Evaluating Drone Imagery for Wildlife Unique Feature Identification

Rozina Myoya[1[0000−0002−3154−1673]], Lesego Matojane[1[0009−0000−6431−6803]], Abiodun Modupe[1[0000−0002−9732−6466]], Vukosi Marivate[1[0000−0002−6731−6267]], and Albert Myburgh[2[0000−0002−6891−1893]]

[1] Department of Computer Science, University of Pretoria, Lynnwood Road, Pretoria 0002, South Africa
rozina.myoya@tuks.co.za,lmatojane@gmail.com,{abiodun.modupe,vukosi.marivate}@up.ac.za
[2] Department of Paraclinical Sciences, the Faculty of Veterinary Sciences & Departments of Zoology and Entomology, and Animals Sciences, University of Pretoria, Lynnwood Road, Pretoria 0002, South Africa
u12034542@tuks.co.za

Abstract. Emerging technologies provide novel ways of assessing ecological systems and species. Drones, also known as unmanned aerial vehicles (UAVs), have emerged as a revolutionary tool in conservation, providing new perspectives and the opportunity for new methods of monitoring elusive species. However, the potential of drones, particularly when combined with modern computational tools, remains largely untapped. In this study, we use a combination of drone technology, computer vision and machine learning approaches to recognise and re-identify individual crocodiles from aerial photos. First, the study focused on the construction and validation of a model capable of detecting crocodiles inside photos while also verifying its robustness for varied applications in real-world scenarios. Second, we investigated the potential of re-identifying individual crocodiles based on unique morphological traits, specifically using posture estimation methods and machine learning approaches such as PCA. The conclusion of these processes enabled us to build the framework for future population monitoring and individual re-identification in the wild. Our findings underlined the revolutionary potential of drone applications in animal conservation but also the necessity for multidisciplinary research to address operational and analytical issues, ensuring the effective protection of ecological diversity and the environment.

Keywords: Drone · CNN · Conservation · Image Recognition · YOLO · PCA · RoboFlow

1 Introduction

The conservation and management of endangered wildlife species, such as crocodiles, requires regular and accurate monitoring. Traditional ground-based methods often fall short, especially when dealing with species that are geographically dispersed or exhibit elusive behaviours. Unmanned aerial vehicles (UAVs), commonly referred to as drones, offer a promising solution, enabling rapid and extensive coverage of habitats whilst constraining the reliance on human observers.

While UAV-based wildlife monitoring has been explored globally, there is still a need for further research and development to optimise its use in varying environments. Additionally, the integration of advanced technologies such as photographic imaging with end-to-end neural networks is able to enhance the effectiveness of UAVs in detecting and tracking wildlife, ensuring more precise data collection for conservation efforts. This study is a pioneering effort in South Africa, utilising indigenous data. The challenges are multifaceted, such as the massive amount of data generated by drones, which necessitates sophisticated processing, and the inherent variability in natural landscapes, which can lead to misidentifications, such as confusing rocks with crocodiles. By utilising indigenous data, the study provides unique insights and perspectives that may not have been captured by previous studies conducted in other regions.

To overcome the challenges alluded to, we developed a computational framework that used machine learning and computer vision methods to analyse drone-captured imagery. Our proposed methodology could be used to detect crocodiles in images and provide basic population monitoring. In addition, we explored the feasibility of creating a system that could identify key crocodilian morphological features (e.g., the snout and front legs) that could facilitate individual re-identification and population health assessment.

The findings from this study can serve as a valuable reference for other countries facing similar conservation challenges, promoting international collaboration and knowledge exchange in wildlife conservation efforts. The study is comprised of 3 key elements:

1. The development of a segmentation process based on photographic imagery collected from UAVs in combination with ground truth data using RoboFlow annotation techniques. This allowed for accurate identification and mapping of different wildlife species and their habitats, aiding in targeted conservation efforts. It also provided a cost-effective and efficient method for monitoring and assessing the population dynamics of endangered wildlife species in South Africa, enabling timely interventions to protect their habitats.

