector wouldn't be feasible to detect them due to the uniform color and texture over the elephant's body, over which the images of the ears overlap [17]. And most of the curve detector algorithms handle these patterns by assuming connected curves. Hence the choice of Generalized Hough Transform, with the ability of handling non-connected curves – a context-free approach. With core efforts to automate the process their system is semi-automated with users required to input positional information like the position of the head [42][43] and various reference points like the location of the eyes, etc. After this step a canny edge detection algorithm charts out the edges that it sees, the user then has to input the start and end point of every curve and nick. These patterns are stored as a sequence for that specific image. The matching algorithm first compares to check for portions of matching sequences of just the presence of these nicks or cuts in the same sequence. By increasing the number of cuts for every successful sequence portion matched (eg. Set of 2 cuts matched then going on for sets of 3 cuts in the matching portion of the sequence), the whole sequence is checked. After this each nick or cut is compared by the shape difference algorithm. Every curve is compared to give a final average dissimilarity value, as the output of the matching process. This method proved to be much more accurate than other curve-matching algorithms.

Table 2.7: Algorithms and techniques used for Animals with unique Deformations

Algorithms and techniques	**Remarks**
String Matching or Encoding	Strings that describe the curvature of the dorsal fins of dolphins. Patterns in the elephant's ears are also stored as

	a sequence of strings
Curve Matching	Used to match the curve sequences on the edges of the elephant's ears.
Edge Detection	To extract the curves of the edges in the elephant's ears.
Generalized Hough Transform	For the ability to handle non-connected curves in the elephant's ear edges.
Context- Free Algorithms	Useful when it comes to edges of elephants ears, because of the undifferentiable colour.

2.8 FACIAL PATTERNS

The animals that are identified by Face recognition are Chimpanzees and Lemurs.

For Chimpanzees that have similar structural features to humans, common face recognition algorithms are observed to produce good results. The popular real-time object detection algorithm by Viola and Jones, combined with techniques that take advantage of Gabor descriptors along with SURF descriptors (taking into account texture and shape features), and Sparse Representation Classification (SRC) over global features along with support vector machines (SVM) over local features for the required classification routines are also used for a fine-tuned recognition algorithm [8]. The Locality Preserving Projection (LPP) method is the dimensionality reduction procedure adopted. With all these finer adjustments to the basic face recognition algorithms [117], there is an improvement in the accuracy for non-frontal images, which is an issue in a previous version of the system. Techniques like Deep convolution Neural Networks (CNN), using Gabor magnitude pictures (GMPs) are also used to improve efficiency. The Stochastic Gradient Technique is also used along with Backpropagation to compute gradients of intermediate layers [52].

For Lemurs images are normalized to extract a local binary Multiscale pattern by a patch-wise method [49]. To normalize for facial hair and ambient lighting, various techniques are used. Inter-class similarity and intra-class variability are tested by considering each class as an individual. Linear Discriminant Analysis (LDA) is the dimensionality reduction technique used. The percentage of correct matches is termed as True Accept Rate (TAR). It is also observed that, in previous related studies of individual identification of chimpanzees, the Gabor features are feasibly used due to the lack of hair on the face compared to lemurs. Observations point that, clustering with classification can improve the identification system over unknown individuals in the wild.

Table 2.8: Algorithms and techniques used for identifying Animals by Face Recognition

<table>
<tr><th>Algorithms and techniques</th><th>Remarks</th></tr>
<tr><td>Locality Preserving Projection (LPP)</td><td>The dimentionality reduction technique used for chimpanzees.</td></tr>
<tr><td>Linear Discriminant Analysis (LDA)</td><td>The dimentionality reduction technique used for lemurs.</td></tr>
<tr><td>Gabor descriptors</td><td rowspan="2">Various techniques used for both chimpanzees and Lemurs to conduct the facial recognition task.</td></tr>
<tr><td>SURF descriptors</td></tr>
</table>

Viola and Jones object detection	
Sparce Representation Classification (SRC)	
Support Vector Machines (SVM)	
Stochastic Gradient Technique	
Matrix Logarithm Technique	
Deep Convolution Neural Networks (CNN)	

2.9 OTHER PATTERNS

The animals with other unique body features covered below are that Bewick Swans and Badgers.

The unique features of the Bewick Swans that described their individuality are drawn manually and various other features are recorded [35] on identity cards designed specifically for this purpose [62]. Hence the number of details that needed to be recorded, depended on the total population in consideration[31]. This is initially done by recording the differences of one individual bird from the others and then resulted in maintaining an exhaustive record of individual features recorded systematically [38] for all birds by having identity cards assigned to each of them. The unique features are the yellow and black patterns found on the upper surface and on either side of each bird's bill. To account for referencing and placing an individual into a fixed location in the population, a code is devised that has types of features and ranges of variations within those features with ranges having arbitrary order of precedence. This method is not a fixed standard, but it allows for a better method to search for an individual and also allowed for a method that can be used for computerization in later stages. The outcomes made possible are, social structure, seasonal and daily movements [23][29], life history, similarity of bill patterns in parents and offspring. Limitations: The observations are heavily dependent on the observer making these observations, hence introducing ambiguity amongst multiple observers and their recordings [39]. The coding technique that is developed to quantify the shapes and patterns encodes observer biases into the code itself. And the system being physical involves a physical search through a huge list of encoded identifiers to deem whether the entry is new or not. Also, it is established that it is very difficult to assign a method for the identification of Cygnets (young swans).

For Badgers and their tail patterns, an accuracy of 95% is achieved for the tail patterns, by direct identification looking at the photographs [50]. Also, the variations in the appearance of the tail at any specific moment, apart from the patterns present on them, signifies various social interactions among individuals. Identification can also be done by the skull shape with less accuracy.

Table 2.9: Algorithms and techniques used for Animals with other Unique features

Algorithms and techniques	**Remarks**
String Matching or Encoding	An arbitrary code for all the variations recorded in Bewick Swans is manually generated.

Manual observation	Recording the tail patterns of badgers manually

2.10 CURRENT TECHNIQUES AND ISSUES IN IDENTIFICATION

Currently, individual identification of Lions is done manually by recording the spot patterns from the photographs, which is tedious, erroneous and time consuming. Though the record keeping and the matching process is done with the help of computational devices [108], the whole process of the data entry is done manually including the entries that record each location of the whisker spots in a grid like convention. The rows of the grid correspond to the rows of the lion's whiskers. Various other information along with the whisker patterns such as the presence of scars, notches on the ears and other metadata of the photograph taken are also manually recorded to aid the identification process.

An issue with this system is that it lacks granularity of the information being entered, leading to ambiguity in the entry of the spot locations within the grid system. This ambiguity is arbitrarily resolved only by the decision-making process by the person who is entering the data onto the system, causing potential disagreement when there is more than one person entering this data into the system. Hence this system has a degree of bias to the decision-making process of the person making the entries. Such an issue may be solved by taking multiple readings with the help of different people [4] making the entries into the system, and having ambiguities handled by having the most repeated entry chosen or to have the system take an average reading of all entries. But even this approach is costly both in terms of effort as well as time. The accuracy of the current manual system has a lot of scope for improvement.

Looking towards computational techniques that input the image as a whole and not just a tabular representation of the spot patterns, one must consider the nonlinear variations in pose that occur in different images of the same individual. Mathematically, the various transformations that may be applied to various degrees are that of rotation, shear, scale variation, distortion (linear angular variation), deformation (non-linear) and occlusions in the extreme case. Such manipulations can be translated back and forth by mapping out the various differences in the transformations and creating an affine map, a Euclidean map or an advanced 3-D model that acts as a map to the information represented in the images. Manually charting out such a mathematical representation is tedious and is open to error for new representations of the same data with the absence of any formal technique to prove for accuracy. Also, such a model is more rigid towards accommodating such new representations or variations [5, 6].

Apart from the issues in pose variation and the prospects of creating a mathematical model to represent all such variations, there exists other complications that come along with the computational analysis of images, such as background noise, lighting, shadows, contrast ratios and the complications of various environmental and climatic conditions that get recorded on these images [7]. Charting out these fluctuations mathematically by separately modeling each kind of noise is possible but costly and will still have room for error accounting for a new possibility of such combined image noises.

Along with these variations that may be estimated, it has also been observed that lions may have scars over their muzzle that results in the damage of their whisker spots. In this case the scar could be taken as an identification feature. A similar situational issue of the presence of flies on the lion's face pose a greater problem of mistaking a fly for a whisker spot. Such issues are difficult and near impossible to be mathematically modeled.

Thus, to automate the creation of a model that combines the mathematical representation of the variability in the lion's pose along with the removal of the various image related and situational noises, we have used deep machine learning techniques, to first isolate the region of interest (the lion's whisker spot region), and then subsequently generate a unique vector for the identification process.

CHAPTER 3 REGION OF INTEREST LOCALIZATION BY FEATURE EXTRACTION TECHNIQUES

3.1 INTRODUCTION

The ideation of development of "Lion Re-ID: A Non-Invasive Method for Individual Identification of Asiatic Lions" is presented in this chapter.

This section presents the proposed architecture (Figure 3.1), including the various components [106] and flow of processes between each component leading to the prediction of the identity of the Lion individual.

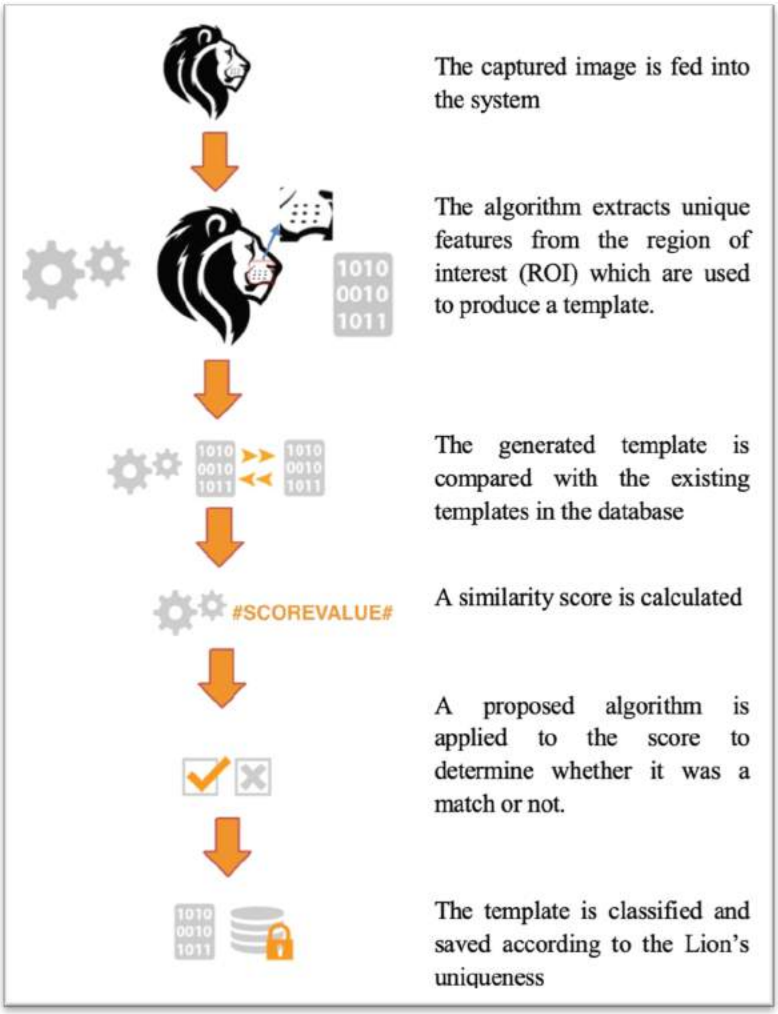


Figure 3.1: Ideation of the identification template generation process

Started as a project to digitize the process of storing and retrieving lion identities from photographs of lions.
The software was supposed to be an upgrade to the current system that identifies these lions by hand using various techniques of mapping the whisker spots by using the convention of localizing the spots of the topmost row with respect to the second and third rows within the whisker spot patterns. Additional identification data such as nicks on the ears, location of scars on the face, and even GPS coordinates were being used.
Hence the upgrade that was proposed was to have a software that could be used to record a more granular reading of the whisker spot locations by allowing the user to mark the spots on the image within the software.
The suggested method to store individuality was to calculate distances and the respective distance ratios so that there may be parity between different images of the same lion having various pose variations [136] [137].
The issues faced at this stage was due to the fact that the distance ratios were not able to have parity for the same individual, hence it was suggested to superimpose [159] the image of the lion whiskers over the 3-D curvature of a 3-D model of a lion face and then calculate the ratios within the stretched (flattened or warped) representation of the 3-D model [109] [156].

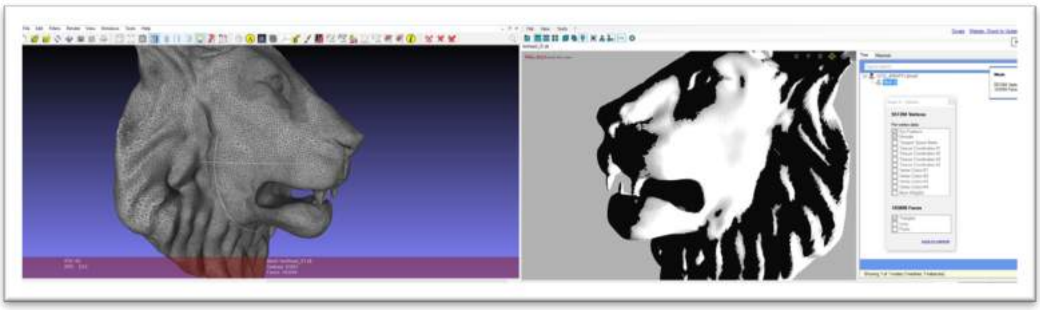


Figure 3.2: Lion-Face 3D mesh manipulation towards texture mapping lion photographs

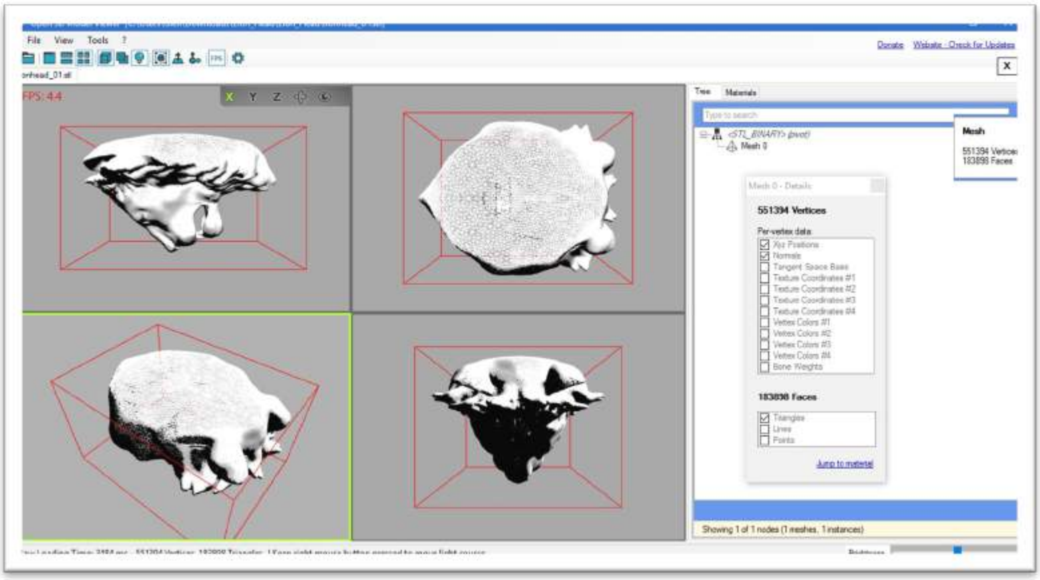


Figure 3.3: Lion-Face 3D mesh manipulation towards texture mapping lion photographs (multi-perspective)

3.2 USER INDUCED INCONSISTENCIES

The issue of the range of variation in the location clicked by the user to represent a single whisker even when a user was asked to mark the same image twice also induced inconsistencies.

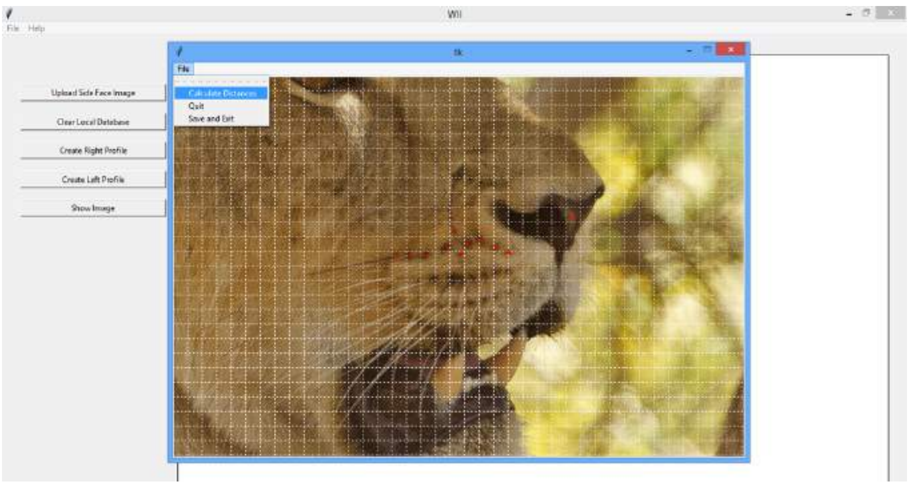


Figure 3.4: User induced inconsistencies during manual whisker marking

All of these techniques were still manually intensive and required a lot of involvement of the user. Thus, it was realized that it was necessary to approach this problem with smarter algorithms that used machine learning approaches to combat such inconsistencies. Hence due to this change in approach we now required the insights of that came along with including the subject of image processing to the problem at hand [147]. This is possible when the images are processed in such a way that a 'perspective' be created with the data that lies within them within them, which in turn can be learned by the machine.

3.3 VARIOUS TECHNIQUES TOWARDS FEATURE EXTRACTION BASED MODEL CREATION

As seen in the previous section, a requirement is imposed of defining features that can be given a specific data representation for a population of similar individuals with the ability to separately identify individuals within this population space [103] [104]. To achieve this by the machine learning pathway, a large dataset is required, this data must be good enough to be used for training such a feature extractor as well to be able to train the model that can tell individuals apart [122].

The first approach was to use the SIFT/SURF algorithm [121] to find the blobs and eventually the centroids of the whisker area of the lion face.

This was the initial idea of automation, but this still carried over the issue of pose variation and the image processing complication of finding the right amount of thresholding for the pixel intensities for the image to make it easier for the blob detection algorithm.

While experimenting and fine tuning this approach, it was observed that due to the thresholding techniques used, there were non-whisker areas also appearing as blobs. These 'stray' blobs (Figure 7) which triggered the blob detection algorithm eventually became noise as the centroids of these blobs were also taken for the distance and distance-ratio calculation. Along with this noise being induced into the system there was also the notion of the loss of data during the process of binary thresholding, this data lost could contain information about the 3-D structure like folds/bumps and other shapes. It can also be said that the data thus lost due to binary thresholding haunts us back in the form of noise of those ‘stray’ blobs. Finally, this method to automation did not produce good results, hence further research into different methods to approach this problem was initiated.

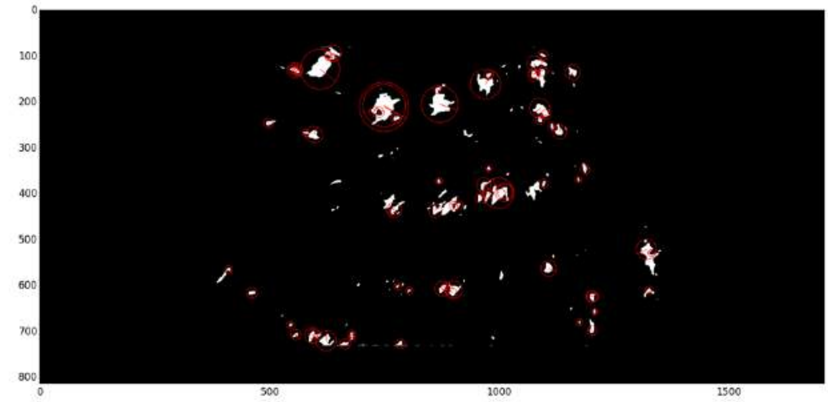


Figure 3.5: ‘Stray’ blobs picked up by the blob detection algorithm due to acute binary thresholding

The next approach was to use Machine Learning techniques to do the heavy lifting for each of the problems faced to go from input image to the output identity, by solving one problem at a time. The first being Pose Estimation of the Lion-Face within the input image. This is the most basic classification task that reduces the ambiguity between the different face orientations possible within a lion photograph.

The first approach within this step was to accommodate all the different poses that may be possible. Following the implementation of the research conducted by [69] it was suggested to combine different models together into a unified single model. The models combined together are: a shape model (the point cloud of the key points within a training set of lion faces), a motion model and finally an appearance model [130]. The resultant unified model was called the Active Template Model.

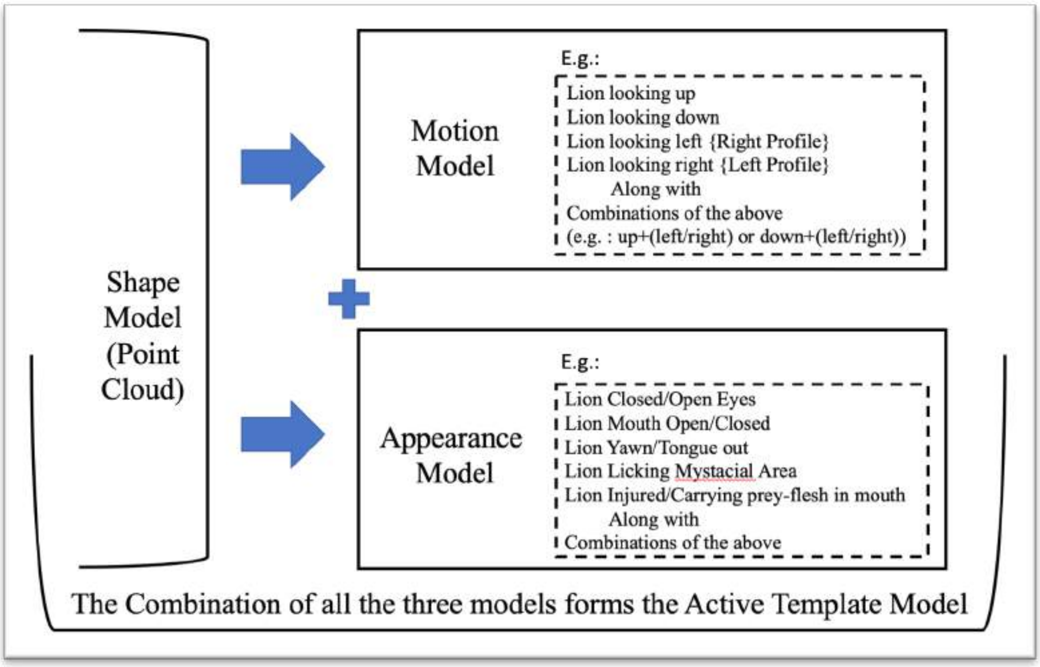


Figure 3.6: Active Template Model

The attempt at arriving at a basic level of automation was the envisioning of an Active Template Model. An Active Template Model, is in effect, the combination of 3 models, that of Shape Model, Motion Model and the Appearance Model [126] [127] [162] [164] [169] [186] (Figure 3.6). The most essential of the three is the shape model which is a distribution variance model of points within a train set point cloud that represents landmark points of a set of annotated images. A set of annotated images with a fixed number of landmark points each will provide as a training set for the shape model. Therefore, each Image in terms of L landmark points can be represented as:

$$s = (x_1, y_1, x_2, y_2 \dots, x_L, y_L)^T \tag{3.1}$$

x and y being the cartesian coordinates of each point in the set of landmarks. Each Landmark point must represent the same location point within the shape instance for each image. For example, taking the case of a 2D face, if a landmark point (x_6,y_6) within a shape instance of an image represents the location of the right pupil, every instance must have the landmark point (x_6,y_6)

to represent the right pupil of the face. Hence all instances have a fixed number of landmarks.

For a 2D image of a known, rigid 3D object, the goal is to find the parameters (position and orientation) of the model which best explain the projection observed in the image [180] [181]. In general, additional parameters are required to account for variability in the object itself [133].

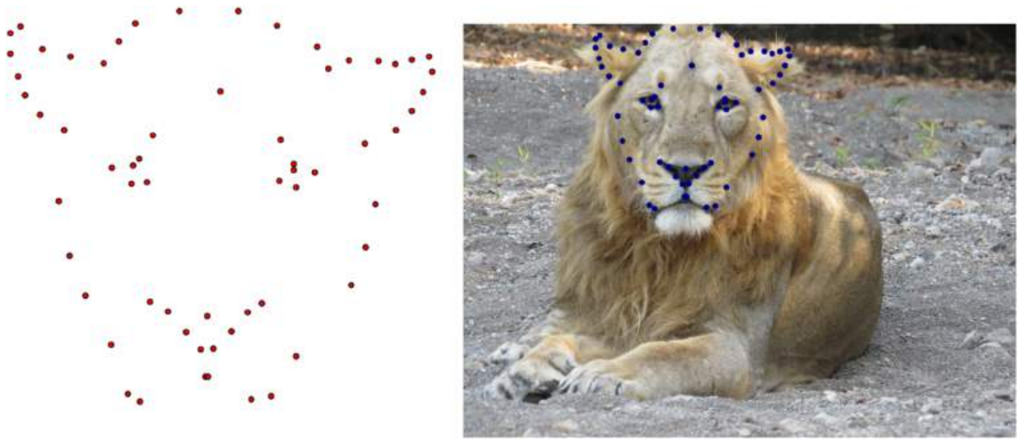


Figure 3.7: Point cloud showing the 63 landmark points of the Active Shape Model

The most relevant approach to finding the parameter values is to extract image primitives and solve the combinatorial problem of establishing a correspondence between these model primitives [150] [166].

The assumption that a robust mechanism exists for extracting well-defined primitives often does not hold especially in the fields of wildlife photography that do not have good datasets that aid in the formation of such models [148].

Active Template Model (ATM) is such method which is inspired by the Lucas-Kanade Affine Image Alignment and the Active Appearance Model [173] [134] [144] [158] [183].

A template image is taken as a seed for the creation of the model along with the list of all the training shapes. The training shapes are formed by the point clouds provided by the training dataset [174] [128] [151]. The reference shape is then created by inferring the holistic features that are extracted before warping the images to the reference shape [129]. The purpose of the reference shape is to normalize the size of the training images.

The normalization is performed by rescaling all the training images so that the scale of their ground truth shapes matches the scale of the reference shape [131] [193].

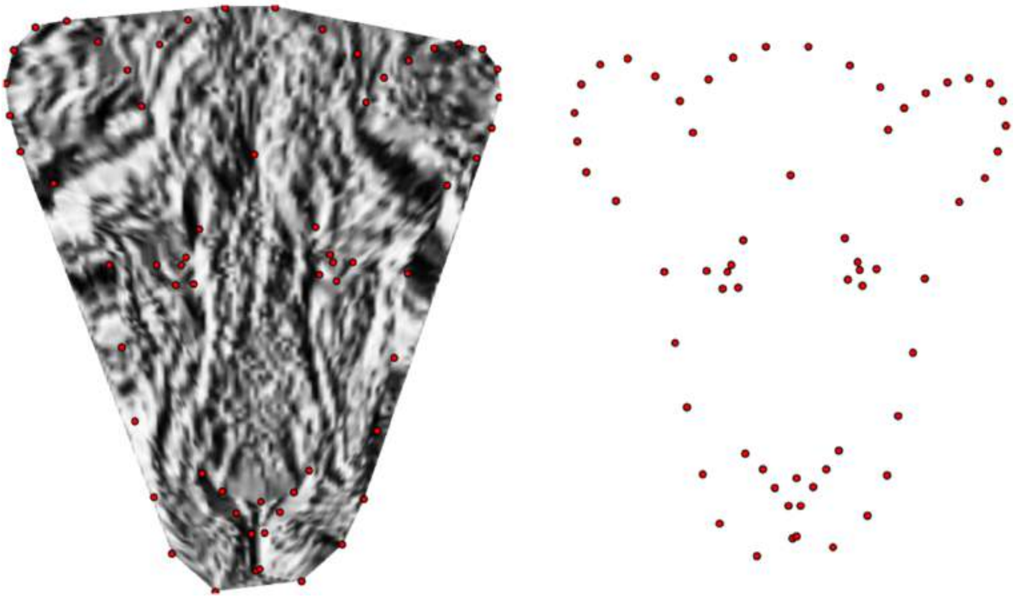


Figure 3.8: Interpolated results with thin plate spline: A visual representation of the Active Template Mode

Moving towards an approach to in encapsulate individuality [161] efforts were taken to initially train a model to learn the location of the whisker spots as key points with respect to the main marked facial key points, which are 17 arbitrarily chosen key-points that represented the major facial landmarks of the lion [123] [171] [172] [185]. The facial regions were mapped with bitmap classes to classify pose variations (Figure 3.9).

The approach of using fixed key-points [98] [99] [100] across the images for the whisker points proved to be totally infeasible as there is no fixed agreement across the photographs on the number of whiskers that may be present on each lion's muzzle.

Simultaneously in other tests, the yolo algorithm performed superiorly compared to the algorithm trained over the 17 facial key-points [153].

Therefore, we went ahead with the yolo algorithm that identify the correct pose of a given lion face [168].

After the step we needed to create the model for recording individuality and to achieve initial results, it was decided to use only the right profiles for the different lions.

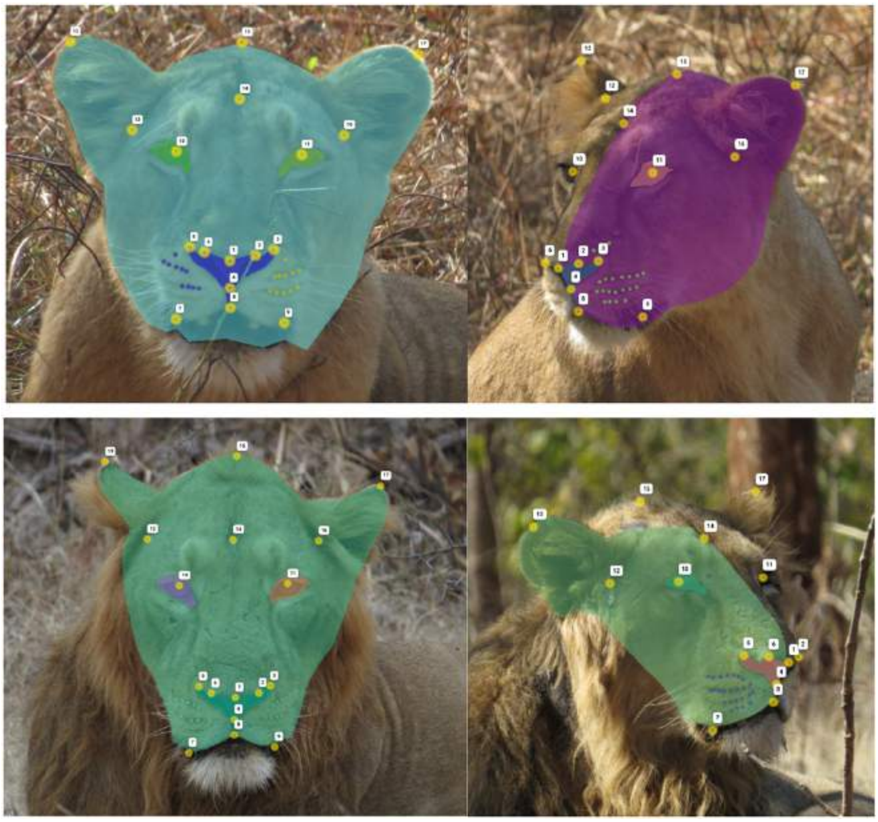


Figure 3.9: Landmarking Along with Bitmap Face-Pose Classes

An initial test of training a simple classifier with various individuals as different classes using the yolo algorithm was performed. The Gaussian Binary Threshold technique was used (Figure 3.10) so that the variance caused by the difference in lighting and the issue of various colour differences could be reduced [155]. The added benefit perceived was the fact that all the images could have only the patterns stored in a binary format.

The results were poor with the classifier not being able to identify the various individuals without error.

Figure 3.10: Gaussian Binary Thresholding Technique with Bitmap regions for individuals as Classes

Furthering the efforts, towards characterising the appearance of the lion face within the photographs, a classification approach using deep learning with the YOLO architecture is tested over the initial dataset containing less number of sample images.

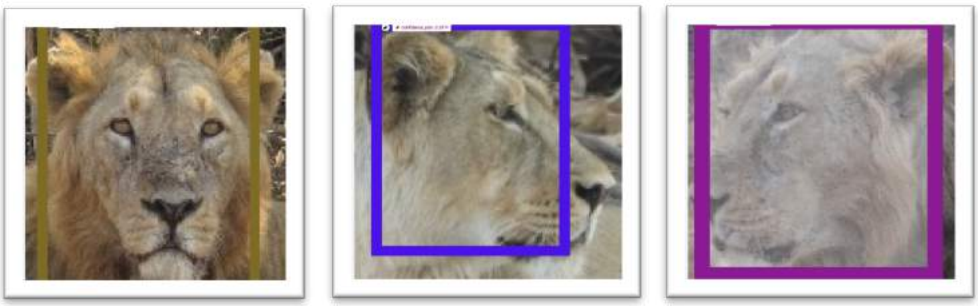


Figure 3.11: Front pose, Left pose and Right pose classification with YOLO

A limited number of 3 classes are targeted, that of, the front pose, the left pose and the right pose. The dataset is then annotated according to the three categories. Once the model is created, the accuracy of classification is thus tested. Due to limited data the results are not the best. Although, this classifier can be considered for its barebones functionality, as the accuracy score is better once the confidence threshold is lowered. A few examples of the model trying to estimate the correct pose is shown in Figure 3.11.

3.4 IDENTIFYING VARIOUS CONSTRAINTS FACED

Through the various techniques explored in the aforementioned sections different bottlenecks and inconsistencies have been faced and multiple solutions to tackle them are proposed by varying degrees of effectiveness in implementation. These inconsistencies and inadequacies in the nature of the dataset including the issues with the images contained in these datasets, are covered in the undermentioned subsections. Following this, the next section gives a brief outlook on how these constraints may be resolved.

3.4.1 THE SIZE OF THE DATASET

This is an account of the various datasets procured, during the period of analysis and thus covers its inadequacies. The initial size of the dataset, was that of 2 Lions with an unequal number of photographs for each Lion, the first having 72 photographs and the second having 5 photographs. The size of the next dataset was of 6 Lions with 3 photographs each. Each lion had a front facing photograph, a right facing photograph and a left facing photograph. There was no significant variation. Each profile for the side of a lion face had only single examples each to train on.

A third dataset procured had 224 images of unidentified lions, with various orientations. This dataset was cleaned of the noisy and unclear photographs, by letting go of 90 such photographs resulting in 134 photographs that were then divided into 46 left images, 53 right images and 38 right images so that a classifier can be trained to identify, if a Lion face image was a left profile image, a right profile image or a front facing image.

Various challenges in the current process:

- Considerable variation of the position, orientation and viewing angle.
- Occlusion of the facial boundary by the Lion's mane and the orientation of the head.
- Noise and artefacts.

Various annotation techniques [92] were also implemented, such as key-points mapping annotation, polygon annotations and bitmap annotations. The only outcome of all these efforts were the successful training of the classifier that could differentiate between left, right and front facing images.

3.4.2 QUALITY OF THE IMAGES:

There was a high variance in resolution ranging from that of 4000x3000 pixels to as small as 300x300 pixels. This was normalized by cropping out the region containing the lion faces and creating two datasets of 256x256 and 128x128 pixel for training.

Other issues with the quality of the images were that of the obstruction of view for the lion faces in the images, such as the occlusion by the leaves, branches and twigs of the various plants and trees. Such images were considered only if the whisker region, and other prominent features like the eyes and the nose were visible, and the remaining unclear images were discarded.

There were also images that were blurry, out-of-focus, or even shaken, rendering them as unusable for this research work. As it needs a set of images with high resolution, as we need to focus on the whisker spots of the lions for identification.

3.4.3 SIZE OF THE REGION OF INTEREST WITHIN AN IMAGE:

The photograph of the lion may be captured with a huge distance from the camera, resulting in a very small representation of the lion face within the photograph. Thus, the location of the lion face in an image poses to be a concern as the percentage of image real-estate covered by the lion face in turn affects the quality of the region of interest that contain the whisker spots. Therefore, an image quality standard, specific to the visibility of whisker spots is needed to be defined.

3.4.4 DOMAIN SPECIFIC INCONSISTENCIES:

The analysis of the data, uncovers various issues that are unique to this domain. Mentioned below are a few such issues related to the whisker spots as seen within image representations in the dataset.

- Whisker specific issues:
    - Clarity of the whiskers

- Partial lighting of whisker areas
- Scars
- Folds along whiskers
- Flies sitting on the Region of Interest (act as whisker spots)

These issues prove to be unique problems to this research and thus need to be addressed in a wholistic fashion.

3.4.5 INFRASTRUCTURE CONSTRAINTS:

The initial infrastructure used, was a machine equipped with the NVIDIA GeForce GT 740 graphics card with 384 CUDA (Compute Unified Device Architecture) cores available. This available resource is a local device provided on campus at the time period. This poses to be a limitation to the RAD (rapid application development) cycle, as the testing and evaluation phases for each iteration takes longer. Therefore the overall time taken to evaluate the models that produce valuable results increases exponentially. Therefore various cloud based solutions with higher configurations are in favour to be explored.

3.4.6 POSSIBILITY OF CORRELATION:

Every lion individual has unique whisker spots on both the right as well as the left faces of their muzzle. From the data it is also inferred that there might be a possibility of correlation between the right and left whisker patterns. Therefore to avoid inducing ambiguity into the system, there needs to be a system that negates the effects of this possibility of correlation.

3.5 HANDLING THE CONSTRAINTS

Each of the constraints recorded in the aforementioned section are analysed and various solutions towards handling these constraints are discussed in this section so as to resolve the same.

3.5.1 SIZE OF THE DATASET:

Towards solving the issue of having less data, a new and more organised dataset is created by including more images from various sources, like the inclusion of African Lions, from the online resource that hosts lion photographs of the Masai mara project. An avenue that has the ability to increase the size of the dataset is to employ data augmentation by various data transformation techniques to generate synthetic images. Various data transformation techniques specific to images, like rotation, stretch, compress, shear, gaussian blur and gaussian noise are good candidates of techniques that can be implemented over the current dataset to increase the size of the dataset through augmentation. Details of the various augmentation techniques implemented are elaborated in Chapter 6.

3.5.2 QUALITY OF THE IMAGE DATA:

Cleaning data is a very crucial to the creation of an effective and accurate model. Therefore if there exists images within the dataset that do not add valuable information towards the training of the model, the training process may introduce noise and errors into the model being created. Therefore, blurred, shaken and occluded images are removed from the dataset to avoid noise to be induced into the system. This procedure is manually conducted so as to insure that the standard of quality is preserved.

3.5.3 RESOLUTION SPECIFICATION FOR THE REGION OF INTEREST:

There are various possibilities by which the lion face may be captured within a photograph. Most of the captured images are not usable due to the lack of clarity and resolution at the region of interest. A resolution specification for the quality of the images does not solve the problem for such lion faces captured at a far distance from the camera. Hence along with a quality specification for an image captured, there is a requirement of specifying the capture distance or the optical zoom distance from the camera to the lion face. Thus, the quality description as per the requirement of this research is defined by the constraint that each whisker spot must be described by at least 10 individual pixels. This creates a single quality parameter that takes into account all kinds of quality inconsistencies like zoom, resolution, distance of lion face from the camera, etc., by specifying the quality required for a single whisker. This is effective, as the relevant information required for training the model resides within the region of interest, that contains the important patterns for individuality created by the whiskers, requiring them to be clearly visible, whatever the original resolution may be.

3.5.4 DATA NORMALIZATION:

By the nature of the problem at hand it is now evident that there are various multidimensional inconsistencies that need to be addressed to be able to locate and analyse the region of interest, thereby paving the path towards uniquely identifying lion individuals. Therefore towards this goal of creating models robust to multidimensional inconsistencies like blur, shear, variable lighting, folds, and flies on the whisker regions and many other domain specific issues, deep learning techniques need to be employed [175] [176] [189]. Hence, to enable workflows with deep learning, the dataset is normalized with regard to the nature of the data that is required to be analysed to represent individuality. Out of the various transformation techniques available for the normalization routine, the perspective based transformation technique is most suitable as the lion face is a 3-spatial-dimensional object represented as a 2-spatial-dimensional image. To achieve this, a normal perspective needs to be fixed so as to normalize the whole dataset with such a perspective based transformation. This is implemented by formalizing an algorithm called the **LI-No^2 normalization algorithm** covered in detail in Chapter 5.

3.5.5 INFRASTRUCTURE IMPROVEMENT:

With respect to the requirement of rapid application development workflows that enable iterative analysis of multiple models, it is preferred to take advantage of the computing power available on cloud based platforms. Therefore various platforms like AWS (Amazon Web Services) as well as Google Colaboratory are used to train the different models iteratively with various model configurations. This infrastructure improvement drastically reduces the time taken to evaluate and compare various models with different configurations. The different cloud platforms, had the provision of various NVIDIA accelerators as mentioned below along with their respective number of CUDA cores (Compute Unified Device Architecture):

- Infrastructure used on the AWS Elastic Compute platform:
    - NVIDIA K80 with 12GB memory and 2496 CUDA cores per GPU (Average model creation time 5 to 6 hours)
- Infrastructure used on the Google Colaboratory platform:
    - NVIDIA Tesla P100 with 16GB memory and 3584 CUDA cores per GPU (Average model creation time 3 to 4 hours)
    - NVIDIA Tesla V100 with 16GB memory and 5120 CUDA Cores per GPU (Average model creation time 1 to 2 hours)

Thus there has been an improvement of the infrastructure and can be quantified by the increase in the number of [71]CUDA cores from 384 CUDA (with an average model creation time of 8 to 12 hours) in the initial stages, all the way up to 5120 CUDA cores. Hence providing the advantage of approximately 11 hours per iteration step of the model creation in terms of time saved. These 11 hours gained per iteration, cumulatively turns out to be a huge advantage to the tune of a few 24-hour-days of time saved.

3.5.6 CORRELATION AVOIDANCE:

To avoid the introduction of ambiguity into the model being trained to estimate the lion identities with reference to the possibility of correlation between the whisker spot patterns that occur on both the left and right faces of the lion's muzzle, only the data pertaining, to the whiskers of the right face are selected. Therefore, a system that takes inputs and estimates individuality by evaluating only the whisker images of the right side of the lion face, avoids all problems related to the possibility of correlations.

3.6 SUMMARY

This chapter explores the various efforts undertaken as well as the issues and inconsistencies, towards digitizing the process of recording as well as retrieving information unique to each lion individual. Towards the goal of region of interest localization, the Active Template Model approach, that includes the active shape model along with the active template model, is explored to characterise the different possibilities of appearance for the lion faces within photographs. These techniques are explored by fixing a set number landmark points of the lion face across the whole dataset used along with thin plate splines. Other than these, approaches of landmarking and bitmap labelling are also explored. The chapter also includes the various constraints that are faced during the initial evaluation process of the various techniques for region of interest localization as well as appearance characterization, towards identity estimation. The chapter thereby, concludes with the different methods that may be employed to combat the various constraints. The further chapters within this thesis document record these methods and techniques in greater detail.

CHAPTER 4 PRELIMINARY ANALYSIS OF WHISKER SPOT PATTERNS

4.1 INTRODUCTION

It is well documented within the literature, that the spots occurring on the mystacial region of a Lion's face, have unique arrangements, rendering these spot patterns to be unique for each Lion individual [187] [192] [194]. Hence even the current system for identifying Lion individuals, relies upon taking manual recordings of these spots, as a definition for the identity of each individual

In this chapter we explore the viability of using the whisker spots as an identification mechanism for Lion individuals by manually marking the whisker spots over photographs of lion faces, by using the coordinate values of these spots across samples of different individuals.

4.2 THE CURRENT MANUAL MARKING METHOD

Within the current technique used for the identification of Lions by the vibrissae patterns, a data sheet [70] for each individual is created manually by examining photographs of the specific individual and then recording the information onto a standard template as shown in Figure 4.1. This type of calibration of the vibrissae patterns even allows for on-field recording of the identification information onto a standard graph.

There are various limitations and prior requirements for this identification mechanism. The need of a trained expert to conduct these recordings is a primary requirement of the procedure. Providing the required training or recruiting trained experts adds to the overall cost of the identification process. The manual recording process also comes along with various limitations like that of ambiguities across the recordings of multiple experts for the same lion individual, as well as the scope for human error in the recorded information. There is also a partial reliance on various nuances of data recorded within the graph used for the calibration, such as the closeness of a marked spot to the edges of a cell within the graph. These nuances, may be easy to be picked up by human domain experts, but suffer the risk of being lost in translation, across multiple representations of the same data.

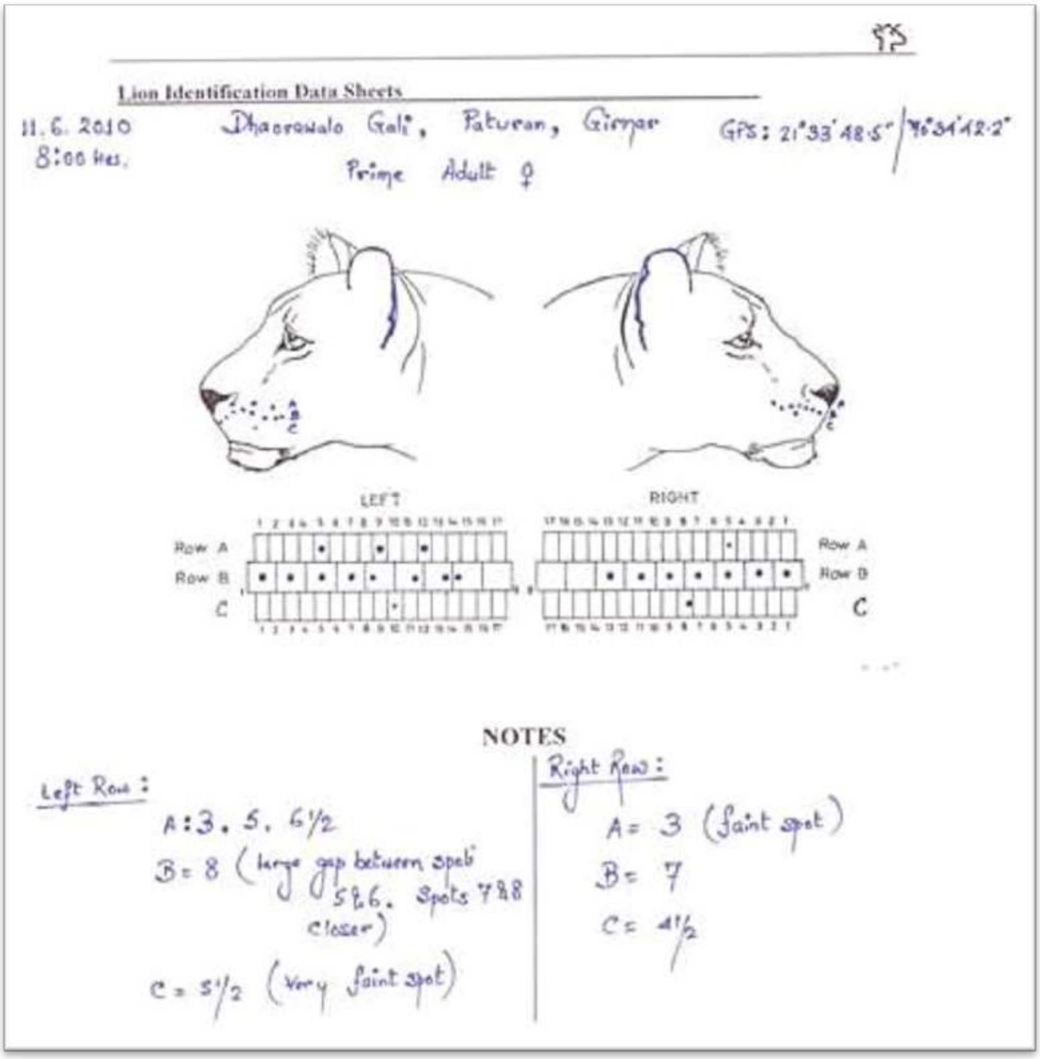


Figure 4.1: Data sheet Performa showing the calibration of the vibrissae patterns onto a graph [70]

Efforts have been taken to digitize this process, leveraging long term database storage and faster search capabilities for the identity information of known lions. Even with these benefits, the shortcomings of the manual marking have been carried over to the digital system, as the recording process is unable to capture the fine grained nuances. This digital rendition of the procedure also heavily relies on the expertise of the user to be able to figure out the accurate representation of the vibrissae patterns based of various viewing angles of the photographs. Hence the accuracy of representation still lies in the hands of the expert. Such a system also lacked the capability to categorise similarly marked individuals into a single identity to represent the same individual, which also leads to the possibility for the creation of multiple identities of the same individual owing to slight variations in the recorded information.

Keeping these shortcomings in mind, we proceeded to develop a semi-automated technique that creates a model for the categorization of the identities of the Lion individuals.

4.3 SEMI-AUTOMATED LION IDENTITY CATEGORIZATION

The creation of the semi-automated system for the lion identity categorization had a few goals that needed to be targeted:

- Making the manual marking system easy for even non-experts.
- Making an interface that handles the ambiguity during the manual picking of coordinate locations of vibrissae spots.
- Creating a representational template pertaining to the Lion's identity from the information generated within each photograph.
- Creating datasets of the representational templates to generate classifier models for the different lion identities, thereby avoiding double identities for the same lion.
- Validating the technique of using vibrissae spots as an identity marker for lions.

Targeting the various goals as outlined above [140], a tool has been developed, which first allows the user to load and crop out the whisker region from a lion photograph. The tool then allows the user to manually click on the whisker spots within the cropped whisker region. To make the process of locating the whisker spots easier, the cropped image is also transformed to a binary image within the interface using a dynamic thresholding technique, thereby reducing the ambiguity. Ambiguities pertaining to the coordinate location of the whisker spots derived by the user-clicked locations are handled by an 8-neighbour region growing technique and calibrated to the centroids for each whisker spot. By this process the coordinate locations of each whisker spot within a particular photograph is recorded.

Distances between every whisker coordinate location is calculated to produce a normalized representation of the vibrissae spot locations thereby creating representational templates that encompass the features of identity. Subsequently, the ratios of the distances between every two pairs of vibrissae spot coordinates are calculated. The angle in radians between pairs of vibrissae spots are also calculated as an added feature. The information thus generated serves as a representational template for a Lion individual within an image. The software tool generates a CSV file recording this information for each image.

A dataset is then created by combining all these representations into a single file. The data imputation process is handled by the method of filling zeros. This dataset is then used to generate various classifier models. The accuracy of these models thus validates the technique of using whisker spot patterns to identify individual lions.

4.3.1 INTERFACE TO RECORD VIBRISSAE SPOT LOCATIONS

To make the process of vibrissae spot selection intuitive and easy to use, the user is presented with an interface, that is capable of panning and zooming so that the user can crop the whisker region from the photograph of the lion. After this process the user is presented with an interface that shows a binary image of the cropped region. The user is required to click on all the visible vibrissae spots, on this panel, allowing the software tool to record these coordinate values so that a representational template may be generated. Using the coordinate points of the whisker spots eliminates the need of an expert user who understands the conventions used previously in the technique of calibrating the whisker spots to a standard graph.