2. Thereafter, With a metadata-rich, annotated dataset, we adopted the YOLO technique, a deep convolutional neural network (CNN) variation that uses bounding boxes to find and detect wildlife species (like crocodiles) in drone-derived images. This made it a flexible tool for monitoring biodiversity in South Africa.

3. Lastly, we used the pose estimation method [5] to investigate unique morphological features detected in aerial images, laying the groundwork for future wildlife monitoring applications.

2 Related work

Species identification through imagery has been a focal point of numerous studies. The process typically involves detecting the animal within an image and subsequently identifying individual animals based on unique morphological features. For instance, the study by [9] describes a deep convolutional neural network (CNN) approach that provided a completely automated pipeline for face detection, tracking, and recognition of wild chimpanzees from long-term video records. Using the detected faces, they created co-occurrence matrices to track changes in the social network structure of an ageing chimpanzee population. The author employed 10 million facial photos from 23 individuals over 50 hours of footage to get an overall accuracy of 92.5% in recognising a wild chimp and 96.2% for gender recognition. However, the exploitation of video data was exceedingly time-consuming, limiting the usability of these technologies at scale.

In the realm of crocodile identification, [3] used convolutional neural networks (CNNs) to identify individual free-roaming mugger crocodiles from drone-captured images, a problem that had persisted for years because data collected from individuals was collected under partially constrained conditions. Consequently, the authors proposed employing a CNN model to identify individual muggers based on their dorsal scute patterns. The CNN model was trained using 88,000 images (collected from 143 individuals residing in 19 distinct locations in western India). Two similar CNN models were utilised: one with an annotated bounding box approach (such as YOLO-v5l) and the other without annotations using Inception-v3. The true positive rate (TPR) and the true negative rate (TNR) were used to validate the performance of the proposed models, which attained 88.8 and 89.6 percent, respectively. At the moment, the implementation of this type of application is severely constrained because of the overwhelming majority of CNN models requiring substantial computing resources for training and performance evaluation. This same limitation was encountered in the study by [11], where the authors noted that CNN accuracy was limited (85%) due to the unavailability of training data for the CNN.

The overarching goal of these computer vision techniques was to bolster research and conservation efforts. For instance, a study on crocodile population modelling in the Olifants River Gorge done by [6] involved manually analysing drone imagery to determine population distribution. Given the time-intensive nature of manual analysis, the aim of this study was to automate this process, thereby streamlining data processing and analysis for conservationists. This type of research underscores the importance of individual-level identification in ecological studies by providing nuanced insights into animal behaviour in various contexts [8]. While biometrics for individual identification (IID) [2] has been extensively researched, studies like [3] have innovatively applied it to recognise unique features, such as the “scutes” on crocodiles.

Building on the foundational work discussed in this study, we collected images of Nile crocodiles (*Crocodylus niloticus*) in captivity using an unmanned aerial vehicle (UAV). We then trained the YOLO-v8s model using UAV imagery that had been segmented and annotated using RoboFlow techniques. The trained model showed high accuracy in identifying crocodiles, however automatic re-identification of individuals proved challenging due to the unavailability of enough annotated images to train the model to identify the key point on individuals using pose estimation. The subsequent sections delve into the details of the data description, the methodology followed, the results derived, and finally concluding remarks and recommendations for future work.

3 Data Description

Data was collected from Nile crocodiles (*Crocodylus niloticus*) in captivity. The data collection was facilitated by the DJI Mavic 3 Series drone. The drone had a 20 Mega-Pixel 1/2" CMOS sensor, 48 MP super-resolution photo and a 8x lossless zoom FHD video. The drone captured 30 frames per second, with a maximum flight time of 45 min with no wind, and a hover time of 38 min. The flight protocol involved taking pictures and videos of the crocodiles at an elevation range of 10 - 40 m. The data consisted on 7870 photographs in JPEG format captured on four different days during the months of January, February and September, which coincided with Summer and Spring in Limpopo province, South Africa. The photographs were captured throughout the day between 07h00 and 16h00. A summary of the photograph features are presented in Table 1.