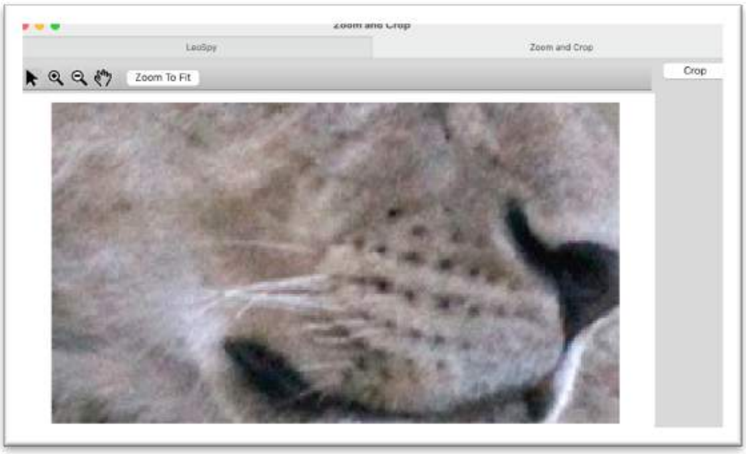


Figure 4.2: Region of Interest selection within whisker coordinate distance and radian calculation tool

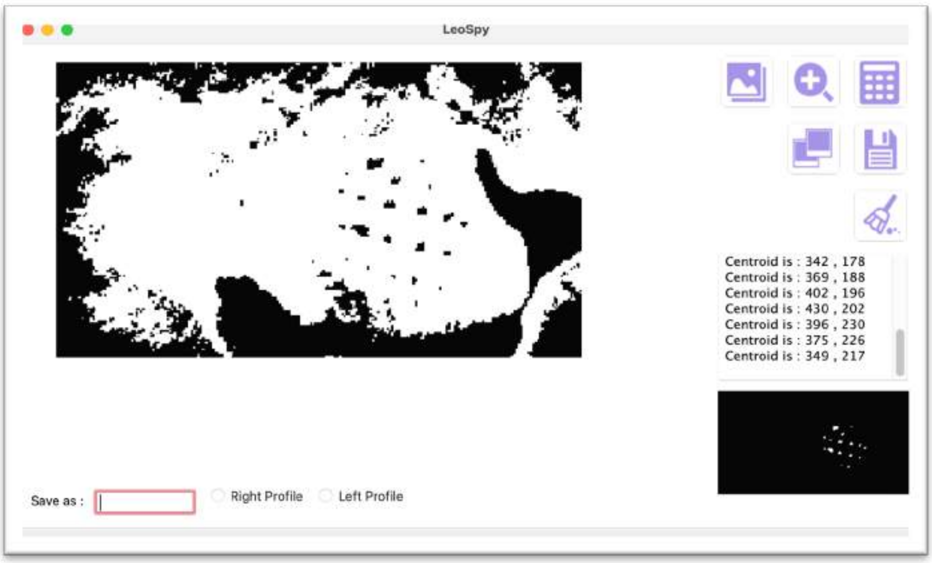


Figure 4.3: Interface of whisker coordinate distance and radian calculation tool showing thresholding, points clicked with region grow, as well as centroid coordinates

Binary Image Thresholding:

To generate a binary image displayed with in the interface, a dynamic thresholding algorithm is used so that the darker regions are clearly differentiable to the lighter regions. Various other thresholding techniques like multi-channel thresholding [124] were tried out before finalizing on the dynamic thresholding technique used (Figure 4.4). Finally, the Gaussian dynamic binary image thresholding is used (Figure 4.5). This produces a clear binary image where the dark portions are distinctly visible, aiding the user to click on the whisker spot locations. This also helps the region growing algorithm to estimate the region of each whisker spot.

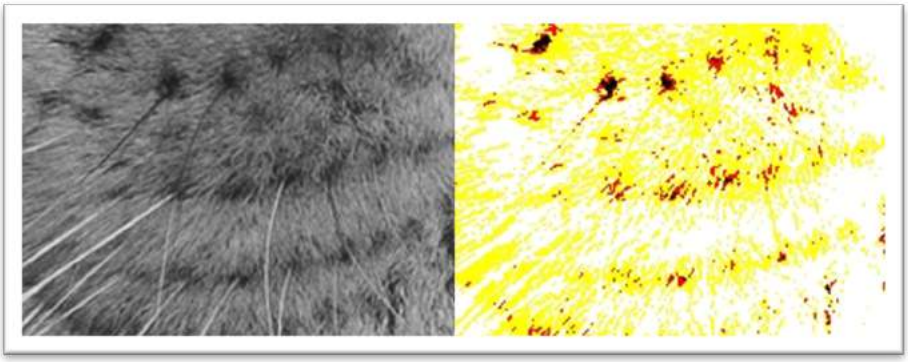


Figure 4.4: Region of Interest: Grayscale image vs Multi-channel thresholded image

Figure 4.5: Gaussian Binary Thresholding

8-Neighbor region growing Algorithm:

To reduce the ambiguity in the process of selecting the coordinate locations of the whisker spots, an 8-Neighbor region growing algorithm was used. At this stage, the interface displays a binary image on a panel allowing the user to click on all the whisker spots that are visible. Since there is a variability in the coordinate locations generated by the user-clicks for each whisker spot, there is a requirement to create a standard to pick these coordinates. A feasible standard in this scenario is to use the centroid coordinates of each whisker spot. For a single whisker spot, its centroid can be calculated if the whole region for that whisker spot is known.

(x-1,y+1)	(x,y+1)	(x+1,y+1)
(x-1,y)	(x,y)	(x+1,y)
(x-1,y-1)	(x,y-1)	(x+1,y-1)

Figure 4.6: 8-Neighbor Filter

The 8-Neighbor region growing algorithm traverses through the neighbouring pixels using an 8 Neighbour filter (Figure 4.6) till a set threshold difference is met, thus estimating the region of each whisker spot (Figure 4.7).

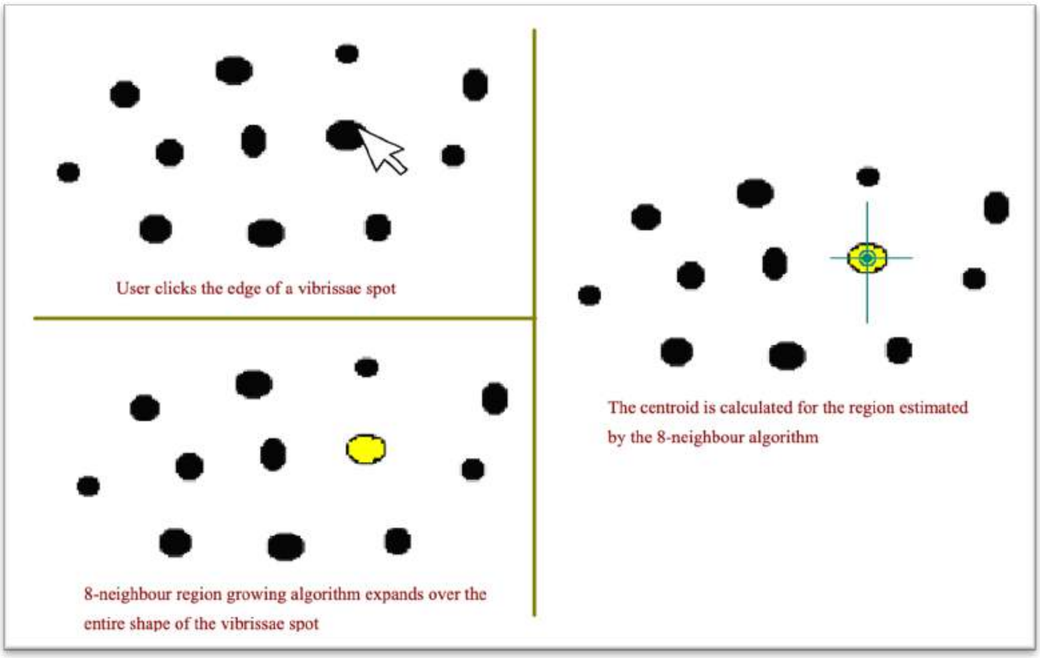


Figure 4.7: Centroid selection using the 8-neighbour region growing algorithm

The threshold difference within the region growing algorithm is evaluated from the value of the pixel at the user-click location. Hence the region grows from the user-click location and stays within the contrast boundary of that whisker spot.

4.3.2 IDENTITY REPRESENTATION TEMPLATE

After the user clicks on all the whisker spots and the centroid coordinates of each spot is recorded, an identity representation template needs to be generated. Such a template will help in the matching process when a similar pattern is found or to search for a pattern with in a known set of patterns. To generate such a template, we require to record the information in a normalized fashion such that it remains invariant to change like scaling, rotation, etc. Thus, using the absolute coordinate locations recorded by the interface the software tool is made to calculate the distances between every point in an all to all fashion. Furthering this step, the ratios of pairs of distances between every two pairs of points are calculated. These ratios are robust to various representational changes (Figure 4.8). Thus, the spot arrangement of a new image recorded with this template of distance ratios can directly be compared with representational templates of known Lion individuals and a match can be found [145]. Along with this method of representing spot patterns, recording angles in radians was also considered. Such a template for each image is generated with a few sample images representing each Lion individual.

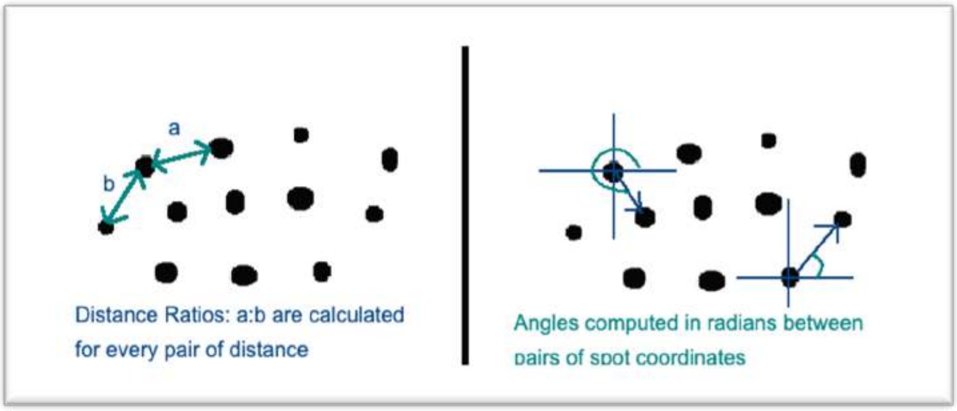


Figure 4.8: Calculation of Distances and Radians

Multiple representational templates for a set of ten lion individuals is generated. Every lion individual had multiple sample images, each of which are marked thrice. Each image is marked thrice with a different (randomly selected) starting vibrissae spot each time, generating three representational templates that contain the spot pattern information within a new but similar representation. Thus resulting in multiple templates for each lion. The tool is designed to generate a file in the CSV format, that stores a record of the calculated distance ratio values along with the angles in radian values.

4.3.3 IDENTITY CLASSIFICATION USING MACHINE LEARNING APPROACHES

To create various classification models that leverage the information contained within the patterns stored in the form of distance ratios, a consolidated dataset needed to be created. The information of each image is stored in separate files in the CSV format by the software tool. These separate files are thus combined together to create a single dataset file. The marked number of vibrissae spots for each image is variable even across the images of the same Lion. This is handled by filling zeros at the locations of missing values, such that the dimensions of the information is fixed across the whole dataset at the time of consolidation. Thus, a consolidated file in the CSV format is created that contains the information of the various lion identities across multiple images represented by the distance ratios as well as the angles in radians computed over the marked vibrissae spots. An initial CSV file is generated containing both the information of the distance ratios as well as the information of the calculated angles in radians. Two separate CSV files are generated from the initial file, these files contains the respective information of the ratios and angles separately.

Hence, there now exists three consolidated datasets with a slight variation of the information contained within them, namely, the first dataset containing the combined data of the distance ratio information as well as the information of the angles in radians, the second dataset containing the information of distance ratios only, and a third dataset containing only the information of the angles in radians.

The model training is carried out with the datasets split into two parts that of training and validation [141]. Entries into each of these two splits are picked at random within every dataset. Various classification models were trained using the training split of the three datasets. The accuracies for every model were calculated against the validation set (Figure 4.9).

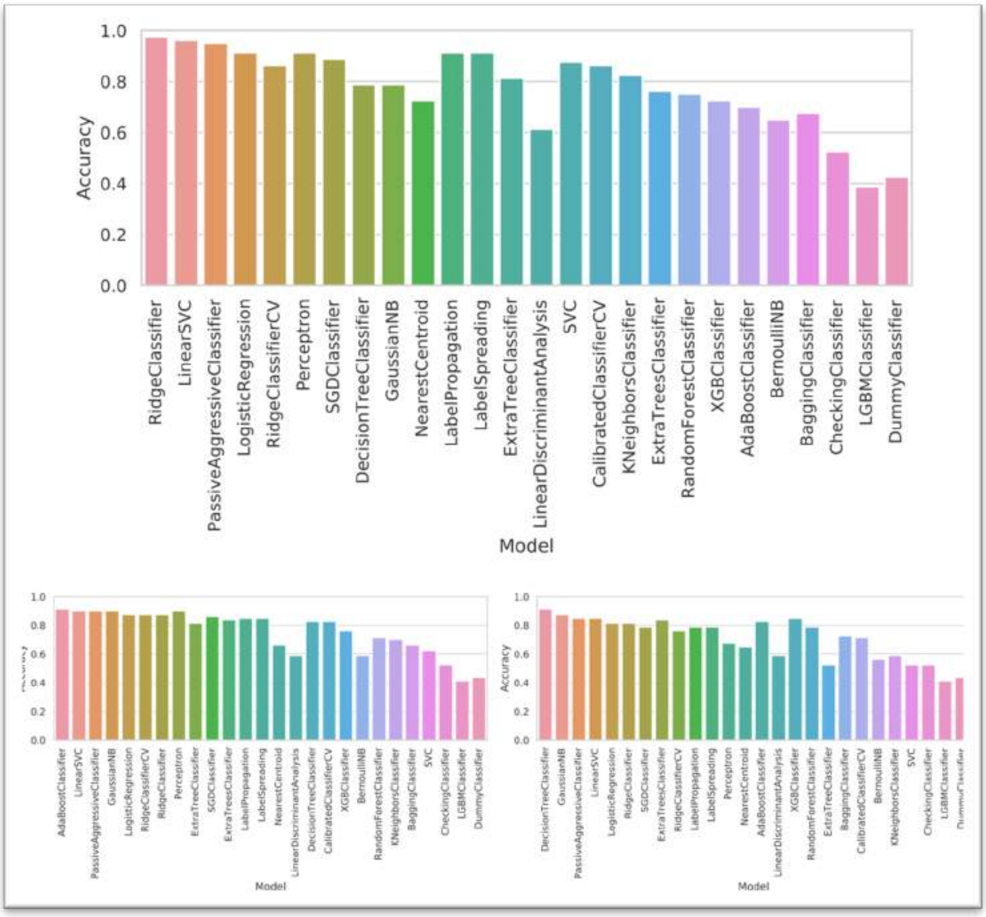


Figure 4.9: Accuracies for every model calculated against validation sets

Many of these models crossed an estimation accuracy threshold of 80%, proving the existence of uniqueness within the information to the patterns of the vibrissae spots of Lion individuals. Hence, these accuracy results on the various classification models shows the existence of learnable unique patterns within whisker spots.

These observations thus validate the furthering of the research work to move towards the identity estimation by computer vision approaches, which is expected to tackle the issues of ambiguity created by human markers. Furthering of the research work in the direction of computer vision approaches is covered in the following chapters.

4.4 SUMMARY

In this chapter, a preliminary analysis of whisker spot patterns has been elaborated on, covering various aspects, firstly, the conventional techniques used in the current system of Lion identification and secondly, the creation of the semi-automated system for the estimation of Lion identities using information of the vibrissae spot patterns located on the Lion's muzzle. This helped to reduce and normalize human errors. While creating the structured dataset for the whisker data, the preferred normalization technique to denote the absence of whiskers as well for the normalization of the length of each data entry, adding zeros to the data was chosen. A comparison of multiple classification models have also been evaluated pointing towards the existence of uniqueness within the vibrissae spot patterns. As a future scope the Groth's Algorithm (Triangulation of stars in astronomy) [46] can also be implemented, creating the triangulation by using the computed angles in radians, thereby adding to this vector analyses to perfect this proof.

CHAPTER 5 DEEP MACHINE LEARNING APPROACH TO LOCALIZE REGION OF INTEREST AND TO CREATE AN EMBEDDING SCHEME FOR INDIVIDUALITY

5.1 INTRODUCTION

Convolutional neural networks (CNNs) are able to train multiple filters of various configurations stacked in layers, such that each layer becomes capable of recognizing patterns of different complexity [138]. Deeper convolutional layers are capable of having their filters trained such that based on the inputs from the previous layers they are able to identify more complex patterns. [8, 9] This capability of convolutional nets is a huge advantage considering the problem at hand [95]. Given sufficient training data, the trained filters of our model are able to isolate the patterns of interest and separate them from the unwanted noise [13]. Thus, being able to extract desired features defined by the loss function as well as our training data with high efficiency.

The convolutional approach is used to tackle the requirement of object localization and detection as well [3]. Object localization is done by including the bounding box information in the output vector for each element in the training set that contains lion faces. Subsequently, this process is replicated of the whisker regions as well (as shown in Figure 4 and Figure 5, below in the methodology section). This enables the prediction of the CNN to use the loss generated from the provided ground truth to increase the accuracy of the predicted bounding box information. The Object detection is done by scanning portions of the image by a sliding window mechanism and then localizing for the presence of the lion face and the whisker region. This sliding window mechanism is optimised by handling these separate windows into one step by passing the whole image into the convolution architecture defined for training. Thus, just the lion face or the whisker region size incorporated is created as a viewport. Hence the size of the window is the size with which the lion face or whisker spot location is trained on.

5.2 THE YOLO ARCHITECTURE

Given any input image larger than this size, the convolutional output resembles a sliding window output within a single step. This increases the output volume but in turn gives us the results for the location of the lion face or the whisker region in one step as described by the YOLO (You Only Look Once) architecture [12].

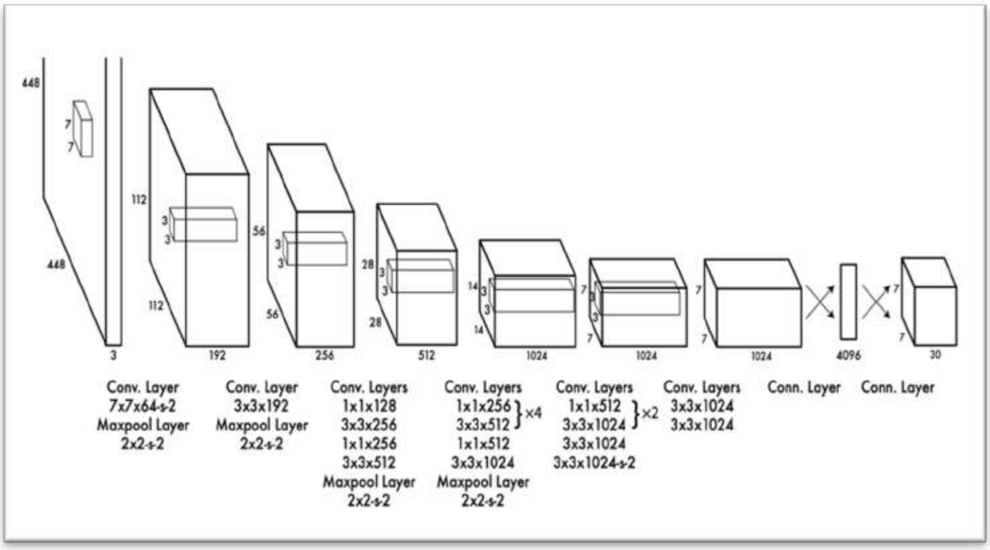


Figure 5.1: Representation of the convolutional layers within the YOLO [12] architecture

The network architecture of the YOLO detector has 24 convolution layers along with 2 fully connected layers. The convolutional layers learn the unique patterns of the classification problem presented to the network and the respective maxpool layers reduces the parameter count passed down to a deeper layer. The fully connected layers are used to flatten the convolutional outputs and then reduce it to the format of the truth representation that has the bounding box information. Given below is a brief representation of the network architecture (Figure 5.1).

The YOLO architecture also handles multiple predictions by thresholding the Intersection over Union (IoU) calculated for each repeated prediction by using non-max suppression over the confidence of each prediction. The YOLO architecture is thus used for our Lion face detector as well as the whisker region detector as the most optimized approach. Described below is the loss function described by the authors of the YOLO architecture (Equation (5.1).

$$\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\,[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2]$$
$$+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\,[(\sqrt{w_i}-\sqrt{\hat{w}_i})^2$$
$$+(\sqrt{h_i}-\sqrt{\hat{h}_i})^2]+\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\,(C_i-\hat{C}_i)^2$$
$$+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{noobj}\,(C_i-\hat{C}_i)^2$$
$$+\sum_{i=0}^{S^2}\mathbb{1}_{i}^{obj}\sum_{c\in classes}(p_i(c)-\hat{p}_i(c))^2 \tag{5.1}$$

In the YOLO loss function (Equation (5.1), *x* and *y* are the coordinates to the centroid of the anchor box, $\hat{x}$ and $\hat{y}$ being the predicted values. Similarly *w* and *h* are the width and height of the anchor boxes along with their predicted values. They are under square root to make predictions for smaller objects more precise . *C* denotes a score of the presence of an object or not. The value of *p(c)* is the classification loss summed over all the classes. To train the network in a reinforced fashion, respective masks for the presence and absence of objects are embedded, $\mathbb{1}_{ij}^{obj}=1$ when the object exists and $\mathbb{1}_{ij}^{obj}=0$ when there is no object inversely, $\mathbb{1}_{ij}^{noobj}$ is 1 when there is no object and 0 when there is an object. S^2 represents the total number of grid cells the input gets divided into. The λ values represent constants with higher values for coordinates (λ_{coord}), to have more focus on the recognition problem. Finally B is the number of anchor boxes defined. The loss is computed for each of the cells within the grid S^2 specified.

5.3 THE INCEPTION NETWORK

To perform individual identification, the capability of learning various patterns and variations at multiple scales is the top requirement. An Inception network is of the best interest as it encompasses multiple convolution filter configurations that are channel concatenated to a single layer (as shown in Figure 5.2). This structure is then repeated over 22 layers [10]. A single inception module is described in the figure below (Figure 5.2).

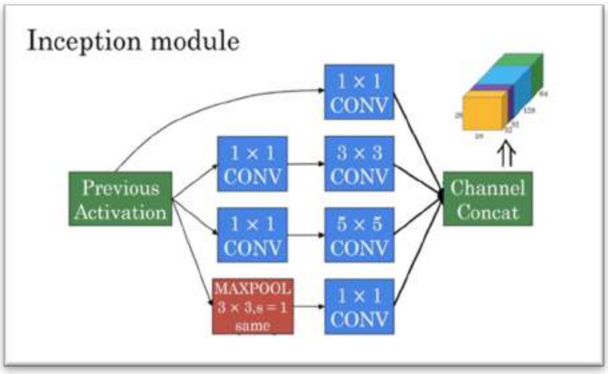


Figure 5.2: The inception model: the constituent convolutional filters within a single layer of the inception network

The inception network is capable of producing embeddings of reduced dimensions for each individual by training with loss functions over this low dimensional (finite) embedding space.

5.4 BUILDING THE ARCHITECTURE TO CREATE UNIQUE EMBEDDINGS FOR EACH LION INDIVDUAL

The identity estimation process is achieved by identifying the uniqueness within the whisker spot patterns over the mystacial region of the lion-face's muzzle area [2][7][9]. This system thus extracts this uniqueness and is able to cluster similar patterns within the embedding space of the learned model.

To create such a system, supervised datasets of labelled examples had to be created. A dataset of labelled lion-faces over whole images for training an initial Convolutional Neural Network that implements the YOLO algorithm was created. A dataset for labelled whisker locations over cropped lion-faces to train a second YOLO network was made. Datasets comprising the images of cropped whisker regions separately for each individual lion were also created [118].

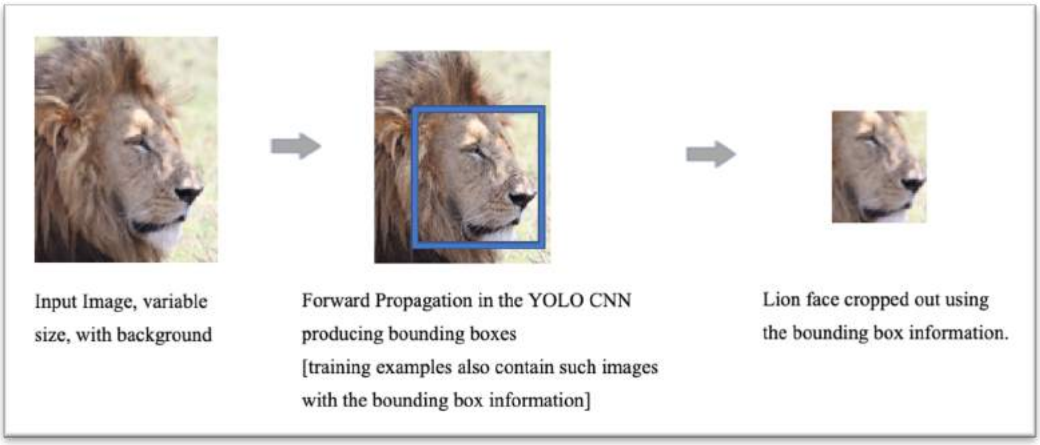


Figure 5.3: Using convolutional neural networks (YOLO) to crop the lion-face from the input image

These datasets for each individual's whisker images provide for training the inception convolutional network by the triplet loss function over precomputed triplets. To create all the datasets the images for each lion was manually downloaded along with the images from the Wildlife Institute of India and from the website that hosts the photographs taken in the Mara Predator Project that has lion photographs documented by individuals over the time period of 2008 to 2013 at the Maasai Mara National Reserve located in Kenya. [15]

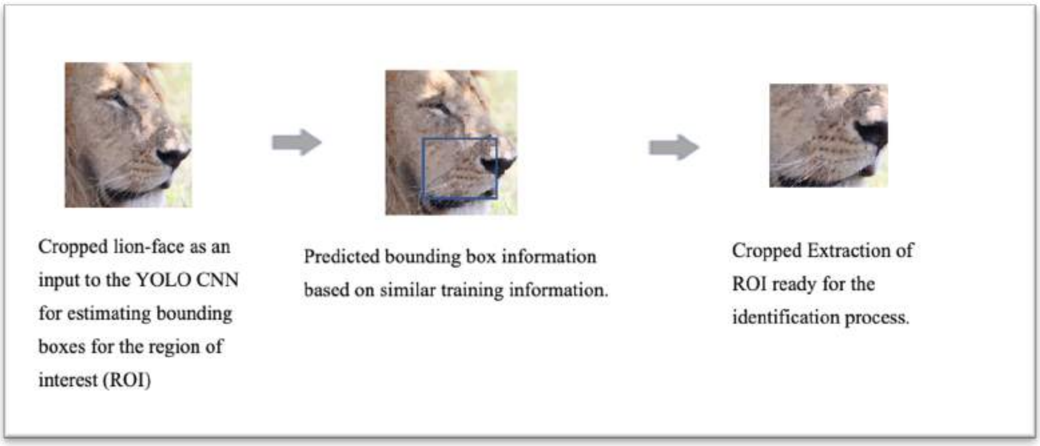


Figure 5.4: Using convolutional neural networks (YOLO) to crop the whisker region (ROI) from the lion-face input

The first step towards the identification process is to locate a Lion face within the Image provided to the system [85]. This is achieved by feeding the input image to a Convolutional Neural Network that implements the YOLO algorithm to get the bounding boxes for the location of the lion face [12].

The training set for this neural network comprises of a set of labelled Images with bounding boxes for three different classes of lion faces: the front face, the right face and the left face. This step also provides for the information of the side of the lion's muzzle from which the whisker patterns are extracted. Forward propagating through this trained network, we get the coordinates to the bounding boxes for the located lion-face in the image. As a final process of this step, we use the bounding box coordinates to crop out only the lion face from the given image (Figure 5.3). This ensures the reduction in background noise for our next step.

Next, to extract the region of interest, which is the whisker area of the lion-face [2], the cropped-out lion face image is fed into another YOLO convolutional neural network trained over cropped lion face images labelled at the whisker area. Forward propagating over a network trained in this fashion gives us the coordinates to the bounding box for our region of interest (ROI).

Using this bounding box information, the whisker pattern is separately cropped out (as described in Figure 5.4).

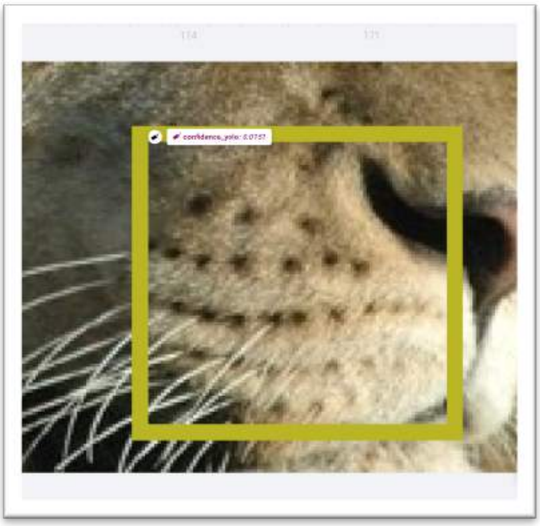


Figure 5.5: Thresholded Confidence of the ROI were fed to the inception convolution network for extracting individuality features

5.4.1 SEMI-AUTOMATED APPROACH FOR DATA NORMALIZATION

To provide a the Convolution network with a uniform representation of data across all the different Lion individuals, there is a requirement to normalize the data such that there are fixed keypoints across the whole dataset. The normalization of the data across the whole search space is an essential requirement, as this is the only way the uniqueness of an individual can be differentiated from other individuals across the dataset. Without normalization we will introduce noise into the system by training the system over features that exist outside our search space of uniqueness which exists within the limited Region Of Interest (ROI). Hence a manual approach is used to achieve this step. This renders the end-to-end process to be fashioned in a semi-automated style, since all the other steps within the system remains automated.

Described below is the detailed account of the manual keypoint marking process that creates the normalized dataset.

The data normalization done to create a normalized representation across the dataset generated is achieved by the **LI-No^2 normalization** that employs the use of perspective transformation techniques with fixed coordinate selection rules across all images [149].

Perspective transformation also known as 'keystone mapping' is a type of image transformation that uses the concepts of homography (uv mapping) by the creation of a transformation matrix as an intermediate step to compute the transformation in terms of perspectives. Therefore this intermediate mapping computes coordinates in a higher dimensional space to achieve perspective invariance [101], allowing the mapping of the original coordinates to any other perspective further [154].

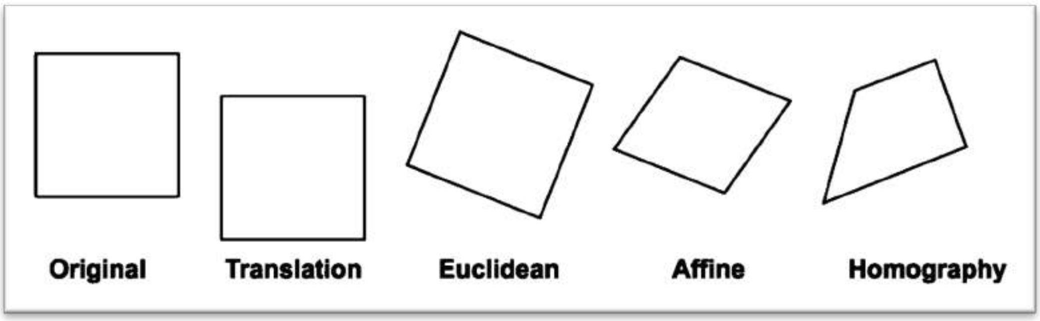


Figure 5.6: Perspective transform in contrast to other motion models

In comparison to the other mothion models, the transformation using to perspective mapping, preserves more information as it has 8 degrees of freedom for transformation, therefore preserving:

- Concurrency
- Collinearity
- Order of contact (intersection, tangency, inflection etc) and
- Cross ratio of collinear points

Fixing a perspective for the dataset, creates a type of normalization across the data for the various angles in which the original data may be captuted. Due to the properties of the perspective transform, that of preserving information while being mapped onto other perspectives, it is preferred that the image data be normalised to a fixed perspective. The normalised perspective is fixed by various governing rules as shown below.

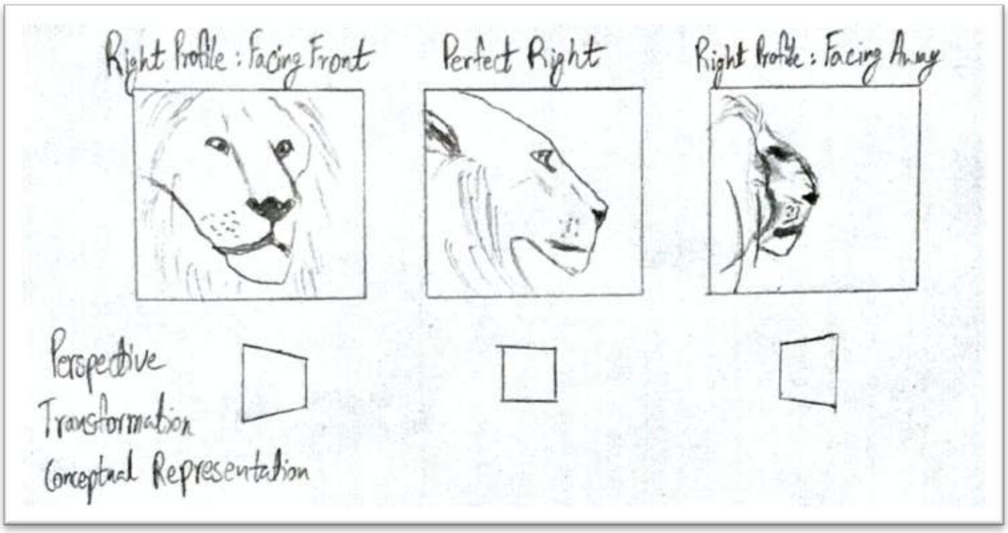


Figure 5.7: Representation of the transformation warps

Only the right profile images are chosen as mentioned before. Normalization coordinate rules are specified such as the coordinate click scheme needs to start at the top right corner.

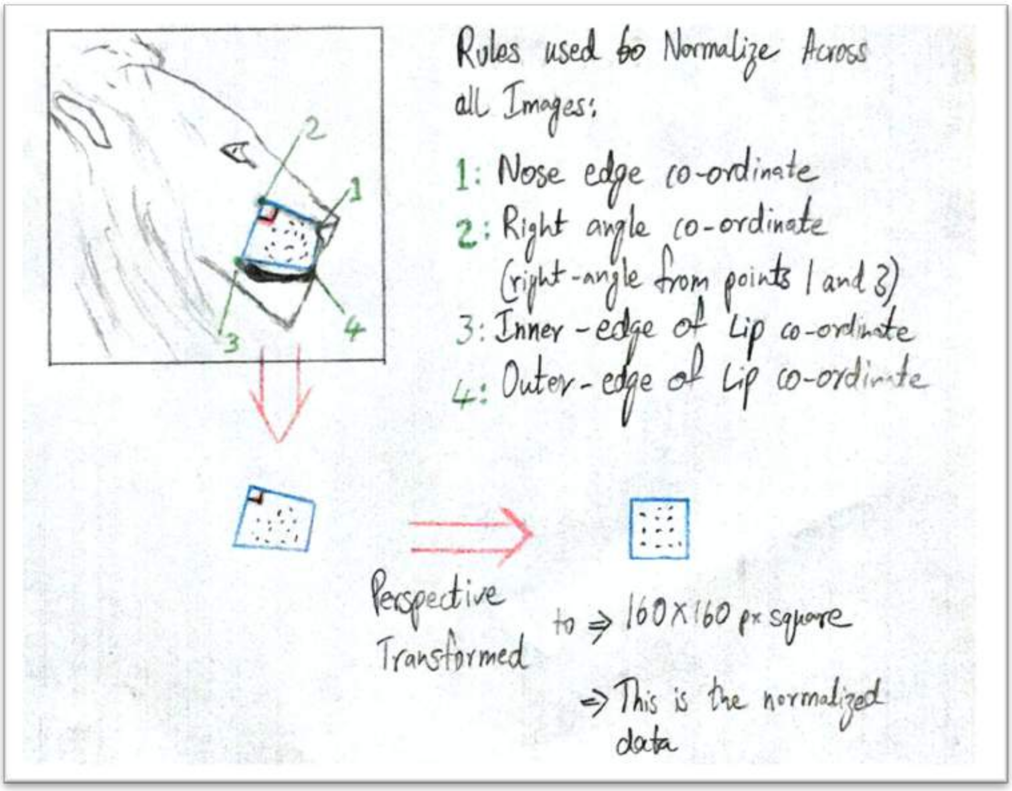


Figure 5.8: Normalization coordinate rules

The first coordinate represents the nose edge, followed by the second coordinate which should be at a right-angle between the first and the third coordinate. The third coordinate represents the inner edge of the lion's lip and the fourth coordinate represents the outer edge of the lion's lip.

Algorithm 1: Lion Identification Non-Invasive Normalization (LI-No^2):

Input: Image(I) with coordinates PerspCoord(list) for a preconditioned perspective transformation
while co-ordinate mapping matches pre-set rules **do**
  TrMx=generate_transformation_matrix(PerspCoord, size(160*160))
  Rectified_Image=Perspective_transform_warp(TrMx,I,size(160*160))
end **while**
Return Rectified_Image

Algorithm 1 is the novel LI-No^2 algorithm introduced in this research work, as a method to efficiently normalize the set of images that includes the region of interest (ROI) to be perspective agnostic by having a fixed perspective to which all the given images are transformed to. This is possible by setting coordinate rules that apply over the 3-dimensional spacial representation of the real world object, which is the landmarks over the lion's face in our case. The coordinate rules are described above and also shown in Figure 5.8. The algorithm describes the process used to implement perspective transformation as the normalization standard over the image dataset. The input to the algorithm is the image to be normalized provided along with the coordinates that are obtained by the manual marking process. The predefined rules of mapping are verified, such as, the order in which the coordinates are provided, etc. An intermediate transformation matrix *TrMx* is generated that contains the mapping information for the list of coordinates passed *PerspCoord* mapped to the preset perspective of normalization with a fixed size of 160X160 pixels. This transformation information is then applied over the input image for every pixel value and its coordinate information to generate the transformed image with the normalised perspective of a fixed size of 160X160.

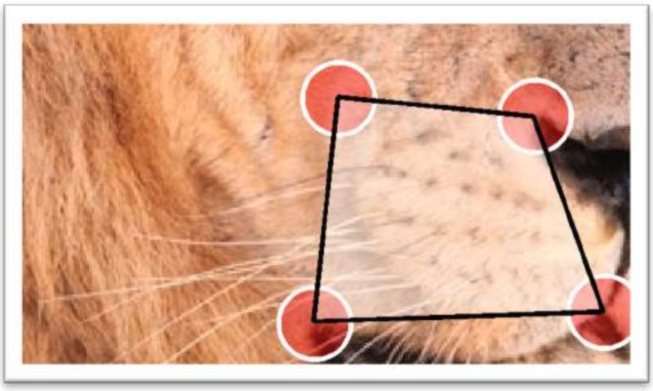


Figure 5.9: Selection tool for the perspective crop

*Figure 5.10: Perspective transformed image at the 160*160 normalization setting*

As seen in the images above (Figure 5.9 and Figure 5.10), the manual marking step within the semi-automated process generates the coordinate locations with respect to the real world representation of the lion face in 3 spacial dimensions, over the 2 dimensional image. These coordinates follow the rules as defined above. The marking of these coordinates also follow the order specified within the rules, which is in a clockwise fashion. This is also a convention used so that there is a direct one to one mapping within the tool that is made, without any need to have an additional rectification step, thus reducing ambiguity as well.

5.4.2 SPECIFIC-DIMENSION VECTOR EMBEDDINGS GENERATED BY THE IDENTITY MODEL USING THE TRIPLET LOSSFUNCTION OVER A SIAMESE NETWORK ARCHITECTURE

After the creation of the normalized dataset across multiple individuals, the required next step is to be able to create a model that is able to represent individuality, which is a model that is able to represent each individual differently. To achieve this, we need to train a network in such a way that it assigns individuality to unique patterns that exists in the extracted region of interest, not accounting for the various image artifacts such as difference in colours due to various lighting intensities and other inconsistencies such as angle variation of the captured pattern.

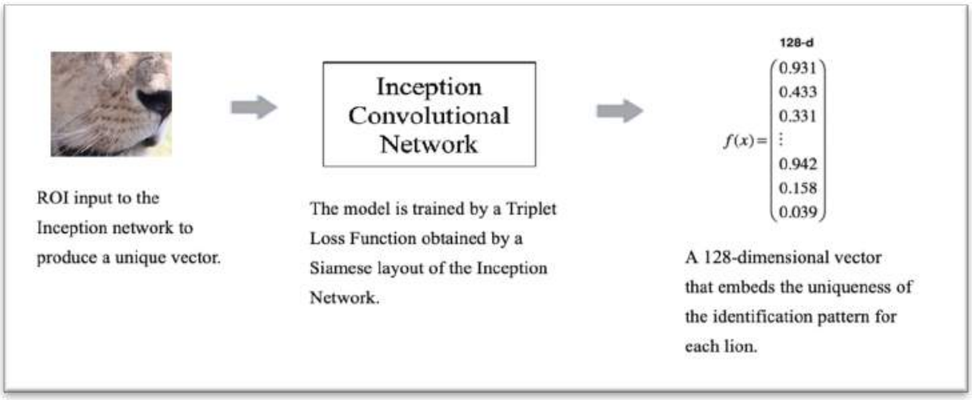


Figure 5.11: Generating 128-Dimentional vectors for each lion with an inception network trained using a triplet loss function

This is achieved by using a Triplet Loss Function and backpropagating (training) this loss function over an inception network, that uses multiple convolution filters for each layer. The inception network outputs a fixed-dimensional vector that will be able to encode the uniqueness in such whisker patterns once trained over multiple triplets of such vectors (as shown in Figure 5.11).

The triplets are created by grouping the vector data of three different images. Out of these three images, two are different images of the same lion and the third is an image of a totally different lion. The triplet loss function is formulated by these images namely, the anchor image, the positive match and the negative match.

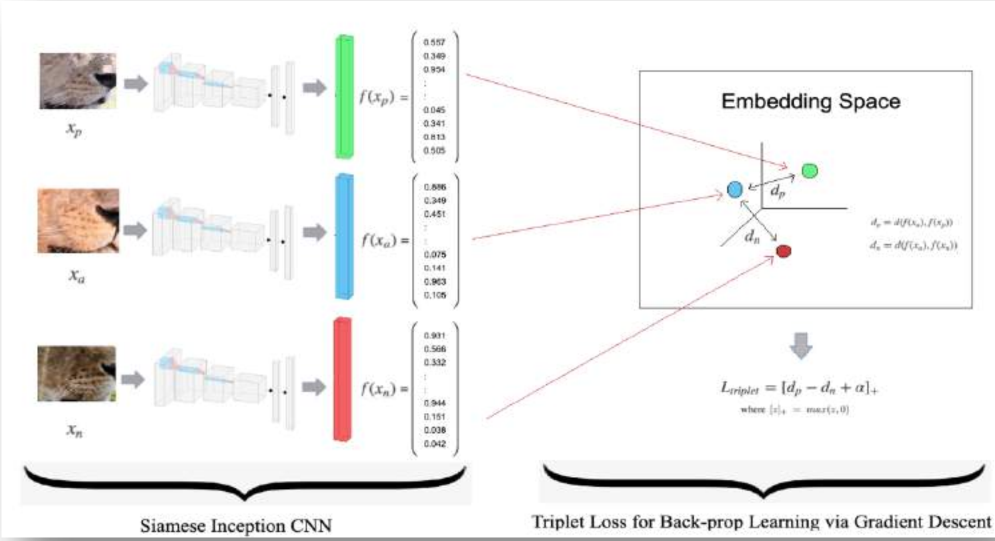


Figure 5.12: An intuitive representation of the Siamese Network for the Triplet loss Function along with the Embedding Space

To form these triplets for each anchor image, the vectors of the positive match as well as the negative match are precomputed each by a forward pass of the respective images through the inception net, so that the loss function can be computed over the current pass of the anchor image. Such a network of using the outputs of multiple forward passes through a same neural network, is known as a Siamese network [89]. Minimizing the difference between the positive pair of the anchor and the positive match and maximizing the difference between the negative pair of the anchor and the negative match is the goal of the triplet loss function [11][14]. An intuitive representation of these processes is shown in Figure 5.12. Backpropagating this loss function through the inception network, creates a model for the various unique patterns and their nature of similarity and dissimilarity. Described below is the triplet loss function (Equation (*5.2)*):

$$J = \sum \left\| f\left(A_{(i)}\right) - f\left(P_{(i)}\right) \right\|_2^2 - \left\| f\left(A_{(i)}\right) - f\left(N_{(i)}\right) \right\|_2^2 + \alpha \qquad (5.2)$$

In the triplet loss function, A is the anchor image, P is the image with the positive match and N is the image with the negative match. The function f, denotes a neural network forward pass that results in a 128-dimensional vector. The L2 norm (Euclidean distance) is taken for each pair with the anchor by computing the sum of squared differences. Finally, α denotes the margin that increases the distance between the negative match pairs and decreases the distance between the positive match pairs. This loss function operates on the triplets for all anchors used for training, hence the summation over the whole function.

Hence while forward propagating for a test image of the cropped region of interest, the network thus outputs a 128-dimensional vector which is used to search for a match within the stored set of vectors of known lions. The matching process is done by finding the Euclidean distances for each pairwise comparison with the vectors of every lion within the database. The lesser the distance the more the similarity, therefore we use a threshold value to find a match.

Algorithm 2: Lion Mystacial Features by Triads (LMFT): Extraction training and Clustering

---

Input: Datasets (LFlab, CLWlab, CLwin), Neural Nets (Y1,Y2,Incep), Triplets Template

while yolo-bounding-box-prediction-loss decreases **do**

   update Y1 with Y1_BackProp(LFlab)

end **while**

while yolo-bounding-box-prediction-loss decreases **do**

   update Y2 with Y2_BackProp(CLWlab)

end **while**

Precompute initial vectors by Incep_ForwardProp(CLwin) => T

Triplet Template (T) => Triads

---

while triplet-loss decreases **do**
    Incep_BackProp(Triads)
end **while**
Return Incep(Triads)

Algorithm 2 describes the process of training the separate networks with respective datasets. The dataset that contains samples of labelled lion faces in whole images is denoted by the name 'LFlab'. 'CLWlab' is the dataset that contains samples of cropped lion-faces with the whisker region labelled. 'CLwin' is the dataset that contains samples of cropped whisker regions, this dataset is subdivided into sets for every individual separately. 'Y1' represents the YOLO neural net that estimates lion face label boxes. 'Y2' represents the YOLO neural net that estimates the whisker region label boxes. Finally, the 'Incep' neural net is the inception network that trains with the triplet loss function which is denoted as 'triplet-loss'. The function with the '_BackProp' postfix denotes the back-propagation training pass of the respective neural network. The function with the '_ForwardProp' postfix denotes the forward-propagation pass of the respective neural network. The 'yolo-bounding-box-prediction-loss' is the loss function that governs the training of the respective YOLO networks to estimate the bounding boxes for the region it is trained for. 'Triplet Template' is a function that creates the triplets based on the lion identities. The trained inception network is the final outcome.

Algorithm 3 describes the various steps taken for the process of lion identification, when the system is provided with a new input image. The input image is denoted by 'I'. The various lion identities are stored as vectors in a database represented as 'LiDb'. The neural net 'Y1' represents the first YOLO net that estimates lion face locations. 'Y2' is the second YOLO net that estimates the whisker locations. Lastly, 'Incep' is the inception network that outputs a 128-dimensional vector that serves as a representation of the lion's identity. The ‘crop_img’ function crops out the predicted region from the original image by taking both the output of the neural network as well as the original image as it’s inputs. ‘Square_diff’ is a function that computes the sum of the squared differences of all values between two vectors of the same dimension. This function computes the difference between the current vector along with all the vectors stored within the database. The ‘Threshold_diff’ value is the threshold value that denotes that an identity is a match if the difference value from the previous step is below this threshold value. Hence, the final value returned is the identity of the identified lion.

Algorithm 3: Lion Mystacial Identity (LMId): Forward pass prediction

---

Input: Image containing Lion Face (I), Datasets of known lion vectors(LiDb), Neural Nets (Y1,Y2,Incep)

Bounding box 1 = Y1(I)

LF = crop_img(bounding box 1, I)

ROI_box=Y2(LF)

ROI=crop_img(ROI_box,LF)

Identity vector=Incep(ROI)

Identity=Square_diff ((**For** all Xi **in** LiDb), Identity vector) < Threshold_diff

Return Identity

5.4.1 FINDING THE BEST TRIPLETS

The training process of the Siamese inception neural network relies on the process of providing the network with training samples in the form of triplets. The learning process within the model takes place when the weights in the embedding space are adjusted to fit the loss generated by the triplet loss function. The triplet loss function is designed to optimise the distances between the embedded representations for the images in a given triplet within the embedding space. This is set, by maximising the distance between the embedded representations of an anchor image with that of the negative match image and by minimizing the distance between the embedded representations of an anchor image with that of the positive match image. Therefore in each of the optimization routines, for the distance maximization as well as for the distance minimization, the weights must be shared. This is why this approach is known as the Siamese neural network. This is done so that the learning based on the loss function may use the weights fixated on the anchor image and then optimise over the representations of the positive as well as the negative match respectively. The distance optimisation is governed by a set margin parameter *'α'*, for both distance minimisation as well as maximisation over the positive match and the negative match to the anchor image respectively.

This process is compute intensive and if it is designed to compute over the whole set of all possible triplets, the learning will be slow and the model might even train over noise, or even learn nothing based on the nature of the triplets provided. Hence, a triplet selection strategy is deployed that identifies triplets that are harder to learn, that is, those triplets, whose representations are further from the truth. The process is known as semi-hard triplet mining [11] [72]. Therefore, the triplets are segregated into easy, hard and semi-hard categories based on how far the embedded representations deviate from the specified ground truth.

If the embedding representations of each image in a triplet set, follow the definitions set by the triplet loss function, the Euclidean distances between the three embeddings, will be within the acceptable margin, reflecting its correctness compared to the ground truth. Thus in this case, the Euclidean distance between the embedded representations of the positive-anchor image pair will be lesser than that of the negative-anchor image pair by a margin distance set by the parameter *'α'*. Therefore the anchor-positive image pair is represented closer to each other and the anchor-negative image pair is represented farther from each other *(5.3)*. Such triplets come into the easy category as there is no correction needed to be made therefore no learning is to be performed over the model to be trained.

$$D\left(f\left(A_{(i)}\right) - f\left(P_{(i)}\right)\right) + \alpha < D\left(f\left(A_{(i)}\right) - f\left(N_{(i)}\right)\right) \qquad (5.3)$$

If the embedding representations of each image in a triplet set, do not follow the definitions set by the triplet loss function, the Euclidean distances between the three embeddings, will not follow any kind of margin, reflecting its deviance from the ground truth. Thus in this case, the Euclidean distance between the embedded representations of the positive-anchor image pair may not be lesser than that of the negative-anchor image pair *(5.4)*. Leading to the possibility that the anchor-negative image pair may be closer to each other by the Euclidean distances of their embedded representations, as opposed to that fact that it should be represented farther from each other. There also exists the possibility that the anchor-positive image pair may be farther to each other as opposed to that fact that it should be represented closer to each other. Such triplets come into the hard category as there is a large correction needed to be made even with respect to the margin parameter *'α'*, therefore steep learning steps are to be performed over the model to be trained.

$$D\left(f\left(A_{(i)}\right) - f\left(N_{(i)}\right)\right) < D\left(f\left(A_{(i)}\right) - f\left(P_{(i)}\right)\right) \qquad (5.4)$$

If the embedding representations of each image in a triplet set, follow the definitions set by the triplet loss function, but are not within the margins set by the parameter *'α'*, the Euclidean distances between the three embeddings, will not be within the acceptable margin, reflecting weak correctness compared to the ground truth. Thus in this case, the Euclidean distance between the embedded representations of the positive-anchor image pair will be lesser than that of the negative-anchor image pair but not by a margin distance set by the parameter *'α' (5.5)*. Therefore the anchor-positive image pair is represented closer to each other and the anchor-negative image pair is represented farther from each other but not by a desired margin. Such triplets come into the semi-hard category as there is a slight correction needed to be made with respect to the margin parameter *'α'*, therefore lighter learning steps are to be performed over the model to be trained.

$$D\big(f\big(A_{(i)}\big) - f(P_{(i)})\big) < D\Big(f\big(A_{(i)}\big) - f\big(N_{(i)}\big)\Big) \\ < D\big(f\big(A_{(i)}\big) - f(P_{(i)})\big) + \alpha \tag{5.5}$$

Therefore, to provide such triplets that allow the model to learn as well as to provide such triplets that are not so difficult and compute intensive resulting in the learning process taking more time, we provide the model with the semi-hard triplets.

5.4.2 THE SOFTMAX FUNCTION

For the efficient and reliable training of the deep neural networks, the SoftMax function is used. This SoftMax function is applied in both the approaches used in this research work to arrive at a prediction. The SoftMax function gives a prediction score values that range in between 0 and 1 for all the available classes. Therefore the class with the maximum value among all the predictions, is considered as the predicted class. This is very useful as it provides the network with gradient differences in the errors computed by the loss functions. This proves efficient in contrast to the argmax function which converts a max value output to 1 and the predictions to the remaining classes as 0. Therefore, if we have an output pertaining to a particular class as Z_k the softmax for the same with respect to all other outputs can be calculated as shown below *(5.6)*.

$$S(Z_k) = \frac{e^{Z_k}}{\sum_{i=1}^{i=H} e^{Z_i}} \tag{5.6}$$

And, the sum of all the SoftMax outputs is 1 *(5.7)*.

$$\sum_{i=1}^{i=H} S(Z_i) = 1 \tag{5.7}$$

Thus the loss computed through this function, known as the SoftMax loss encompasses specific losses pertaining to each class with respect to every other class, and therefore proves to be very efficient while training networks using back propagation. The training of the embedding space benefits from the advantages of using the SoftMax function especially in the approach that relies on only using the SoftMax loss function compared to using the Siamese architecture with the triplet loss function.