Table 1. Dataset Metadata

No. Images:	7870	Resolution:	4056×3040
Data Set Size:	29.7 GB	Color Space:	RGB
Average Image Size:	7 MB	Format:	JPEG
Dates Recorded:	26/01/2022, 03/02/2022, 12/08/2022, 15/08/2022		
Time of day:	07h00 - 16h00 SAST		

The data also consisted of 1 hour and 7 minutes of video footage of the same crocodiles taken on 30/03/2023. A total of 4029 frames were extracted from 77 videos. A summary of the extracted photograph's features are presented in Table 2.

In addition, data from the study conducted by [3] on crocodiles photographed in India using a dron4.

3.1 Exploratory Data Analysis (EDA) on the Unlabeled Images

While EDA on unlabeled data may seem less intuitive than its labeled counterpart, its significance in the machine learning pipeline cannot be understated.

Table 2. Extracted Image Dataset Metadata

No. of images:	3593	Resolution:	1920 x 1080 (9%) 3840 x 2160 (33%) 5120 x 2700 (58%)
Dataset size:	71.8 GB	Format:	JPG
Colour Space:	RGB		
Date Recorded:	30/03/2023		

Image attributes such as quality, size, aspect ratio, and pixel intensity can profoundly influence model performance [7, 10]. Our EDA process invovled: (i) Data cleaning to remove corrupted files and ensure image readability; (ii) Image processing, including resizing images to 640 x 640 due to computational constraints on Google Co-Lab; (iii) K-means clustering to categorise images based on shared attributes; and lastly, (iv) Computation of image statistics to understand brightness and contrast distributions.

From the data collected, similarities were deduced from images falling within the same cluster. This gave a good indication of the different categories of images that existed within the dataset and also gave a good indication of the primary images where crocodiles were present. The average and variance of the pixel intensity gave an indication of the average brightness and contrast of the images, respectively [10]. When aiming to perform tasks such as object identification, the brightness and contrast features of the images were used to identify where the crocodiles may have been. Furthermore, it was found that considerable variability existed within the images, which was considered an advantage when conducting fine-grained tasks such as individual identification [4].

Lastly, to investigate the image quality, manual inspection was conducted by taking a sample of random photos and determining if distinct features could be identified by zooming into the crocodiles. From Figure 1 we can see that the distinct dorsal scute patterns were not clearly visible. This posed a challenge in the development of an individual re-identification model that we aimed to address using the method of pose estimation.

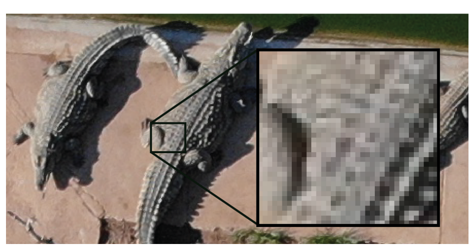

Fig. 1. Aerial image excerpt dataset depicting crocodiles along with a zoomed-in close-up

4 Methodology

Utilising the RoboFlow platform[1], we annotated images with instance segmentation, drawing bounding boxes around adult crocodiles larger than 1 m in length, as illustrated in Figure 2. Our research was then divided into two primary phases, namely, *Phase I: Crocodile Detection*, and *Phase II: Crocodile Re-Identification.*

Fig. 2. Annotated image of crocodile using smart polygon bounding box on RoboFlow

1. **Phase I: Crocodile Detection**: This phase aimed to develop a model proficient in detecting crocodiles within images. It was essential to ensure that the model was not only robust but also versatile enough to be applied in various scenarios beyond its initial training context. The pipeline for this phase is presented in Figure 3.
2. **Phase I I: Crocodile Re-Identification**: This phase delved into the application of pose estimation, where key points on crocodile individuals were annotated and the spatial differences between these points were calculated to determine if an individual could be identified through their unique morphological feature measurements. The pipeline for this phase is presented in Figure 4.