5.5 THE SUPPORT VECTOR MACHINE

After designing the model for the feature extractor that generates unique embedded representation for each Lion individual by using a well-trained embedding space, the next and final step is to create a classification model that is able to use the embedded representations of each lion individual and map them correctly to the name or identity value of the lion individual that it actually represents. The embeddings generated for each lion by the trained feature extractor serve as template representations for the lion individual in consideration. Therefore, to create a classifier model a Support Vector Machine (SVM) algorithm is used as it serves as a robust classifier widely used in the industry. Also, each lion is now represented in a numerical fashion with a fixed length structure, allowing for a more structured classification approach. A generalised representation of a support vector classifier for two classes is shown below *(*5.8*)*, showing how the support vector classifier targets to minimise over terms $w, b\ and\ \zeta$, by the formula as shown, representing the loss function itself. The parameters w and b are the margin and the bias parameters within any prediction given by $y_i(w^T\phi(x_i) + b)$. Therefore the goal is to make the correct predictions for most of the samples.

$$\begin{gathered} \min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i \\ \text{subject to} \quad y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i, \\ \zeta_i \geq 0, i = 1, \ldots, n \end{gathered} \tag{5.8}$$

The parameter ζ_i is the allowed distance from the support vector hyperplane to a misclassified sample, since all samples may not be correctly classified by a hyperplane. The function ϕ is the identity function. As an intuitive understanding, the goal of the suppor vector classifier is to maximise the margin between the classes by minimising $||w||^2 = w^T w$, by inducing a penalty for any misclassification. Therefore the distance of ζ_i is allowed from the perfect prediction of 1 by the prediction function $y_i(w^T \phi(x_i) + b)$. This penalty is controlled by the term C thst specifies the strength of the penalty therefore having an effect of regularization.
This method of using the Support Vector Machines (SVMs) for the classification routine is employed for both the approaches examined in this research and the results of which are described in the forthcoming chapter.

5.6 SUMMARY

This chapter has covered in detail the various building blocks that is used to create the Lion Re-Identification System within this doctoral research work. Architectures such as the YOLO architecture, the Siamese Network Architecture, as well as the concepts behind the triplet loss function, with various triplet selection strategies have been described. An introduction to the LI-No^2 normalization algorithm that uses various techniques to normalise the perspective across all the images within the dataset by a semi-automated technique is also covered here in the chapter. A prime focus of this chapter is to give a sense of how all the various techniques come together to thereby create an embedded representation for each Lion individual thus providing a structured template of fixed dimensions as a feature vector that encapsulates the identity of a lion individual. Furthering this, it is also described of how a support vector machine is used as a classifier to the various identity labels thus generating a model that is able to map the embedded identity representation to the actual identity name, therefore providing an identity prediction interface to the outside world. The upcoming chapters thereby provide the various discussions on the results thus generated using the techniques described within this chapter.

CHAPTER 6 EVALUATION OF THE DEEP LEARNING TECHNIQUES

6.1 INTRODUCTION

This chapter documents the records of evaluation for different models created for the identity estimation process using the deep learning based computer vision techniques performed over various configurations iteratively. A detailed description of the creation of the datasets using various augmentation techniques is also covered in this chapter. The performance results of the different models evaluated are thus documented. To reduce the complexity of the system and to avoid the possibility of correlation between the whisker patterns of both the sides of the lion's muzzle, these results are achieved by training the models over the data of the whisker patterns of only the right side of every lion providing for a singular identification signature.

6.2 THE APPROACH:

For the purpose of creating a model to perform the inference of lion identities, it is necessary to create a dataset of image examples for multiple lions. These examples are used to perform training over a suitable deep learning architecture that will then serve as the model used for inference. As mentioned in the further sections of this chapter, the initial problem of data scarcity is handled by the process of data augmentation, achieving a 900% increase in the size of the dataset.

To generate encoded representations [157] of the lion individuals from their normalized photographic representations, an inception convolution neural network is used. During the training routine, the output layer of this neural network is trained by a SoftMax loss process using backpropagation, hence the network gets trained as a classifier for the multiclass classification over the different lion identities provided. As the output requirement is an encoded representation of the identities, we need to extract the embedding space representation of a model trained over the lion identities. Hence, it is desired that the embeddings of the multi class classifier trained over the lion identities be used [73] [75] [76] [77]. To have an embedding of a desired dimension size, the construction of the inception network is done in such a way that the number of nodes of the second-last fully connected layer is equal to the desired dimension size. This is the layer which is located just before the output layer. The setting of the number of nodes as the number of dimensions of the embedding space, within this layer needs to be done prior to the training process. The values of the weights within the nodes of this layer taken as a vector, serves as the encoded representation [97]. These encoded representations in the dimensions specified initially, will be the encoded representation of the lion identities once the network is trained as a classifier of lion identities (Figure 6.1).

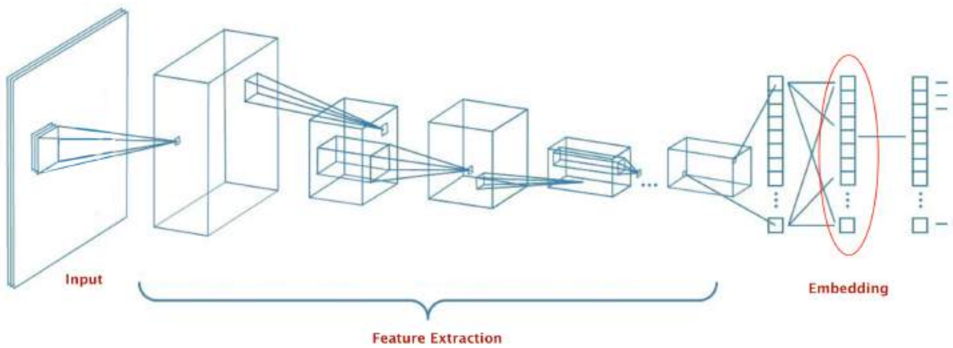


Figure 6.1: The conceptual representation of an embedding within a deep classifier network

Data for new lions can easily be introduced into the system without the retraining of the deep learning architecture. This is possible due to the strategy of using the embedding representations as a feature vector, representing the lion individuals. Thus the convolutional network initially trained as a classifier is repurposed as a feature extractor.

Furthering the creation of the system for the identity estimation of the lion individuals, the support vector machine (SVM) multi-class classification algorithm is used. Therefore, once the feature extractor is created, input images forward-passed through it, generates unique representations as feature vectors with specific dimensions. And since the feature vector is trained for classifying lion individuals, this embedded representation, can be considered as a representation of individuality. Thus an image representation of the same lion will have its vibrissae spot pattern represented in a close proximity to other image representations of the same lion within this embedding space. The inverse is also true, such that the embedded representations of different lions are far from each other. This embedding information is exactly what gets represented by the values generated by a forward pass within the encoding vector. Hence, this encoded information of individuality is used to train a multi-class classifier, so that a model is thus created that can assign identity labels to these encodings received from the feature extractor, once an image is passed. For the current implementation, a support vector machine (SVM) is used.

This is a detailed overview to the approach used for creating the re-identification system for lion individuals. Different models are created by incorporating various architectural changes and dimensional changes, and are thus evaluated. One such architectural change evaluated in this work is the incorporation of a Siamese network architecture for the purpose of generating a triplet loss function. Considering the problem at hand, there are multiple issues that can be addressed such as, the multiclass nature, along with the fact that new classes need to be added at a later stage during the usage of the system, and also, the added issue of the scarcity of data. The usage of the triplet loss function to backpropagate over the network for the training of the feature extractor, addresses these issues, this is due to the fact that, the embedding space in effect represents individuality as separate clusters for each individual. Hence the creation of triplets from the data, that of an anchor, a positive match and a negative match, generates a larger set of data. This, along with the loss equations, creates clusters withing the embedding space trained over the lion identities. Therefore due to the capability of robust creation of clusters by the triplet loss function trained over a Siamese network, the results of this architecture is documented, with a final comparison between both the approaches with and without triplet loss function.

6.3 THE DATASET

An initial normalized dataset using the perspective transformation technique (LI-No^2) is created with the techniques described in section 5.4.1. Batchwise operations over this initial state of the acquired data is used to generate the final form of the datasets to be used for training over the features that encode individuality [96] [170]. These batchwise augmentation operations include various blur augmentations such as Gaussian blur and Median blur and various noise augmentations, that of Gaussian noise in two different configurations. For the blur augmentation operations, the Gaussian blur transformation is performed with a configuration of sigma max of '10' and sigma min of '2' and the Median blur transformation is performed with a configuration of the kernel size set to '5'. For the noise augmentation operations, the first Gaussian noise transformation is performed with a configuration taking the standard deviation of the noise to be of a value '5' and the mean of the noise to be of the value '25', next, the second Gaussian noise transformation is performed with a configuration taking the standard deviation of the noise to be of a value '25' and the mean of the noise to be of the value '5'. The flow diagram for these augmentation operations along with the various batches, that of the original, intermediate as well as the final dataset, along with the various transformation configurations are shown in Figure 6.2.

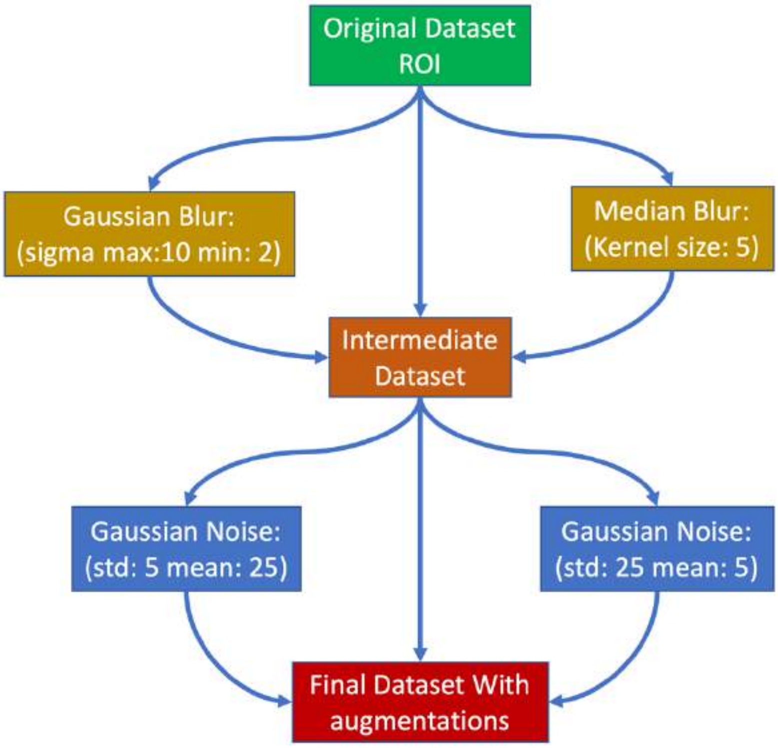


Figure 6.2: Batchwise augmentation flow diagram showing the transformation configurations

As shown in the Figure 6.2, there are two stages of augmentation, the first stage that creates the intermediate dataset and the second stage that creates the final dataset (Figure 6.3).

Figure 6.3: A few samples taken from the final dataset over which the LI-No² transformation and the various augmentation routines are performed

At each stage, the augmentation process creates three copies of the dataset it starts with, and applies various transformations to two copies and leaves the third copy untouched. Hence, at each augmentation stage, the new dataset that gets generated after each transformation step and contains the original data untouched by any transformation, and hence the size of this new dataset gets tripled in size compared to the dataset it starts with. At the second level of transformation, we also get data with various combinations of transformations with the transformed data of the first level.

6.3.1 DATASET FOR THE FEATURE EXTRACTOR

For the purpose of training and validation of the feature extractor, an initial dataset consisting of the normalised images using the perspective transformation technique (LI-No^2), consists of 46 lion individuals within the obtained dataset of 51 lion individuals. A minimum of 3 image examples represent each lion individual. For the purposes of testing the whole system the images of 5 individuals were not provided for the training of the feature extractor. Therefore in total, the size of the initial dataset amounts to 161 images for the 46 individuals. Augmenting this batch consisting of the initial dataset with the blur transformations, creates an intermediate dataset batch with a size amounting to 483 images, which is a 300% increase in the size of the original dataset. Applying the next stage of augmentation, to the intermediate dataset batch with the noise transformations, creates the final dataset batch with a size amounting to 1449 images, which is a 900% increase in the size of the original dataset. Therefore, through the augmentation techniques, a 900% increase in dataset size is achieved (Table 6.1).

Table 6.1: Quantities within the dataset created

Data Description	Quantity
Unique lion individuals within the database	51
Minimum training samples per individual lion	3
Total images of all lions	186

Number of individuals within the dataset for training feature extractor	46
Total images for feature extractor training dataset	161
Dataset augmented with blur transformations	483 (300%)
Final dataset with both blur and noise transformations	1449 (900%)

This dataset is created for the purpose of training the feature extractor. A detailed description of the creation of this dataset is shown below (Table 6.2).

Table 6.2: Information of the dataset used to train the feature extractor

Lion Code	Lion Name	Transformed ROI	Blur Augmentations	Noise Augmentations
AF1	Ajani	4	12	36
AF2	Amber	3	9	27
AF3	Arria	3	9	27
AF4	Charm	3	9	27
AF5	Doto	3	9	27
AF6	Engiyaa	3	9	27
AF7	Enkume	3	9	27
AF8	Esiriwua	3	9	27
AF9	Hasani	5	15	45
AF10	Kaka	3	9	27
AF11	Kinna	3	9	27
AF12	Lemayian	3	9	27
AF14	Mama-kali	4	12	36
AF15	Maskio	3	9	27
AF16	Matajo	4	12	36
AF17	Mickey	4	12	36
AF20	Naengop	3	9	27
AF22	Naisiae	3	9	27
AF23	Nakato	3	9	27
AF24	Namunyak	3	9	27
AF25	Nariku-inkgera	4	12	36
AF26	Ngare	3	9	27
AF27	Nguro	3	9	27
AF28	Nkasiogi	3	9	27
AF29	Olbarnoti	3	9	27
AF30	Olchore	4	12	36
AF31	Opi	3	9	27
AF32	Romeo	4	12	36
AF33	Saba	3	9	27
AF35	Safi	4	12	36
AF36	Saitoti	5	15	45
AF37	Samir	3	9	27
AF38	Sangiki	3	9	27
AF39	Saruni	5	15	45
AF40	Selenkay	4	12	36
AF41	Senteu	3	9	27
AF42	Sero	3	9	27
AF43	Shambe	4	12	36
AF44	Siena	4	12	36

AF45	Simaloi	3	9	27
AF46	Spot	4	12	36
AF47	Spring	3	9	27
AF48	Summer	5	15	45
AF49	Supu	5	15	45
AF50	Tikki	4	12	36
AF51	White-eye	3	9	27
	Total	161	483	1449

6.3.2 DATASET FOR THE CLASSIFIER AS A DATABANK OF KNOWN INDIVIDUALS

The feature extractor once trained, performs the task of a template generator, by encoding the embedding space representations onto vectors having the exact dimensions as that of the embedding space. Thus each vector represents distinct locations within the embedding space. These encoded vectors embed representations of individuality, such that the Euclidean distances between the vector representations of the same individual will be at a minimum and the distances will be at a maximum between the vectors of different individuals. To create a model that learns such differences and similarities for a set of known lions for the purpose of their reidentification, a Support Vector Machine (SVM) Classifier is trained on the template vector representations for a dataset containing the normalized images of a known set of lions. To train and validate the classifier, a dataset as a set of known individuals comprising of sample images of 18 individuals are taken from the initial dataset of the normalised images along with the normalized images of the 5 individuals that were separated out initially. Both these sets are normalized using the perspective transformation technique (LI-No^2), and comprise a minimum of 3 image examples for each lion individual. A set of 2 images each, for 10 individuals are withheld from this dataset to create the test set (section 6.3.3). Therefore the dataset of known individuals, contains the samples of 23 lion individuals with a total of 82 images. The augmentation process which is conducted over the training dataset for the feature extractor is repeated for this data set as well. Hence, augmenting the initial dataset with the blur transformations, creates an intermediate batch dataset with a size amounting to

246 images, which is a 300% increase in the size of the original dataset. Applying the next stage of augmentation, to the intermediate batch dataset creates the final dataset batch with a size amounting to 738 images, which is a 900% increase in the size of the original dataset. Thus, as seen previously, through the augmentation techniques, a 900% increase in dataset size is achieved (Table 6.3).

Table 6.3: Information of the dataset used to train the classifier as a databank (Individuals marked in Red are new individuals to the system)

Lion Code	Lion Name	Transformed ROI	Blur Augmentations	Noise Augmentations
AF9	Hasani	3	9	27
AF13	Lolparpit	3	9	27
AF14	Mama-kali	4	12	36
AF16	Matajo	4	12	36
AF17	Mickey	4	12	36
AF18	Mooza	3	9	27
AF19	Moswen	3	9	27
AF21	Naibor	3	9	27
AF25	Nariku-inkgera	4	12	36
AF26	Ngare	4	12	36
AF30	Olchore	4	12	36
AF32	Romeo	4	12	36
AF34	Sadala	3	9	27
AF35	Safi	4	12	36
AF36	Saitoti	3	9	27
AF39	Saruni	3	9	27
AF40	Selenkay	4	12	36
AF43	Shambe	4	12	36
AF44	Siena	4	12	36
AF46	Spot	4	12	36
AF48	Summer	3	9	27
AF49	Supu	4	12	36
AF50	Tikki	3	9	27
	Total	82	246	738

The Table 6.3 gives a brief portrayal of the dataset used to train the classifier to be used as a databank of known lions for the final re-Identification process. Hence, to test the capability of the system to generate unique identification features for new individuals, the normalised images of 5 new lion individuals are provided within the dataset created, over which the feature extractor has not been trained on. To distinguish these individuals from the rest, their Lion codes are marked in red colour within Table 6.3.

6.3.3 THE TEST DATASET

Maintaining a minimum of 3 normalized image samples for each lion individual, there exists individuals that have more than 3 samples. Strategically, the 5 lion individuals previously separated also have more than 3 samples. Hence a test set of 10 individuals, is created with a set of 2 sample images for each lion individual. These sample images are not provided to any prior training set. Out of the sample normalized images of these 10 individuals, the feature extractor is not trained on a single sample of the 5 lion individuals that are previously separated. Thus this test set also tests the viability of using the pretrained feature extractor as a template generator for known lions. Table 6.4 gives a portrayal of the test dataset, with the lion codes in red colour depicting the data of the lion individuals not used to train the feature extractor initially.

Table 6.4: Information of the Test dataset

Lion Code	Lion Name	No of Test ROI
AF9	Hasani	2
AF13	Lolparpit	2
AF18	Mooza	2
AF19	Moswen	2
AF21	Naibor	2
AF34	Sadala	2
AF36	Saitoti	2
AF39	Saruni	2
AF48	Summer	2
AF50	Tikki	2

6.4 THE RESULTS

In a broad overview, this work covers two approaches of training the feature extractor. The first approach covers the method of creating the feature extraction model by directly repurposing a trained deep inception network classifier, using the SoftMax loss function as the classifier algorithm. The second approach covers the method of modifying the deep inception network architecture to incorporate the form of a Siamese architecture, enabling the training of the feature extractor with a triplet loss function. For each of these approaches multiple models with different configurations [132] are created and evaluated. Thus creating two sets of models with the various configurations specified.

In the following sections, the lion identity estimation reports comprising of three parts have been generated for each model. These three sections include the confusion matrix of identity classification, a graph showing the confidence levels for each prediction and finally a classification report documenting the precision, recall, the f1-score and the accuracy. The details within these reports give an insight into the performance of any specific model.

To generate the classification report for a model, each classification output is analysed and counted within buckets of 4 different types. The types of classifications include, True Negative (*tn*) classifications, False Positive (*fp*) classifications, False Negative (*fn*) classifications and finally True Positive (*tp*) classifications. The number for each type of classification is counted to then calculate the values of precision, recall, f1-score and accuracy. The equations to each of the calculations of these values are shown below.

$$Precision = \frac{tp}{tp + fp} \tag{6.1}$$

$$Recall = \frac{tp}{tp + fn} \tag{6.2}$$

$$F_1 = \frac{precision \cdot recall}{precision + recall} \tag{6.3}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{6.4}$$

The following two sections comprise the documentation of the performances for the models of the two approaches that are previously mentioned. Within each section of the respective approach, the reports to the various model configurations have been documented.

6.5 SOFTMAX LOSS AND THE DEEP INCEPTION FEATURE EXTRACTOR:

This section documents the reports of the models with the different configurations that were tested for the initial approach using the feature extractor that is a repurposed deep classifier trained over lion identities using the SoftMax loss function. This approach does not have any major architectural modifications except the repurposing of the deep inception network classifier to extract the feature vector. The various subsections within this approach documents the reports to the various embedding size configurations. For each embedding size configuration, the model state for three levels to the number of epochs are considered and reports for each are documented respectively.

6.5.1 CONFUSION MATRICES FOR THE EMBEDDINGS TRAINED BY THE SOFTMAX LOSS FUNCTION:

A summarised representation of the confusion matrices of the predictions produced by various feature extractor model configurations with embeddings trained by the SoftMax loss function is shown below. The values denote the identities estimated by each model over 10 individuals of Lions on two test images for each individual. The 5 individuals that are not included in the training set of the feature extractor (AF13, AF18, AF19, AF21 and AF34) are included among the 10 individuals selected for the purpose of testing.

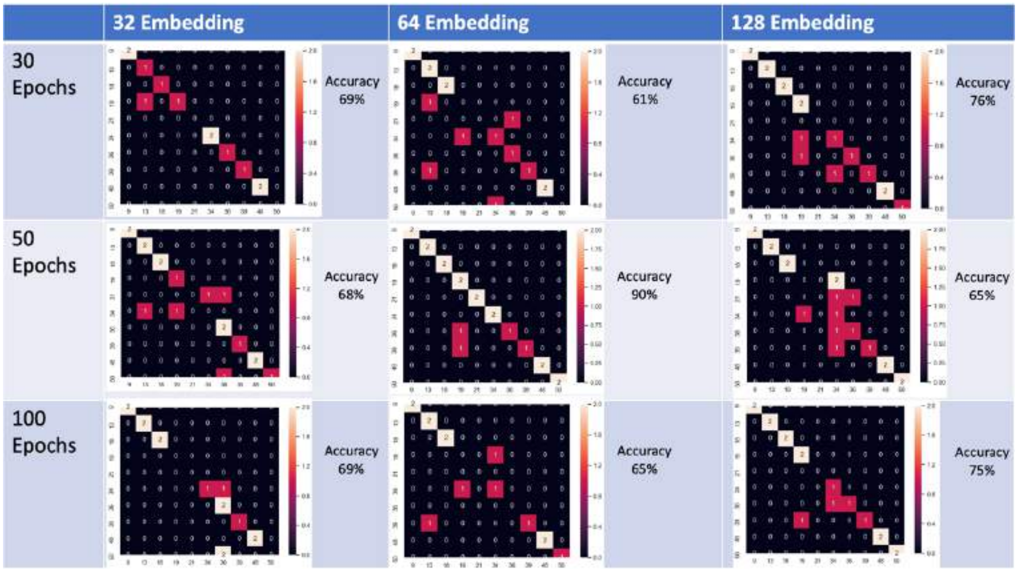


Figure 6.4: Confusion matrices of the identity estimation system for all the feature extractor model configurations of the embeddings trained by the SoftMax loss function (y-axis = Real Lion ID Number; x-axis = Lion ID estimated by the model)

6.5.2 GRAPHS OF CONFIFENCE SCORES FOR THE EMBEDDINGS TRAINED BY THE SOFTMAX LOSS FUNCTION:

A summarised representation for the graphs plotting the confidence scores for the various model configurations using the feature extractor with embeddings trained by the SoftMax loss function is shown below. The plots denote the confidences for the identities estimated by each model over 10 individuals of

Lions on two test images for each individual. The 5 individuals that are not included in the training set of the feature extractor (AF13, AF18, AF19, AF21 and AF34) are included among the 10 individuals selected for the purpose of testing.

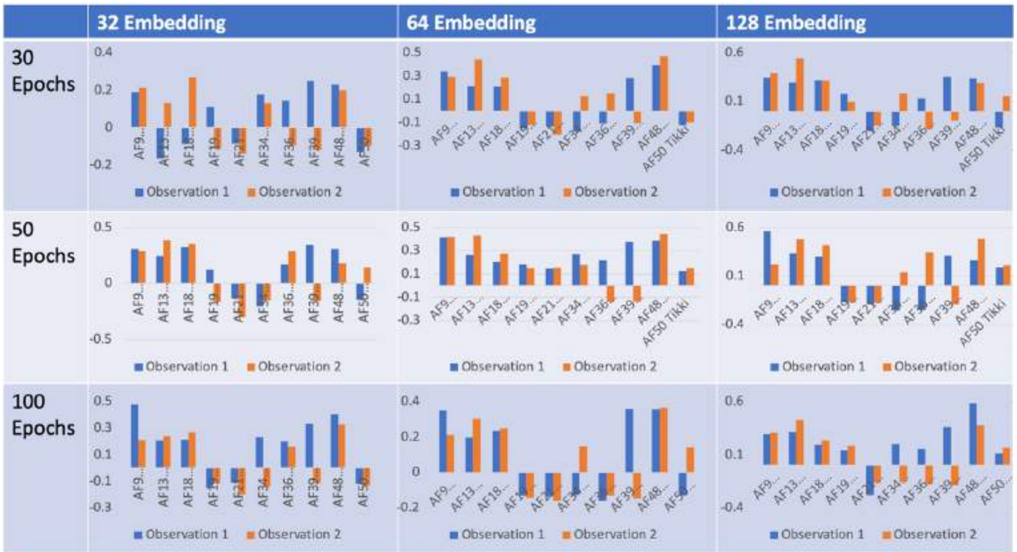


Figure 6.5: Confidence graphs of the identity estimation system for all the feature extractor model configurations of the embeddings trained by the SoftMax loss function (y-axis = Confidence; x-axis = Lion Individuals)

6.5.3 DETAILS OF THE BEST PERFORMING MODEL FOR THE EMBEDDINGS TRAINED BY THE SOFTMAX LOSS FUNCTION:

The results from the tests conducted show that, for a fully functioning system, the feature extractor model configuration with an embedding size of 64 dimensions trained over 50 epochs by the SoftMax loss function, displays an identity prediction accuracy of 90%. This accuracy is estimated by testing the model's performance over the test set with the data of 10 designated individuals, of which, the data to 5 individuals were not provided as training samples for the feature extractor. This is therefore the system configuration that performs the best within the approach of using the embeddings of the feature extractor trained by the SoftMax loss function in comparison to the other model configurations tested within the same approach. Hence a detailed report of this specific configuration is given below, which also is a reference to the full reporting process undertaken.

Table 6.5: Classification report of the identity estimation system with the feature extractor model configuration of 64 embedding dimensions trained over 50 epochs with the embeddings trained by the SoftMax loss function

	precision	recall	f1-score	support
AF9 Hasani	1	1	1	2
AF13 Lolparpit	1	1	1	2
AF18 Mooza	1	1	1	2
AF19 Moswen	0.5	1	0.66666667	2
AF21 Naibor	1	1	1	2
AF34 Sadala	1	1	1	2
AF36 Saitoti	1	0.5	0.66666667	2
AF39 Saruni	1	0.5	0.66666667	2
AF48 Summer	1	1	1	2
AF50 Tikki	1	1	1	2
accuracy	0.9	0.9	0.9	0.9
macro avg	0.95	0.9	0.9	20
weighted avg	0.95	0.9	0.9	20

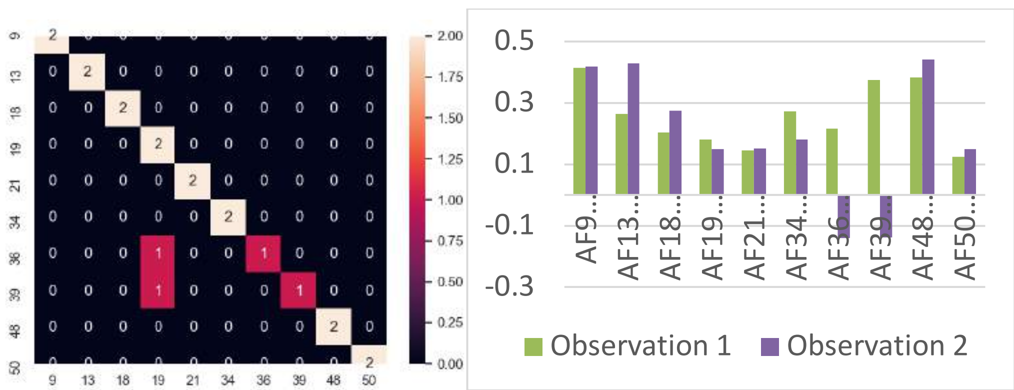


Figure 6.6: Confusion matrix and confidence graph for the identity estimation system with the feature extractor model configuration of 64 embedding dimensions trained over 50 epochs with the embeddings trained by the SoftMax loss function. Confusion matrix: (y-axis = Real Lion ID Number; x-axis = Lion ID estimated by the model). Confidence graph: (y-axis = Confidence; x-axis = Lion Individuals).

Due to the capability of this system to achieve an identity prediction accuracy of 90%, this specific configuration of the feature extractor model with the embeddings trained by the SoftMax loss function is used as the baseline for comparison with the models using the forthcoming approach.

6.6 TRIPLET LOSS AND THE FEATURE EXTRACTOR USING THE TRIPLET SIAMESE ARCHITECTURE:

This section documents the reports of the models with the different configurations that were tested for a modified approach of using the feature extractor that is a repurposed deep classifier trained over lion identities using a Triplet Siamese architecture with a triplet loss function. This approach has the architectural modification, that of, the triplet Siamese architecture with shared weights to learn over the representations of positive and negative matches with the representations of anchor images, thus optimizing the triplet loss function as described in Chapter 5. This is a unique method to train the classifier, due to the operating concept of each identity being represented as its own cluster within the embedding space of the classifier. The deep inception network classifier thus trained by the Siamese triplet method, is then repurposed to extract feature vectors.

The various subsections within this approach documents the reports to the various embedding size configurations. For each embedding size configuration, the model state for three levels to the number of epochs are considered and reports for each are documented respectively.

6.6.1 CONFUSION MATRICES FOR THE EMBEDDINGS TRAINED BY THE TRIPLET LOSS FUNCTION:

A summarised representation of the confusion matrices of the predictions produced by various feature extractor model configurations with embeddings trained by the triplet loss function is shown below. The values denote the identities estimated by each model over 10 individuals of Lions on two test images for each individual. The 5 individuals that are not included in the training set of the feature extractor (AF13, AF18, AF19, AF21 and AF34) are included among the 10 individuals selected for the purpose of testing.

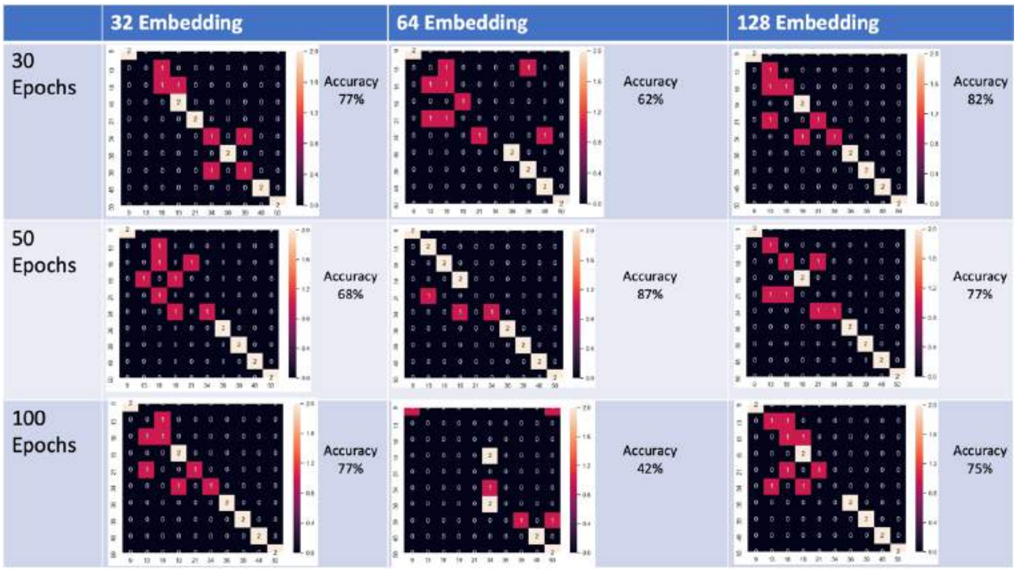


Figure 6.7: Confusion matrices of the identity estimation system for all the feature extractor model configurations of the embeddings trained by the triplet loss function (y-axis = Real Lion ID Number; x-axis = Lion ID estimated by the model)

6.6.2 GRAPHS OF CONFIFENCE SCORES FOR THE EMBEDDINGS TRAINED BY THE TRIPLET LOSS FUNCTION:

A summarised representation for the graphs plotting the confidence scores for the various model configurations using the feature extractor with embeddings trained by the triplet loss function is shown below. The plots denote the confidences for the identities estimated by each model over 10 individuals of Lions on two test images for each individual. The 5 individuals that are not included in the training set of the feature extractor (AF13, AF18, AF19, AF21 and AF34) are included among the 10 individuals selected for the purpose of testing.

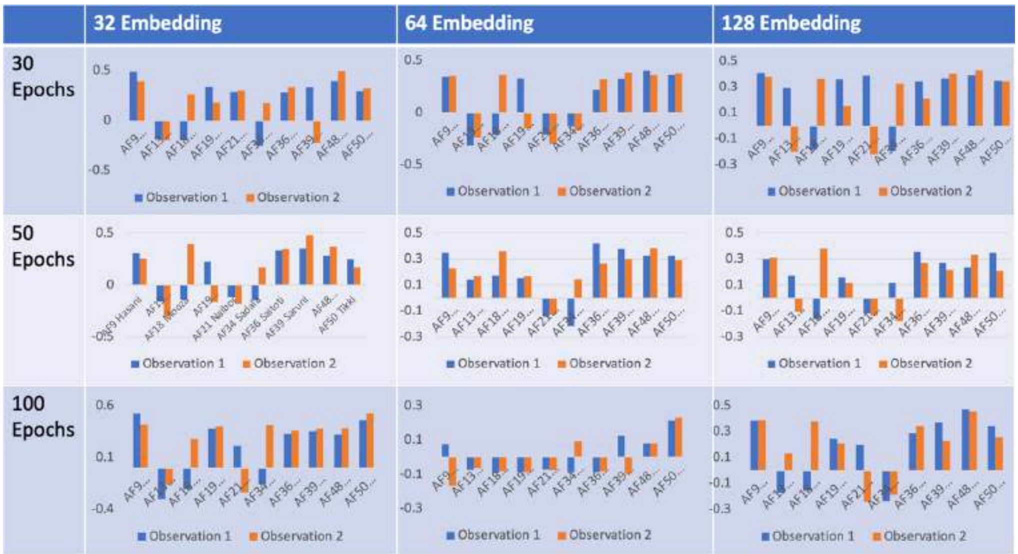


Figure 6.8: Confidence graphs of the identity estimation system for all the feature extractor model configurations of the embeddings trained by the triplet loss function (y-axis = Confidence; x-axis = Lion Individuals)

6.6.3 DETAILS OF THE BEST PERFORMING MODEL FOR THE EMBEDDINGS TRAINED BY THE TRIPLET LOSS FUNCTION:

The results from the tests conducted show that, for a fully functioning system, the feature extractor model configuration with an embedding size of 64 dimensions trained over 50 epochs by the triplet loss function, displays an identity prediction accuracy of 87%. This accuracy is estimated by testing the model's performance over the test set with the data of 10 designated individuals, of which, the data to 5 individuals were not provided as training samples for the feature extractor. This is therefore the system configuration that performs the best within the approach of using the embeddings of the feature extractor trained by the triplet loss function in comparison to the other model configurations tested within the same approach. Hence a detailed report of this specific configuration is given below, which also is a reference to the full reporting process undertaken.

Table 6.6: Classification report of the identity estimation system with the feature extractor model configuration of 64 embedding dimensions trained over 50 epochs with the embeddings trained by the triplet loss function

	precision	recall	f1-score	support
AF9 Hasani	1	1	1	2
AF13 Lolparpit	0.66666667	1	0.8	2
AF18 Mooza	1	1	1	2
AF19 Moswen	0.66666667	1	0.8	2
AF21 Naibor	0	0	0	2
AF34 Sadala	1	0.5	0.66666667	2
AF36 Saitoti	1	1	1	2
AF39 Saruni	1	1	1	2
AF48 Summer	1	1	1	2
AF50 Tikki	1	1	1	2
micro avg	0.89473684	0.85	0.87179487	20
macro avg	0.83333333	0.85	0.82666667	20
weighted avg	0.83333333	0.85	0.82666667	20

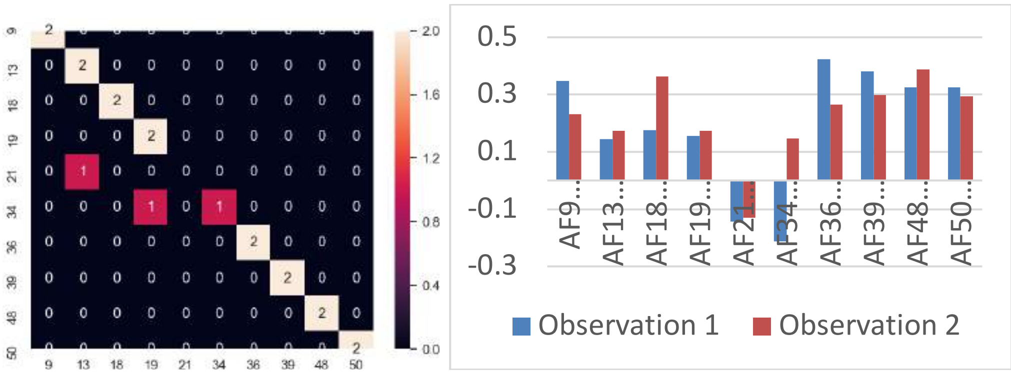


Figure 6.9: Confusion matrix and confidence graph for the identity estimation system with the feature extractor model configuration of 64 embedding dimensions trained over 50 epochs with the embeddings trained by the triplet loss function. Confusion matrix: (y-axis = Real Lion ID Number; x-axis = Lion ID estimated by the model). Confidence graph: (y-axis = Confidence; x-axis = Lion Individuals).

Due to the capability of this system to achieve an identity prediction accuracy of 87%, this specific configuration of the feature extractor model with the embeddings trained by the triplet loss function is used as the baseline for comparison with the models using the previous approach.

6.7 THE COMPARATIVE REPORT

Different comparisons have been studied to analyse both the approaches, gaining, various insights into the qualities, shortcomings as well as the benefits of the two approaches. This section, covers in detail all such comparisons documenting how each approach fared for the different model configurations that were tested.

Combining the reports to the results of the best performing models of both approaches, results to an ensembled prediction may be simulated. The reports of which are shown below.

Table 6.7: Classification report of the identity estimation system with the combined models of the best feature extractor configurations for both approaches of the embeddings trained by the SoftMax loss function as well as of the embeddings trained by the triplet loss function

	precision	recall	f1-score	support
AF9 Hasani	1	1	1	4
AF13 Lolparpit	0.8	1	0.88888889	4
AF18 Mooza	1	1	1	4
AF19 Moswen	0.57142857	1	0.72727273	4
AF21 Naibor	1	0.5	0.66666667	4
AF34 Sadala	1	0.75	0.85714286	4
AF36 Saitoti	1	0.75	0.85714286	4
AF39 Saruni	1	0.75	0.85714286	4
AF48 Summer	1	1	1	4
AF50 Tikki	1	1	1	4
micro avg	0.8974359	0.875	0.88607595	40
macro avg	0.93714286	0.875	0.88542569	40
weighted avg	0.93714286	0.875	0.88542569	40

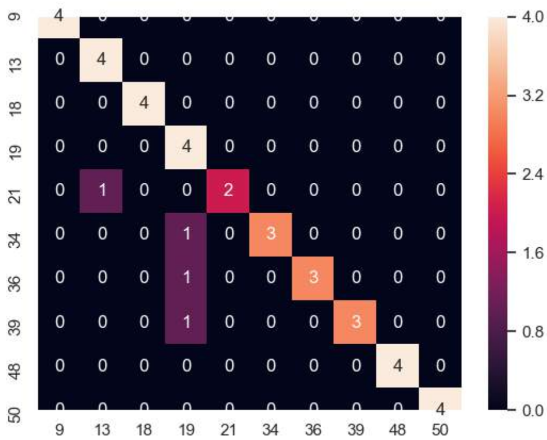


Figure 6.10: Confusion matrix of the identity estimation system with the combined models of the best feature extractor configurations for both approaches of the embeddings trained by the SoftMax loss function as well as of the embeddings trained by the triplet loss function (y-axis = Real Lion ID Number; x-axis = Lion ID estimated by the model)

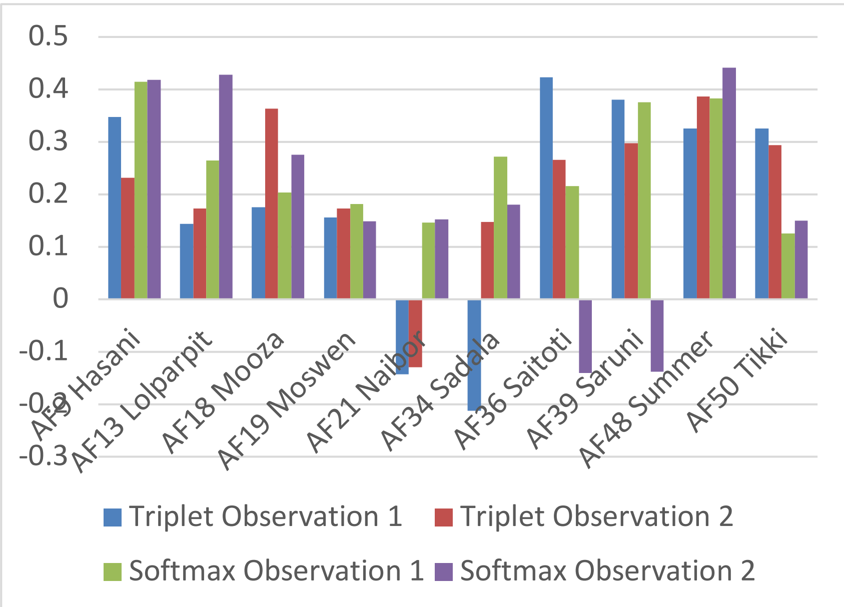


Figure 6.11: Confidence graph of the identity estimation system with the combined models of the best feature extractor configurations for both approaches of the embeddings trained by the SoftMax loss function as well as of the embeddings trained by the triplet loss function (y-axis = Confidence; x-axis = Lion Individuals)

The combined prediction of the best models within both the approaches as a simulated ensemble model for prediction has an 89% accuracy score. Although this may not be the best score among all the three approaches, i.e., including the combined approach, it encapsulates the strengths of both approaches, the lesser complexity of the softmax approach as well as the robustness of the triplet approach.

To get an overview of the various configurations as well as their accuracy scores over the test dataset of the 10 designated lion individuals, comparative graphs showing all the different configurations within each approach is generated. These overview graphs for both the approachs, that of, using the feature extractor with its embedding space trained by the softmax loss function as well as the approach of using the feature extractor with its embedding space trained by the triplet loss function is shown below (Figure 6.13 & Figure 6.12).

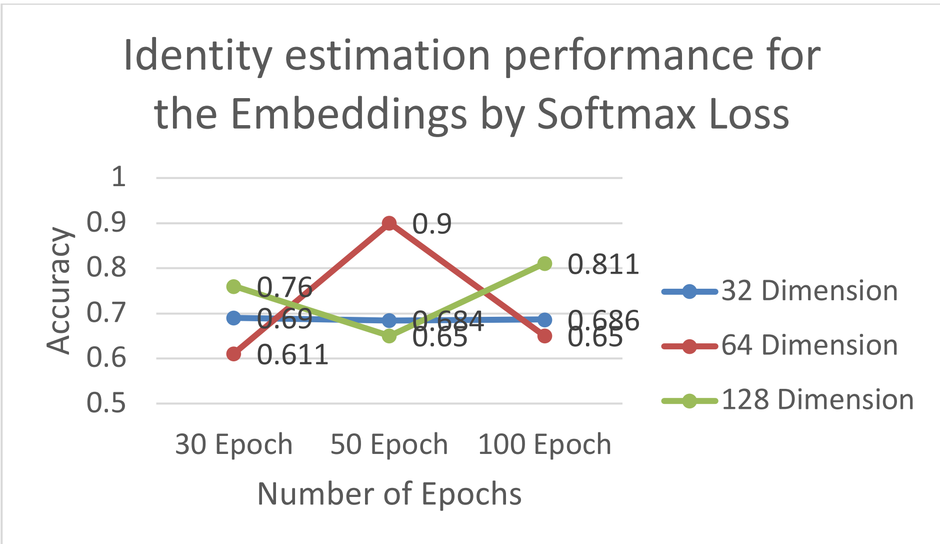


Figure 6.12: Graph showing accuracies using the Feature Extractor based on SoftMax Embedding strategy

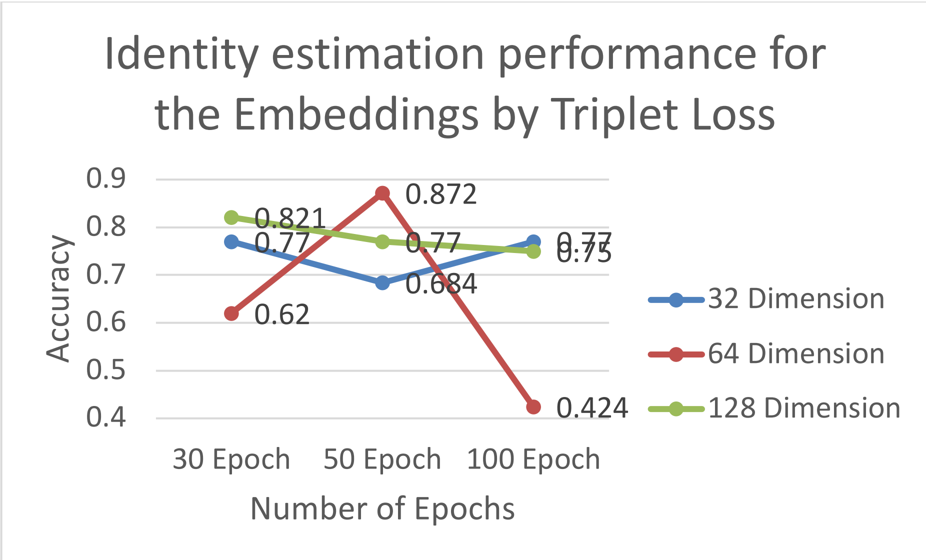


Figure 6.13: Graph showing accuracies using the Feature Extractor based on Triplet Loss Embedding strategy

Furthering the comparison it is explored how both approaches fare when put togther head to head, for the various configurations of the feature extractor models that were tested. The graphs shown below, portray the comparison between both the approaches taken, the approach of using the feature extractor with its embedding space trained by the softmax loss function as well as the approach of using the feature extractor with its embedding space trained by the triplet loss function. All the model configurations that were set up for testing over the designated test set of 10 individuals were conducted for both the approaches.

In the first setting, for the embedding size of 32 dimensions, the identity estimation system that uses the feature extractor with its embedding space trained by the triplet loss function has better accuracy scores overall, when compared to the identity estimation system that uses the feature extractor with its embedding space trained by the softmax loss function. This is seen in two configurations for both the approaches, that of, both the models trained over the configuration of 30 epochs as well as for the configuration of 100 epochs. For the configuration of 50 epochs, the models of both the approaches in the

32 embedding dimension setting, produces an equal accuracy score of identity estimation over the test set.

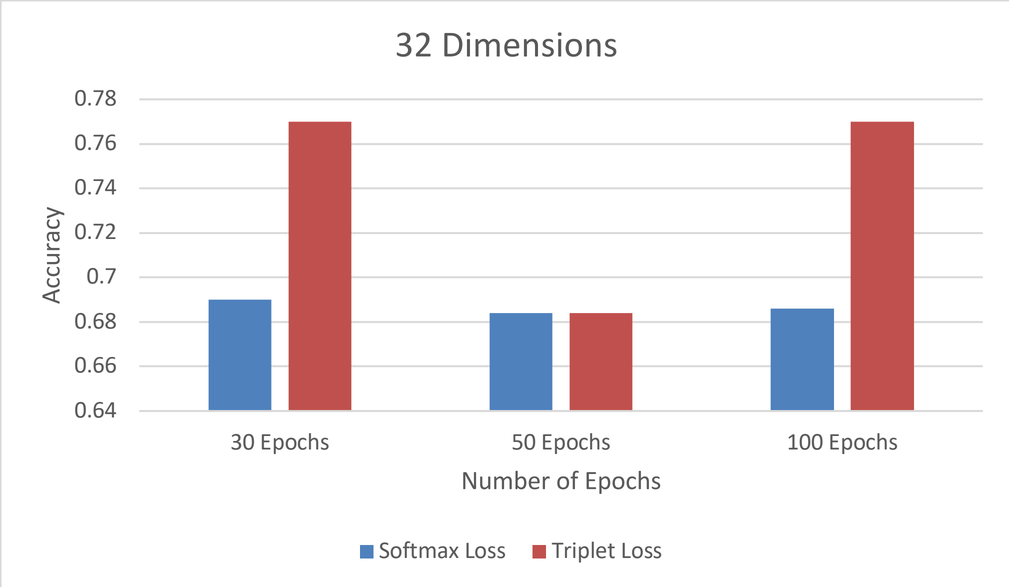


Figure 6.14: SoftMax vs Triple approach: Accuracy comparison for the Embedding size of 32 dimensions

Moving towarts the next setting, that of, having an embedding size of 64 dimensions, the identity estimation system that uses the feature extractor with its embedding space trained by the softmax loss function has better accuracy scores overall, when compared to the identity estimation system that uses the feature extractor with its embedding space trained by the triplet loss function. This is seen in two configurations for both the approaches, that of, both the models trained over the configuration of 50 epochs as well as for the configuration of 100 epochs. For the configuration of 30 epochs, the model that uses the feature extractor embeddings trained over the triplet loss function produces an accuracy score higher than that of the model that uses the feature extractor embeddings trained over the softmax loss function, although the accuracy scores of both the approaches are very similar for the 30 epoch configuration. The configuration with 50 epochs is the configuration at which both the approaches performits best, with the softmax loss approach having an accuracy of 90% and the triplet loss approach having an accuracy of 87% with a combined simulated ensemble accuracy score of 89%

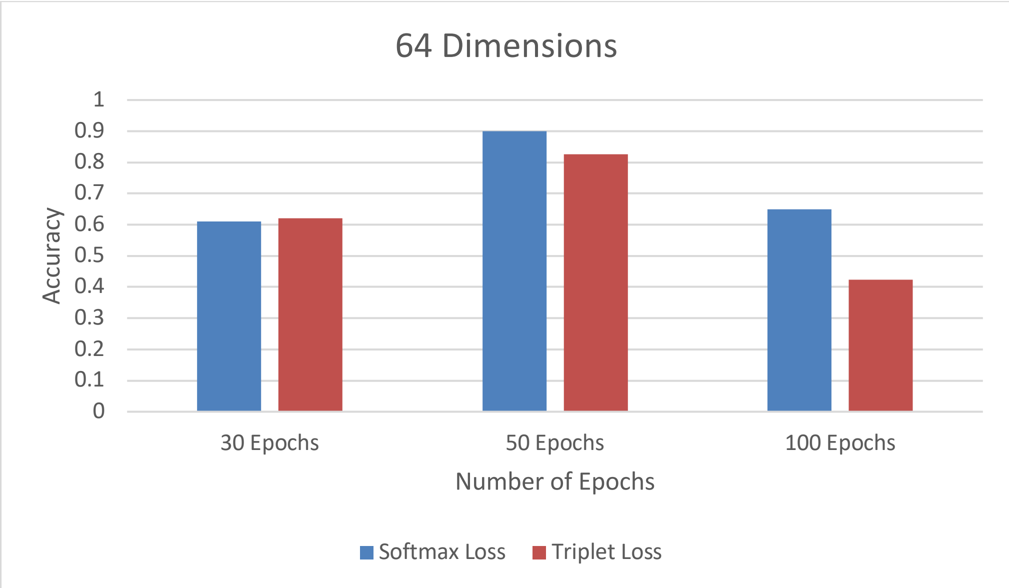


Figure 6.15: SoftMax vs Triple approach: Accuracy comparison for the Embedding size of 64 dimensions

In the final setting, with the embedding size of 128 dimensions, the identity estimation system that uses the feature extractor with its embedding space trained by the triplet loss function has better accuracy scores overall, when compared to the identity estimation system that uses the feature extractor with its embedding space trained by the softmax loss function. This is seen in two configurations for both the approaches, that of, both the models trained over the configuration of 30 epochs as well as for the configuration of 50 epochs. For the configuration of 100 epochs, the model that uses the feature extractor embeddings trained over the softmax loss function produces an accuracy score higher than that of the model that uses the feature extractor embeddings trained over the triplet loss function.

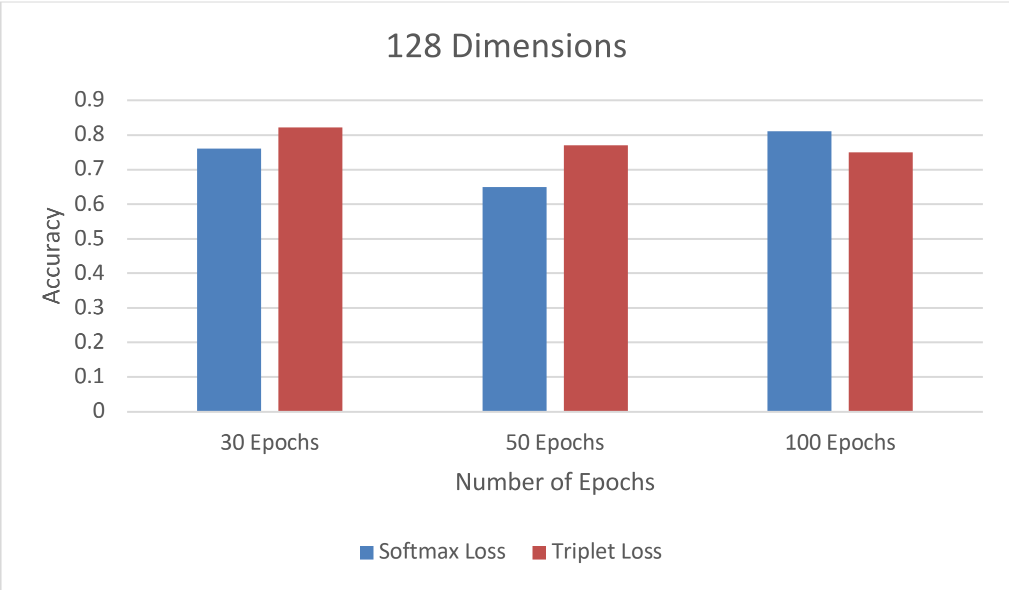


Figure 6.16: SoftMax vs Triple approach: Accuracy comparison for the Embedding size of 128 dimensions

Through these comparisons we can see that, the overall performance via the accuracy parameter for the approach using the triplet loss function, is more robust. This is evident by the fact that, out of the 9 different model configurations settings, there are 5 model configuration settings where the approach using the triplet loss function has a better accuracy score, there are 3 model configurations settings where the approach using the SoftMax loss function has a better accuracy score and there is 1 model configuration setting where both the approaches have the same accuracy score. Hence 5 out of the 9 times, the approach using the triplet loss function performs better.

Looking further we can see that by increasing the number of dimensions, the accuracy scores for the approach using the triplet loss function increases at the lower epoch settings as well. Although the configuration with the best overall performance is that of the feature extractor with the embedding size of 64 dimensions and the number of epochs set to 50.

6.8 DISCUSSION

Given the issues of data scarcity, it was tackled first by various augmentation techniques and later by the additional step of creating triplets within the approach of using the feature extractor with the embeddings trained by the triplet loss function. To have a brief understanding, while starting with an initial data set of 138 images, considering the 46 lion individuals with a minimum of 3 images each, the first data augmentation technique using the blur transformations generates a dataset with 414 images, a 300% increase, and the next augmentation step using the noise transformations generates a dataset of 1242 images, which is a 900% increase in dataset size compared to the original dataset. Furthering these dataset size gains, for the approach of using the feature extractor with embeddings trained by the triplet loss function, sets of triplet combinations are created as input for the training, this step exponentially increases the size of the input dataset given for training resulting in 15,09,030 unique triplets that are generated as the set of input data generated for training. Therefore the procedure of using triplets as a feature engineering step also increased the viability of training a deep model. Thus these techniques make it possible to efficiently tackle the data scarcity.

An exclusive validation approach of excluding 5 individuals from the training set of the feature extractor ensured that during testing, the identity embeddings of these line individuals were newly generated by way of the unique features extracted by the trained feature extractor. The performance of the feature extractor over the identities of 10 lion individuals used for testing, 5 of which were unseen by the feature extractor.

The dataset created for the purposes of training the model for lion re-identification is the first of its kind within the research space and the serves as an initial benchmark dataset. For further operations and development within this field of lion re-ID, this dataset will be made available so that it will serve as a reference for the furthering of this work.

The performance parameter used to evaluate the model's identity estimation capability is the accuracy parameter. This is calculated across the identification task of estimating the identity of 10 individuals by providing perspective crop images and unseen by the whole model (both the feature extractor as well as the classifier) which is the test set. The accuracy metric calculated over this test set serves as the performance parameter across the various models. Multiple models are generated by varying the embedding size of the feature extractor across different training epochs. The results thus gained are shown below by means of various graphs as well as confusion matrices.

Due to the nature of the triplet loss function that trains the inception network, the uniqueness of the patterns learned for each lion within the training set develops a separate cluster within the embedding space of the model learned. This is predominantly due to the hyperparameter α (margin parameter) within the triplet loss function, as well as the differences between the patterns that determine the location of these clusters within the embedding space. Nevertheless, the triplet loss approach is more robust.

To further prove robustness of the approach of using the feature extractor with embeddings trained by the triplet loss function, we may test the model against fewer augmentation processes along with more number of unique training data images.

CHAPTER 7 CONCLUSION AND FUTURE SCOPE

The initial target of getting an accuracy score of above 80% using Computer Vision for the estimation of individual lion identities has been met with a best score of 90% accuracy (using the feature extractor trained by the SoftMax loss classification approach) and a combined score of 89% accuracy (using both variants of feature extractors: the extractor trained by the Triplet Loss over the Siamese network layout as well as the extractor trained by the repurposed SoftMax loss classifier approach).

Further experimentation is done by tuning over the alpha parameter of the Triplet loss function that acts like the margin of distance between the positive and negative matches for an anchor. There were no significant gains in identification accuracy. The current margin value of 0.2 has been used for all Triplet loss-based feature extractor models.

The unique method of training by repurposing embedding spaces enables the system towards individual lion's identity and to have its own cluster within the embedding space of the learned model.

The various objectives towards the goal of individual identification have been met, including the formulation of the Computer Vision approach towards the Lion individual identification.

- Creation of a framework to generate normalized data
- Method for repurposing the embedding space of deep networks to serve as a feature extractor
- Partial automation towards identifying Asiatic Lions

At least two images of an individual is required for creating a new entry in the database, reducing the scope of ambiguity.

This work thus provides for a Computer Vision based non-invasive identification mechanism for lions that will aid the various efforts towards the monitoring and hence towards the conservation of this beloved species.

In order to prepare better models for the identification of individuals in a focal species, a better understanding of traditional algorithms must be taken into consideration. Hence, this review article presents a study of various traditional and available algorithms in order to obtain a foundation for preparing better algorithms both in time and space complexities, with respect to the existing ones.

What has been observed from the previously available sections in the article, suggests that there are a few major takeaways that keep showing up in most of the individual identification approaches.

Semi-automated methods have issues of observer ambiguity or operator subjectivity just as the manual method, suggesting observer biases to be carried over to the observations being made, and in turn creating ambiguities in the information recorded. The workarounds observed are that there must be least two independent observers/operators that are required to record the information for identification [65]. The issue extends even further, with questionable consistency of collected data over time due to different data collection methods. Also, identification based on researcher knowledge requires a lot of training, which is expensive and still has the problem of ambiguity due to biases.

The quality of the image depends on a variety of factors and plagues almost all the non-invasive techniques that use images or image based approaches as the basis for identification. This makes image quality the main factor that determines identifiability, creating all kinds of observer and observation ambiguities. Another factor related to the quality of the photographs is the distinctiveness of an individual, by which highly distinctive individuals don't need as much photographic quality as much as those with lesser distinctiveness. The method of workaround used for this issue is by having multiple judges to rate both aspects on multiple variables, and then deriving an agreement statistic of every possible pair of judges (Friday et al., 2000).

Weather conditions and factors like the time of the day also need to be taken into consideration for the best quality of the image captured, with preferences for no rain, and neither images of dusk nor dawn [37].

An issue that still prevails is that there is no system in place to quantify the degree of change in appearance over time in an individual animal [49][160]. Therefore, this work focuses on the identification of adult individuals that are generally between the range of 3 to 8 years of age.

The preliminary studies and the takeaways from the discussion suggest that, improvement in the various individual identification systems can be achieved with a few workarounds for the various processes involved. Such as, the indexing process of the identity system, can speed up the process of searching [62]. Also, while handling multiple observers without proper data recording standards, a fixed length identification code can be used to reduce the ambiguity [66].

There are few workarounds observed to combat the issue of the quality of the image which are mentioned here. One approach is the creation of a new measure of image quality and using it as a metric for evaluation, so that there may be an improvement in the confidence of the system. The idea behind this workaround is to quantify the various problems in image quality by generating a score based on the various factors like, the position of the animal, the angle of the camera, the tilt, scaling, orientation, partial capture involving glare or environmental artefacts, resolution and other image quality problems [27].

An additional specification that at least two images of an individual is required for creating a new entry in the database, reducing the scope of ambiguity due to image quality [59]. This specification of minimum number of inputs for a new entry is a necessary factor while looking into approaches that implement machine learning to create mathematical models for each individual as templates.

During the process of training the system, most of the times there is the issue of overfitting unique patterns over the model created for each individual. A novel workaround observed to address this problem is the generation of a species-space model, like a single model for the whole species [55].

For animals with an anatomy of having a muzzle, there is an issue of the possibility of correlation of patterns on both sides of the animal's muzzle, that has been addressed in previous research and the only workaround used was to consider the patterns of only one side of the muzzle for identification processes [44][9]A suggested proposition would be that of having the combined pattern of both the sides taken to be used for identification. A requirement of two images each for both sides of the muzzle as well as two front facing images could be imposed as a minimum to create an entry for a new individual.

Therefore, a standardized system model that encompasses all such requirements that wards off most ambiguities across different time sessions and across multiple researchers is the way forward. Such a system also reduces the required training of the researchers within this field, with an effect of potential acceleration in training. It is required to be able to quantify the uniquely identifiable aspects of appearances with higher accuracy by reducing noise and possible ambiguity hence, various normalization techniques must be implemented.

Towards the creation of fully automated individual identification systems, it can thus be concluded that, for non-invasive individual animal identification, modelling the system mathematically with the problem specific computational complexity in mind is necessary because there does not exist a 'one size fits all' approach [49].

The unique method of training by triplets enables the system towards individual lion's identity and to have its own cluster within the embedding space of the learned model. Furthermore, reducing the overhead of creating separate mathematical models that handle each and every problem that is native to identification within photographs, such as lighting, contrast, shadows, etc., along with other occlusions due to the various features of the lion face's anatomy. The triplet loss function reduces the problem of identification to a problem of vector embedding within the embedding space of the model being trained. This model also has the advantage of getting even better accuracy on the exposure to more lions as this data may be used to create further triplets to revise the learned clusters. Furthering this work, the system can be improved by the inclusion of both the right as well as the left mystacial areas of the lion's muzzle and studies may be conducted to test for correlation between these patterns for each lion.

7.1 FUTURE WORK

A major improvement in such a system would be the capability that could be added by a mathematical model that quantifies the degree of change in appearance over time for an individual animal [49]. This increases the accuracy for longitudinal studies that conduct repeated observation of parameters over time [94] [167]. These studies shed light on life history variables and the heritability of traits.

Additional attributes that contribute to the uniqueness and individuality of lions may also be considered to be included for creating more robust models for re-identification. Attributes such as cuts or notches on the lions ears, are such features that can be mapped towards providing more unique data per lion individual.

A fully automated system can be created by training a system to recognize perspectives for the angle of image capture. Towards a broader reach and the furthering of this work, the generation of a species-space model [107] [93] [188], like a single model for the whole species can also be done. Adding to this, working with video streams [81] [82] [195] as well as live streams is a future need to be worked out, towards the development of real-time systems [165]. Looking even further, as the world gets more virtual by the day, the 3D imaging for each lion individual as well as high resolution Iris scans for each lion individual may be considered for more granular records of individuality.

Considering the commercial angle that can be gained by the fruitfulness of this work, the methods and processes used to identify lion individuals come under the domain known as **Fine Grained Visual Categorization (FGVC)** [74] within the field of computer science. Hence the techniques under these methods may be fine-tuned towards any industry that requires fine-grained categorization such as the domain of quality control that is an integral part of multiple industries. Various industries will require varying degrees of granularity for the categorization used for the quality control procedures. For example, the clothing industry will require a lower degree of granularity whereas the pharmaceutical industry will require a higher degree of granularity of the quality control procedures. Hence the various techniques used in this work can be tweaked and fine-tuned towards such use-cases for a commercial benefit.

Furthermore, the analysis of the closeness of various clusters may reveal other correlations that may suggest similarity of these whisker patterns between close relatives of lion individuals [190]. This work thus provides for an efficient non-invasive identification mechanism for lions that will aid the various efforts towards the monitoring and hence towards the conservation of this beloved species.

REFERENCES

[1] Pennycuick, C. J., & Rudnai, J. (1970). A method of identifying individual lions Panthera leo with an analysis of the reliability of identification. Journal of Zoology, 160(4), 497-508.

[2] Miththapala, S., Seidensticker, J., Phillips, L. G., Fernando, S. B. U., & Smallwood, J. A. (1989). Identification of individual leopards (Panthera pardus kotiya) using spot pattern variation. Journal of Zoology, 218(4), 527-536.

[3] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[4] Friday, N., Smith, T. D., Stevick, P. T., & Allen, J. (2000). Measurement of photographic quality and individual distinctiveness for the photographic identification of humpback whales, Megaptera novaeangliae. Marine Mammal Science, 16(2), 355-374.

[5] Burghardt, T., Thomas, B., Barham, P. J., & Calic, J. (2004). Automated visual recognition of individual african penguins.

[6] Arzoumanian, Z., Holmberg, J., & Norman, B. (2005). An astronomical pattern-matching algorithm for computer-aided identification of whale sharks Rhincodon typus. Journal of Applied Ecology, 42(6), 999-1011.

[7] Anderson, C. J., Da Vitoria Lobo, N., Roth, J. D., & Waterman, J. M. (2010). Computer-aided photo-identification system with an application to polar bears based on whisker spot patterns. Journal of Mammalogy, 91(6), 1350-1359.

[8] Loos, A., & Ernst, A. (2013). An automated chimpanzee identification system using face detection and recognition. EURASIP Journal on Image and Video Processing, 2013(1), 49.

[9] Osterrieder, S. K., Salgado Kent, C., Anderson, C. J., Parnum, I. M., & Robinson, R. W. (2015). Whisker spot patterns: a noninvasive method of

individual identification of Australian sea lions (Neophoca cinerea). Journal of Mammalogy, 96(5), 988-997.

[10] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[11] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

[12] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[13] Nguyen, H., Maclagan, S. J., Nguyen, T. D., Nguyen, T., Flemons, P., Andrews, K., ... & Phung, D. (2017, October). Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. In 2017 IEEE international conference on data science and advanced Analytics (DSAA) (pp. 40-49). IEEE.

[14] Zhang, X., Yu, F. X., Kumar, S., & Chang, S. F. (2017). Learning spread-out local feature descriptors. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4595-4603).

[15] Mara Predator Project: http://www.livingwithlions.org/mara/ Accessed: 16-10-2020.

[16] Ahmed, S., Gaber, T., Tharwat, A., et al.: ‘Muzzle-based cattle identification using speed up robust feature approach’. IEEE Int. Conf. on Intelligent Networking and Collaborative Systems (INCOS), Taipei, 2015, pp. 99–104, doi: 10.1109/INCoS.2015.60

[17] Ardovini, A., Cinque, L., Sangineto, E.: ‘Identifying elephant photos by multi-curve matching’, Pattern Recognit.’, 2008, 41, pp. 1867–1877, doi: 10.1016/j.patcog.2007.11.010

[18] Atanbori, J., Duan, W., Murray, J., et al.: 'Automatic classification of flying bird species using computer vision techniques', Pattern Recognit. Lett., 2016, 81, (1), pp. 53–62, doi: 10.1016/j.patrec.2015.08.015

[19] Bolger, D.T., Morrison, T.A., Vance, B., et al.: 'A computer-assisted system for photographic mark–recapture analyses', Methods Ecol. Evol., 2012, 3, pp. 813–822, doi: 10.1111/j.2041-210X.2012.00212.x

[20] Carlos, J., Reyesvera, U., Possani-Espinosa, A.: 'Scorpions: Classification of poisonous species using shape features'. Int. Conf. on Electronics Communications and Computers (CONIELECOMP), Cholula, February 2016, pp. 125–129, doi: 10.1109/CONIELECOMP.2016.7438563

[21] Chen, J., Wen, Q., Qu, W., et al.: 'Panda facial region detection based on topology modelling'. 5th Int. Congress on Image and Signal Processing (CISP), Chongqing, Sichuan, China, October 2012, pp. 911–915, doi: 10.1109/CISP.2012.6469668

[22] Ernst, A., Ku blbeck, C.: 'Fast face detection and species classification of African great apes'. Eigth IEEE Int. Conf. on Advanced Video and Signal- Based Surveillance, Klagenfurt, September 2011, pp. 279–284, doi: 10.1109/AVSS.2011.6027337

[23] Hannuna, S.L., Campbell, N.W., Gibson, D.P.: 'Identifying quadruped gait in wildlife video'. IEEE Int. Conf. Image Processing (ICIP), Genova, Italy, September 2005, vol. 1, pp. 710–713, doi: 10.1109/ICIP.2005.1529850

[24] Hiby, L., Lovell, P., Patil, N., et al.: 'A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins', Biol. Lett., 2009, 5, pp. 383–386, doi: 10.1098/rsbl.2009.0028

[25] Holmberg, J., Norman, B., Arzoumanian, Z., et al.: 'Estimating population size, structure, and residency time for whale sharks Rhincodontypus through collaborative photo-identification', Endangered Species Res., 2009, 7, (1) pp. 39–53, doi: 10.3354/esr00186

[26] Jill, M.L., Buler, J.J., Frank, R.M.: 'Geographic position and landscape composition explain regional patterns of migrating landbird distributions during spring stopover along the northern coast of the Gulf of Mexico', Landsc. Ecol., 2016, 31, (8), pp. 1697–1709, doi: 10.1007/s10980-016-0354-1

[27] Kelly, M.J.: 'Computer-aided photograph matching in studies using individual identification: an example from Serengeti cheetahs', J. Mammal., 2001, 82, pp. 440–44, doi: 10.1644/1545-1542(2001)082<0440:CAPMIS>2.0.CO;2

[28] Kiapour, M.H., Jagadeesh, W., Di, V., et al.: 'Mine the fine: Fine-grained fragment discovery'. IEEE Int. Conf. on Image Processing (ICIP), Quebec City, QC, September 2015, pp. 3555–3559, doi: 10.1109/ICIP.2015.7351466

[29] Krüger, B., Yasin, H., Zsoldos, R., et al.: 'Retrieval, recognition and reconstruction of quadruped motions'. Int. Conf. on Computer Graphics Theory and Applications (GRAPP), Lisbon, Portugal, January 2014, pp. 1–8, doi: 10.5220/0004710903560363

[30] Lahiri, M., Tantipathananandh, C., Warungu, R., et al.: 'Biometric animal databases from field photographs: identification of individual zebra in the wild'. First ACM Int. Conf. on Multimedia Retrieval, Trento, Italy, April 2011, pp. 6:1s–66:8, doi: 10.1145/1991996.1992002

[31] Li, W., Song, D.: 'Automatic bird species filtering using a multimodal approach', IEEE Trans. Autom. Sci. Eng., 2015, 12, pp. 553–564

[32] Lu, A., Hou, X., Liu, C.L., et al.: 'Insect species recognition using discriminative local soft coding'. 21st Int. Conf. on Pattern Recognition (ICPR), Tsukuba, November 2012, pp. 1221–1224, doi: 10.4018/978-1-5225-5204-8.ch073

[33] Martin-Smith, K.M.: 'Photo-identification of individual weedy sea dragons Phyllopteryx taeniolatus and its application in estimating population dynamics', J. Fish Biol., 2011, 78, pp. 1757–1768, doi: 10.1111/j.1095-8649.2011.02966.x

[34] Petherick, J.C.: ‘Animal welfare issues associated with extensive livestock production: the northern Australian beef cattle industry’, Appl. Anim. Behav. Sci., 2005, 9, (3), pp. 211–234, doi: 10.1016/j.applanim.2005.05.009

[35] Piczak, K.: ‘Recognizing bird species in audio recordings using deep convolutional neural networks’. Int. Conf. CLEF Working Notes, 2016, pp. 273–288, Doi: 10.1109/CVPR.2015.7298658

[36] Sagonas, C., Panagakis, Y., Zafeiriou, S., et al.: ‘Robust statistical frontalization of human and animal faces’, Int. J. Comput. Vis., 2016, pp. 1– 22, doi: 10.1007/s11263-016-0920-7

[37] Sherley, R.B., Burghardt, T., Barham, P.J., et al.: ‘Spotting the difference: towards fully automated population monitoring of African penguins Spheniscus demersus’, Endangered Species Res., 2010, 11, pp. 101–111, doi: 10.3354/esr00267

[38] Song, D., Qin, N., Xu, Y., et al.: ‘System and algorithms for an autonomous observatory assisting the search for the ivory-billed woodpecker’. IEEE Int. Conf. on Automation Science and Engineering, Arlington, VA, August 2008, pp. 200–205

[39] Takeki, A., Trinh, T.T., Yoshihashi, R., et al.: ‘Detection of small birds in large images by combining a deep detector with semantic segmentation’. IEEE Int. Conf. Image Processing, Phoenix, AZ, USA, September 2016, pp. 3977–3981, doi: 10.1109/ICIP.2016.7533106

[40] Van Horn, G., Branson, S., Farrell, R., et al.: ‘Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection’. IEEE Conf. on Computer Vision and Pattern Recognition, Boston, MA, June 2015, pp. 595–604, doi: 10.1109/CVPR.2015.7298658

[41] Wilson, D.J., Lyver, P.O.B., Greene, T.C., et al.: ‘South Polar Skua breeding populations in the Ross Sea assessed from demonstrated relationship with Adélie Penguin numbers’, Polar Biol., pp. 1–16, doi:10.1007/ s00300-016-1980-4

[42] Zhang, W., Sun, J., Tang, X.: 'From tiger to panda: animal head detection', IEEE Trans. Image Process., 2011, 20, pp. 1696–1708, doi: 10.1109/TIP.2010.2099126

[43] Zhang, W.: 'Cat head detection: how to effectively exploit shape and texture features', Lect. Notes Comput. Sci., 2008, 5305, pp. 802–816, doi: 10.1007/978-3-540-88693-8_59

[44] Anderson, C. (2007). Individual identification of polar bears by whisker spot patterns.

[45] Araabi, B. N., Kehtarnavaz, N., McKinney, T., Hillman, G., & Wu¨rsig, B. (2000). A string matching computer-assisted system for dolphin photoidentification. Annals of Biomedical Engineering, 28(10), 1269-1279.

[46] Arzoumanian, Z., Holmberg, J., & Norman, B. (2005). An astronomical pattern-matching algorithm for computer-aided identification of whale sharks Rhincodon typus. Journal of Applied Ecology, 42(6), 999-1011.

[47] Burghardt, T. (2008). Visual animal biometrics: automatic detection and individual identification by coat pattern (Doctoral dissertation, University of Bristol).

[48] Chase, F. W., Hardie, B. E., Kern, M. M., Pittman, S. E., Harden, L. A., & Dorcas, M. E. (2015). Evaluation of two individual identification techniques for spotted salamanders (Ambystoma maculatum). Herpetological Review, 46, 192-196.

[49] Crouse, D., Jacobs, R. L., Richardson, Z., Klum, S., Jain, A., Baden, A. L., & Tecot, S. R. (2017). LemurFaceID: a face recognition system to facilitate individual identification of lemurs. Bmc Zoology, 2(1), 2.

[50] Dixon, D. R. (2003). A non-invasive technique for identifying individual badgers Meles meles. Mammal Review, 33(1), 92-94.

[51] Foster, G., Krijger, H., & Bangay, S. (2007). Zebra fingerprints: towards a computer-aided identification system for individual zebra. African Journal of Ecology, 45(2), 225-227.

[52] Freytag, A., Rodner, E., Simon, M., Loos, A., Ku¨hl, H. S., & Denzler,

J. (2016, September). Chimpanzee faces in the wild: Log-euclidean cnns for predicting identities and attributes of primates. In German Conference on Pattern Recognition (pp. 51-63). Springer, Cham.

[53] Gamble, L., Ravela, S., & McGarigal, K. (2008). Multi-scale features for identifying individuals in large biological databases: an application of pattern recognition technology to the marbled salamander Ambystoma opacum. Journal of Applied Ecology, 45(1), 170-180.

[54] Hillman, G. R., Wursig, B., Gailey, G. A., Kehtarnavaz, N., Drobyshevsky, A., Araabi, B. N., ... & Weller, D. W. (2003). Computer-assisted photo-identification of individual marine vertebrates: a multi-species system. Aquatic Mammals, 29(1), 117-123.

[55] Hughes, B., & Burghardt, T. (2017). Automated visual fin identification of individual great white sharks. International Journal of Computer Vision, 122(3), 542-557.

[56] Jonas, M., Bowman, J. L., & Nazdrowicz, N. H. (2011). Using spot pattern to identify individual Long-tailed salamanders. Herpetol. Rev, 42(4), 520-522.

[57] Karanth, K. U., & Nichols, J. D. (1998). Estimation of tiger densities in India using photographic captures and recaptures. Ecology, 79(8), 2852-2862.

[58] McConkey, S. D. (1999). Photographic identification of the New Zealand sea lion: a new technique.

[59] Mizroch, S. A., Beard, J. A., & Lynde, M. (1990). Computer assisted photo-identification of humpback whales. Report of the International Whaling Commission, 12, 63-70.

[60] Pennycuick, C. J. (1978). Identification using natural markings. In Animal marking (pp. 147-159). Palgrave, London.

[61] Plaˇia¸su, R., Hartel, T., Baˇncilˇa, R. I., & Cogaˇlniceanu, D. (2005). The use of digital images for the individual identification of amphibians. Studii ¸si Cercetˇari, 10, 137-140.

[62] Scott, D. K. (1978). Identification of individual Bewick’s swans by bill

patterns. In Animal marking (pp. 160-168). Palgrave, London.

[63] Sharp, T., Saunders, G., & Mitchell, B. (2007). Marking of pest animals used in research. Institute for Laboratory Animal Research Journal, 44, 1-15.

[64] Sherley, R. B., Burghardt, T., Barham, P. J., Campbell, N., & Cuthill, I. C. (2010). Spotting the difference: towards fully-automated population monitoring of African penguins Spheniscus demersus. Endangered Species Research, 11(2), 101-111.

[65] Vincent, C., Meynier, L., & Ridoux, V. (2001). Photo-identification in grey seals: legibility and stability of natural markings. Mammalia, 65(3), 363-372.

[66] Whitehead, H. (1990). Computer assisted individual identification of sperm whale flukes. Report of the International Whaling Commission, 12, 71-77.

[67] Bakliwal, K., & Ravela, S. (2020). The Sloop System for Individual Animal Identification with Deep Learning. arXiv preprint arXiv:2003.00559.

[68] BELLO, R. W., TALIB, A. Z. H., & MOHAMED, A. S. A. B. (2020). Deep learning-based architectures for recognition of cow using cow nose image pattern. Gazi University Journal of Science, 33(3), 831-844.

[69] Matthews, I., & Baker, S. (2004). Active appearance models revisited. International journal of computer vision, 60(2), 135-164.

[70] Jhala, Y. V., Banerjee, K., Chakrabarti, S., Basu, P., Singh, K., Dave, C., & Gogoi, K. (2019). Asiatic lion: Ecology, economics, and politics of conservation. Frontiers in Ecology and Evolution, 7, 312.

[71] Jo, J., Jeong, S., & Kang, P. (2020, February). Benchmarking GPU-accelerated edge devices. In 2020 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 117-120). IEEE.

[72] Xuan, H., Stylianou, A., & Pless, R. (2020). Improved embeddings with easy positive triplet mining. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2474-2482).

[73] Pambala, A., Dutta, T., & Biswas, S. (2020). Generative model with semantic embedding and integrated classifier for generalized zero-shot learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1237-1246).

[74] Hanselmann, H., & Ney, H. (2020). ELoPE: Fine-grained visual classification with efficient localization, pooling and embedding. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1247-1256).

[75] Park, G., Han, C., Yoon, W., & Kim, D. (2020). MHSAN: Multi-Head Self-Attention Network for Visual Semantic Embedding. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1518-1526).

[76] Rahimpour, A., & Qi, H. (2020). Class-discriminative feature embedding for meta-learning based few-shot classification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3179-3187).

[77] Gairola, S., Shah, R., & Narayanan, P. J. (2020). Unsupervised image style embeddings for retrieval and recognition tasks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3281-3289).

[78] Khatun, A., Denman, S., Sridharan, S., & Fookes, C. (2020). Semantic consistency and identity mapping multi-component generative adversarial network for person re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2267-2276).

[79] Chen, H., Lagadec, B., & Bremond, F. (2020). Learning discriminative and generalizable representations by spatial-channel partition for person re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2483-2492).

[80] Kumar, D., Siva, P., Marchwica, P., & Wong, A. (2020). Unsupervised domain adaptation in person re-id via k-reciprocal clustering and large-

scale heterogeneous environment synthesis. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2645-2654).

[81] Matiyali, N., & Sharma, G. (2020). Video person re-identification using learned clip similarity aggregation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2655-2664).

[82] Li, M., Xu, H., Wang, J., Li, W., & Sun, Y. (2020). Temporal aggregation with clip-level attention for video-based person re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3376-3384).

[83] Yuan, Y., Chen, W., Chen, T., Yang, Y., Ren, Z., Wang, Z., & Hua, G. (2020). Calibrated domain-invariant learning for highly generalizable large scale re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3589-3598).

[84] Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods in Ecology and Evolution, 10(4), 461-470.

[85] Schneider, S., Taylor, G. W., & Kremer, S. C. (2020). Similarity learning networks for animal individual re-identification-beyond the capabilities of a human observer. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (pp. 44-52).

[86] Ravoor, P. C., & Sudarshan, T. S. B. (2020). Deep Learning Methods for Multi-Species Animal Re-identification and Tracking–a Survey. Computer Science Review, 38, 100289.

[87] Murali, N., Schneider, J., Levine, J., & Taylor, G. (2019, January). Classification and re-identification of fruit fly individuals across days with convolutional neural networks. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 570-578). IEEE.

[88] Schneider, S., Taylor, G. W., & Kremer, S. (2018, May). Deep learning object detection methods for ecological camera trap data. In 2018 15th Conference on computer and robot vision (CRV) (pp. 321-328). IEEE.

[89] Nepovinnykh, E., Eerola, T., & Kalviainen, H. (2020). Siamese network based pelage pattern matching for ringed seal re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (pp. 25-34).

[90] Kulits, P., Wall, J., Bedetti, A., Henley, M., & Beery, S. (2021, June). ElephantBook: A Semi-Automated Human-in-the-Loop System for Elephant Re-Identification. In ACM SIGCAS Conference on Computing and Sustainable Societies (pp. 88-98).

[91] Beery, S., Morris, D., & Yang, S. (2019). Efficient pipeline for camera trap image review. arXiv preprint arXiv:1907.06772.

[92] Beery, S., Van Horn, G., Mac Aodha, O., & Perona, P. (2019). The iwildcam 2018 challenge dataset. arXiv preprint arXiv:1904.05986.

[93] Beery, S., Cole, E., Parker, J., Perona, P., & Winner, K. (2021, June). Species distribution modeling for machine learning practitioners: A review. In ACM SIGCAS Conference on Computing and Sustainable Societies (pp. 329-348).

[94] Beery, S., Wu, G., Rathod, V., Votel, R., & Huang, J. (2020). Context r-cnn: Long term temporal context for per-camera object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 13075-13085).

[95] Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., & Clune, J. (2021). A deep active learning system for species identification and counting in camera trap images. Methods in Ecology and Evolution, 12(1), 150-161.

[96] Das, T., Bruintjes, R. J., Lengyel, A., van Gemert, J., & Beery, S. (2021). Domain Adaptation for Rare Classes Augmented with Synthetic Samples. arXiv preprint arXiv:2110.12216.

[97] Moskvyak, O., Maire, F., Armstrong, A. O., Dayoub, F., &

Baktashmotlagh, M. (2019). Robust re-identification of manta rays from natural markings by learning pose invariant embeddings. arXiv preprint arXiv:1902.10847.

[98] Moskvyak, O., Maire, F., Dayoub, F., & Baktashmotlagh, M. (2021). Keypoint-aligned embeddings for image retrieval and re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 676-685).

[99] Moskvyak, O., Maire, F., Dayoub, F., & Baktashmotlagh, M. (2020). Learning landmark guided embeddings for animal re-identification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (pp. 12-19).

[100] Moskvyak, O., Maire, F., Dayoub, F., & Baktashmotlagh, M. (2021). Semi-supervised Keypoint Localization. arXiv preprint arXiv:2101.07988.

[101] Moskvyak, O., & Maire, F. (2017). Learning geometric equivalence between patterns using embedding neural networks. In 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.

[102] Holmberg, J. A., Parham, J. R., & Blount, A. (2021). Feasibility Analysis: Using Artificial Intelligence to Match Photographed Lateral Ridges of Gray Whales. Camarillo (CA): US Department of the Interior, Bureau of Ocean Energy Management, Pacific OCS Region. OCS Study BOEM, 59, 29.

[103] Nepovinnykh, E., Eerola, T., Kälviäinen, H., & Radchenko, G. (2018, September). Identification of saimaa ringed seal individuals using transfer learning. In International Conference on Advanced Concepts for Intelligent Vision Systems (pp. 211-222). Springer, Cham.

[104] Chelak, I., Nepovinnykh, E., Eerola, T., Kalviainen, H., & Belykh, I. (2021). Eden: Deep feature distribution pooling for saimaa ringed seals pattern matching. arXiv preprint arXiv:2105.13979.

[105] Weideman, H., Stewart, C., Parham, J., Holmberg, J., Flynn, K.,

Calambokidis, J., ... & Lepirei, J. (2020). Extracting identifying contours for African elephants and humpback whales using a learned appearance model. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1276-1285).

[106] Parham, J., Stewart, C., Crall, J., Rubenstein, D., Holmberg, J., & Berger-Wolf, T. (2018, March). An animal detection pipeline for identification. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1075-1083). IEEE.

[107] Berger-Wolf, T. Y., Rubenstein, D. I., Stewart, C. V., Holmberg, J. A., Parham, J., Menon, S., ... & Joppa, L. (2017). Wildbook: Crowdsourcing, computer vision, and data science for conservation. arXiv preprint arXiv:1710.08880.

[108] Dunbar, S. G., Anger, E. C., Parham, J. R., Kingen, C., Wright, M. K., Hayes, C. T., ... & Baumbach, D. S. (2021). HotSpotter: Using a computer-driven photo-id application to identify sea turtles. Journal of Experimental Marine Biology and Ecology, 535, 151490.

[109] Sun, Y., & Murata, N. (2020). CAFM: A 3D morphable model for animals. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (pp. 20-24).

[110] Duyck, J., Finn, C., Hutcheon, A., Vera, P., Salas, J., & Ravela, S. (2015). Sloop: A pattern retrieval engine for individual animal identification. Pattern Recognition, 48(4), 1059-1073.

[111] Marstaller, J., Tausch, F., & Stock, S. (2019). Deepbees-building and scaling convolutional neuronal nets for fast and large-scale visual monitoring of bee hives. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 0-0).

[112] Tausch, F., Schmidt, K., & Diehl, M. (2020). Current achievements and future developments of a novel AI based visual monitoring of beehives in ecotoxicology and for the monitoring of landscape structures. bioRxiv.

[113] Tausch, F., Stock, S., Fricke, J., & Klein, O. (2020). Bumblebee Re-

Identification Dataset. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (pp. 35-37).

[114] Bruslund Haurum, J., Karpova, A., Pedersen, M., Hein Bengtson, S., & Moeslund, T. B. (2020). Re-identification of zebrafish using metric learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (pp. 1-11).

[115] Pedersen, M., Haurum, J. B., Bengtson, S. H., & Moeslund, T. B. (2020). 3D-ZEF: A 3D zebrafish tracking benchmark dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2426-2436).

[116] Deepthi, D. R., Krishna, G. R., & Eswaran, K. (2007). Automatic pattern classification by unsupervised learning using dimensionality reduction of data with mirroring neural networks. arXiv preprint arXiv:0712.0938.

[117] Vijayarani, S., & Vinupriya, M. (2015). An efficient algorithm for facial image classification. Int J Signal Process Image Process Pattern Recogn, 8(11), 121-134.

[118] Al-Mahadeen, B., AlTarawneh, M. S., & AlTarawneh, I. H. (2010). Signature region of interest using auto cropping. arXiv preprint arXiv:1004.3549.

[119] Lewis, J. C. (1970). Wildlife census methods: a resume. Journal of wildlife diseases, 6(4), 356-364.

[120] Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences, 115(25), E5716-E5725.

[121] Karami, E., Prasad, S., & Shehata, M. (2017). Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images. arXiv preprint arXiv:1710.02726.

[122] Bolger, D. T., Morrison, T. A., Vance, B., Lee, D., & Farid, H. (2012).

A computer-assisted system for photographic mark–recapture analysis. Methods in Ecology and Evolution, 3(5), 813-822.

[123] Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). A semi-automatic methodology for facial landmark annotation. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 896-903).

[124] Hassanein, A. S., Mohammad, S., Sameer, M., & Ragab, M. E. (2015). A survey on Hough transform, theory, techniques and applications. arXiv preprint arXiv:1502.02160.

[125] Gierer, A., & Meinhardt, H. (1972). A theory of biological pattern formation. Kybernetik, 12(1), 30-39.

[126] Alabort-i-Medina, J., & Zafeiriou, S. (2017). A unified framework for compositional fitting of active appearance models. International Journal of Computer Vision, 121(1), 26-64.

[127] Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. IEEE Transactions on pattern analysis and machine intelligence, 23(6), 681-685.

[128] Tzimiropoulos, G., Alabort-i-Medina, J., Zafeiriou, S. P., & Pantic, M. (2014). Active orientation models for face alignment in-the-wild. IEEE transactions on information forensics and security, 9(12), 2024-2034.

[129] Cootes, T. F., & Taylor, C. J. (1992). Active shape models—'smart snakes'. In BMVC92 (pp. 266-275). Springer, London.

[130] Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. Computer vision and image understanding, 61(1), 38-59.

[131] Antonakos, E., Alabort-i-Medina, J., & Zafeiriou, S. (2015). Active pictorial structures. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5435-5444).

[132] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[133] Papandreou, G., & Maragos, P. (2008, June). Adaptive and constrained

algorithms for inverse compositional active appearance model fitting. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE.

[134] Rahtu, E., Salo, M., & Heikkila, J. (2005). Affine invariant pattern recognition using multiscale autoconvolution. IEEE Transactions on pattern analysis and machine intelligence, 27(6), 908-918.

[135] PETERSEN, J. B. (1972). An identification system for zebra (Equus burchelli, Gray). African Journal of Ecology, 10(1), 59-63.

[136] Lucas, B. D., & Kanade, T. (1981, April). An iterative image registration technique with an application to stereo vision.

[137] Burghardt, T., & Calic, J. (2006). Analysing animal behaviour in wildlife videos using face detection and tracking. IEE Proceedings-Vision, Image and Signal Processing, 153(3), 305-312.

[138] Trnovszky, T., Kamencay, P., Orjesek, R., Benco, M., & Sykora, P. (2017). Animal recognition system based on convolutional neural network. Advances in Electrical and Electronic Engineering, 15(3), 517-525.

[139] Kumar, S., Singh, S. K., Singh, R., & Singh, A. K. (2018). Animal Biometrics: Techniques and Applications. Springer.

[140] Greenberg, S., Godin, T., & Whittington, J. (2019). Design patterns for wildlife-related camera trap image analysis. Ecology and evolution, 9(24), 13706-13730.

[141] Delac, K., Grgic, M., & Liatsis, P. (2005, June). Appearance-based statistical methods for face recognition. In 47th International Symposium ELMAR-2005 (pp. 151-158).

[142] Yu, X., Wang, J., Kays, R., Jansen, P. A., Wang, T., & Huang, T. (2013). Automated identification of animal species in camera trap images. EURASIP Journal on Image and Video Processing, 2013(1), 1-10.

[143] Cheema, G. S., & Anand, S. (2017, September). Automatic detection and recognition of individuals in patterned species. In Joint European

Conference on Machine Learning and Knowledge Discovery in Databases (pp. 27-38). Springer, Cham.

[144] Alabort-i-Medina, J., & Zafeiriou, S. (2014). Bayesian active appearance models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3438-3445).

[145] Chandrasekaran, A., & Thuraisingham, B. (2007, April). Fingerprint matching algorithm based on tree comparison using ratios of relational distances. In The Second International Conference on Availability, Reliability and Security (ARES'07) (pp. 273-280). IEEE.

[146] Anderson, C. J. R., Roth, J. D., & Waterman, J. M. (2007). Can whisker spot patterns be used to identify individual polar bears?. Journal of Zoology, 273(4), 333-339.

[147] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.

[148] Liang, X., Liu, S., Shen, X., Yang, J., Liu, L., Dong, J., ... & Yan, S. (2015). Deep human parsing with active template regression. IEEE transactions on pattern analysis and machine intelligence, 37(12), 2402-2414.

[149] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1701-1708).

[150] Saragih, J. M., Lucey, S., & Cohn, J. F. (2011). Deformable model fitting by regularized landmark mean-shift. International journal of computer vision, 91(2), 200-215.

[151] Wu, Y. N., Si, Z., Fleming, C., & Zhu, S. C. (2007, October). Deformable template as active basis. In 2007 IEEE 11th International Conference on Computer Vision (pp. 1-8). IEEE.

[152] Plăiaşu, R., Hartel, T., Băncilă, R. I., & Cogălniceanu, D. (2005). The

use of digital images for the individual identification of amphibians. Studii şi Cercetări, 10, 137-140.

[153] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

[154] Baker, S., & Matthews, I. (2001, December). Equivalence and efficiency of image alignment algorithms. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 (Vol. 1, pp. I-I). IEEE.

[155] Hoffmann, U., Naruniec, J., Yazdani, A., & Ebrahimi, T. (2008). Face detection using discrete gabor jets and color information. In Proceedings of SIGMAP 2008-International Conference on Signal Processing and Multimedia Applications (No. CONF, pp. 76-83).

[156] Ashraf, A. B., Lucey, S., & Chen, T. (2010, June). Fast image alignment in the fourier domain. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 2480-2487). IEEE.

[157] Cristinacce, D., & Cootes, T. F. (2006, September). Feature detection and tracking with constrained local models. In Bmvc (Vol. 1, No. 2, p. 3).

[158] Antonakos, E., Alabort-i-Medina, J., Tzimiropoulos, G., & Zafeiriou, S. P. (2015). Feature-based lucas–kanade and active appearance models. IEEE Transactions on Image Processing, 24(9), 2617-2632.

[159] Heckbert, P. S. (1989). Fundamentals of texture mapping and image warping. Master's thesis. University of California, Berkeley, 2, 3.

[160] Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91). PMLR.

[161] Gross, R., Matthews, I., & Baker, S. (2005). Generic vs. person specific active appearance models. Image and Vision Computing, 23(12), 1080-1093.

[162] Martins, P., Henriques, J. F., & Batista, J. (2020). Gradient Shape

Model. International Journal of Computer Vision, 128, 2828-2848.
[163] Jain, A. K., & Li, S. Z. (2011). Handbook of face recognition (Vol. 1). New York: springer.
[164] Antonakos, E., Alabort-i-Medina, J., Tzimiropoulos, G., & Zafeiriou, S. (2014, October). Hog active appearance models. In 2014 IEEE International Conference on Image Processing (ICIP) (pp. 224-228). IEEE.
[165] Ross, D. A., Lim, J., Lin, R. S., & Yang, M. H. (2008). Incremental learning for robust visual tracking. International journal of computer vision, 77(1), 125-141.
[166] Burghardt, T., & Campbell, N. (2007, December). Individual animal identification using visual biometrics on deformable coat patterns. In International Conference on Computer Vision Systems: Proceedings (2007).
[167] Clutton-Brock, T., & Sheldon, B. C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. Trends in ecology & evolution, 25(10), 562-573.
[168] Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1867-1874).
[169] Saragih, J., & Göcke, R. (2009). Learning AAM fitting through simulation. Pattern Recognition, 42(11), 2628-2636.
[170] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008, October). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition.
[171] Belhumeur, P. N., Jacobs, D. W., Kriegman, D. J., & Kumar, N. (2013). Localizing parts of faces using a consensus of exemplars. IEEE transactions on pattern analysis and machine intelligence, 35(12), 2930-2940.

[172] Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. Josa a, 4(3), 519-524.

[173] Baker, S., & Matthews, I. (2004). Lucas-kanade 20 years on: A unifying framework. International journal of computer vision, 56(3), 221-255.

[174] Alabort-i-Medina, J., Antonakos, E., Booth, J., Snape, P., & Zafeiriou, S. (2014, November). Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 679-682).

[175] Mil'Shtein, S., Pillai, A., Shendye, A., Liessner, C., & Baier, M. (2008, May). Fingerprint recognition algorithms for partial and full fingerprints. In 2008 IEEE Conference on Technologies for Homeland Security (pp. 449-452). IEEE.

[176] Hill, A., & Taylor, C. J. (1992). Model-based image interpretation using genetic algorithms. Image and Vision Computing, 10(5), 295-300.

[177] González, S., & Duarte, J. M. B. (2007). Non invasive methods for genetic analysis applied to ecological and behavioral studies in Latino-America. Revista brasileira de Zootecnia, 36, 89-92.

[178] Arandjelovic, M., Head, J., Rabanal, L. I., Schubert, G., Mettke, E., Boesch, C., ... & Vigilant, L. (2011). Non-invasive genetic monitoring of wild central chimpanzees. PLoS One, 6(3), e14761.

[179] TABERLET, P., & LUIKART, G. (1999). Non-invasive genetic sampling and individual identification. Biological journal of the linnean society, 68(1-2), 41-55.

[180] Chernov, N., Lesort, C., & Simányi, N. (2004). On the complexity of curve fitting algorithms. Journal of complexity, 20(4), 484-492.

[181] Evangelidis, G. D., & Psarakis, E. Z. (2008). Parametric image alignment using enhanced correlation coefficient maximization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(10), 1858-1865.

[182] Martin-Smith, K. M. (2011). Photo-identification of individual weedy seadragons Phyllopteryx taeniolatus and its application in estimating

population dynamics. Journal of Fish Biology, 78(6), 1757-1768.

[183] Wang, S., Yi, X., & Chen, Y. (2017, May). Piecewise affine warp based frontal face synthesizing and application on face recognition. In 2017 29th Chinese Control And Decision Conference (CCDC) (pp. 3098-3102). IEEE.

[184] Burghardt, T., & Calic, J. (2006, September). Real-time face detection and tracking of animals. In 2006 8th seminar on neural network applications in electrical engineering (pp. 27-32). IEEE.

[185] Asthana, A., Zafeiriou, S., Cheng, S., & Pantic, M. (2013). Robust discriminative response map fitting with constrained local models. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3444-3451).

[186] Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. International journal of computer vision, 1(4), 321-331.

[187] Speed, C. W., Meekan, M. G., & Bradshaw, C. J. (2007). Spot the match–wildlife photo-identification using information theory. Frontiers in zoology, 4(1), 1-11.

[188] Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2012). Subspace learning from image gradient orientations. IEEE transactions on pattern analysis and machine intelligence, 34(12), 2454-2466.

[189] Lu, C., & Tang, X. (2015, March). Surpassing human-level face verification performance on LFW with GaussianFace. In Twenty-ninth AAAI conference on artificial intelligence.

[190] Rudnai, J. A. (2012). The social life of the lion: a study of the behaviour of wild lions (Panthera leo massaica [Newmann]) in the Nairobi National Park, Kenya. Springer Science & Business Media.

[191] Burghardt, T., Calic, J., & Thomas, B. T. (2004, October). Tracking Animals in Wildlife Videos Using Face Detection. In EWIMT.

[192] Biometrics, P. (2014). Understanding biometric performance evaluation. URL: https://precisebiometrics. com/wp-content/uploads/2014/11/White-Paper-Understanding-Biometric-

Performance-Evaluation. pdf (pristupljeno: srpanj 2018.)[10].

[193] Alabort-i-Medina, J., & Zafeiriou, S. (2015). Unifying holistic and parts-based deformable model fitting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3679-3688).

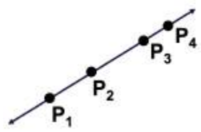
[194] Kumar, S., & Singh, S. K. (2017). Visual animal biometrics: survey. IET Biometrics, 6(3), 139-156.

[195] Bolme, D. S., Beveridge, J. R., Draper, B. A., & Lui, Y. M. (2010, June). Visual object tracking using adaptive correlation filters. In 2010 IEEE computer society conference on computer vision and pattern recognition (pp. 2544-2550). IEEE.

[196] Walker, K. A., Trites, A. W., Haulena, M., & Weary, D. M. (2011). A review of the effects of different marking and tagging techniques on marine mammals. Wildlife Research, 39(1), 15-30.

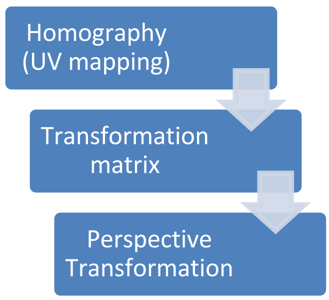
APPENDIX A

In section 5.4.1, it is mentioned about the technique of normalizing by perspectives by applying perspective transformations over all the images to match a certain perspective. The reason to use perspective transformations by keystone mapping techniques is due to the advantages of the 8-degrees of freedom for transformation and hence resulting in the information preserved. One such ratio thus preserved, is the cross ratio of collinear points, as shown below in the image of points on a straight line



$$\frac{\|\mathbf{P}_3 - \mathbf{P}_1\| \, \|\mathbf{P}_4 - \mathbf{P}_2\|}{\|\mathbf{P}_3 - \mathbf{P}_2\| \, \|\mathbf{P}_4 - \mathbf{P}_1\|}$$

The formation of perspective transformation can be seen as:



Therefore the values of the ratios to the same are preserved even after the perspective transformations are applied.

LIST OF PUBLICATIONS

GLEN BENNET HERMON

EDUCATION

2014-2016	M-Tech in Artificial Intelligence and Neural Networks, University of Petroleum & Energy Studies (UPES), Dehradun, India.
2010-2014	B.Tech in Information Technology, Karunya University, Coimbatore, India.

JOURNAL PUBLICATIONS

1. TITLE: "**Non-Invasive Techniques for Identification of Individuals within a Species: A Computational Review**"
   AUTHORS: ***Glen Bennet Hermon****, *Durgansh Sharma.*
   JOURNAL: ECOLOGY, ENVIRONMENT AND CONSERVATION
   VOLUME: Published in the 2021 May Supplement Issue; Page: S124–S138
   DIGITAL LINK:
   http://www.envirobiotechjournals.com/article_abstract.php?aid=11332&iid=329&jid=3

2. TITLE: "**Unique Lion Identification Using Triplet Loss and Siamese Networks**"
   AUTHORS: *Glen Bennet Hermon*, Durgansh Sharma*
   JOURNAL: Revue d'Intelligence Artificielle (RIA)
   VOLUME: Volume 34, Number 6, 2020; Page: 693-700
   DOI: https://doi.org/10.18280/ria.340603
   DIGITAL LINK:
   http://www.iieta.org/journals/ria/paper/10.18280/ria.340603

Thesis First Draft

ORIGINALITY REPORT

5%	4%	2%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.zooreach.org Internet Source	<1%
2	Santosh Kumar, Sanjay Kumar Singh. "Visual animal biometrics: survey", IET Biometrics, 2017 Publication	<1%
3	Santosh Kumar, Sanjay Kumar Singh, Rishav Singh, Amit Kumar Singh. "Animal Biometrics", Springer Science and Business Media LLC, 2017 Publication	<1%
4	www.bmva.org Internet Source	<1%
5	dokumen.pub Internet Source	<1%
6	digital-library.theiet.org Internet Source	<1%
7	d-nb.info Internet Source	<1%