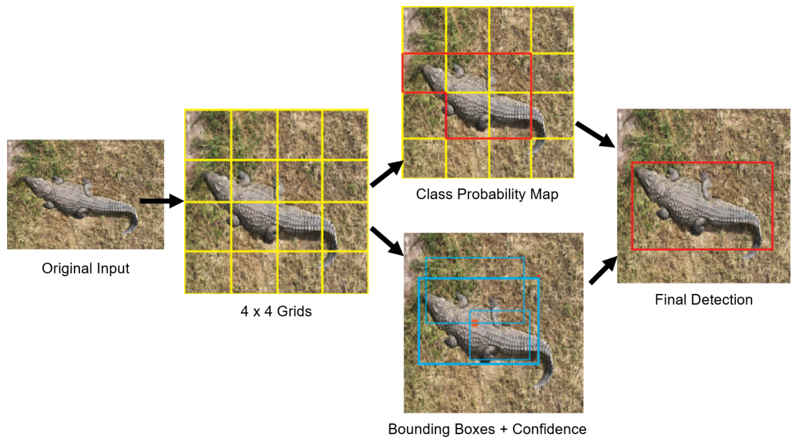

Fig. 3. Phase 1: Crocodile identification

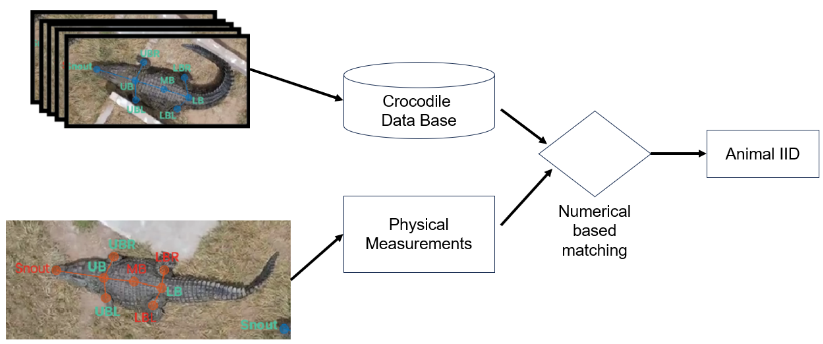

Fig. 4. Phase 2: Individual Identification Process Pipeline.

For model development, we employed the YOLO-v8s model, with its baseline hyperparameters detailed in Table 3. The NVIDIA Tesla T4 GPU was used, which is based on a Turing architecture and comes with 16 GB GDDR6 VRAM. This GPU is available on Google Co-Lab.

Model evaluation hinged on the Mean Average Precision (mAP) measure, commonly used in object detection. mAP is defined as:

$$mAP = \frac{1}{n} \sum_{k=1}^{n} AP_k \tag{1}$$

where, AP_k represents the average precision, i.e., the area under the Precision-Recall (P-R) curve for class k and n. Typically, mAP is computed based on the Intersection Over Union (IOU), which is defined as the overlap between the

Table 3. Baseline Model Hyperparameters

Hyperparameter	YOLO-v8s
Optimiser:	Stochastic Gradient Descent
Learning Rate:	0.01
Mini Batch size:	16
Epoch:	25
Input Image size:	640 x 640
Loss function:	Binary Cross-Entropy with Logits Loss

predicted and ground truth bounding box divided by the union of the predicted and ground truth bounding box (See Equation 2).

$$IOU = \frac{Overlap\ between\ the\ Predicted\ and\ Ground\ Truth\ BB}{Union\ of\ the\ Predicted\ and\ Ground\ Truth\ BB} \tag{2}$$

Furthermore, the model performance was evaluated using the True Positive Rate (TPR) and True negative Rate (TNR) metrics, which are defined by:

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

$$TNR = \frac{TN}{TN + FP} \tag{4}$$

5 Results

5.1 Phase I: Crocodile Detection

The primary aim of Phase I was to develop a robust model capable of detecting crocodiles within images, ensuring its applicability beyond the training context, and deploying it for real-world applications. The methodology for this phase involved:

1. Annotating images containing crocodiles using the RoboFlow platform to create a training dataset for the YOLO-v8s model.
2. We then trained the model using the annotated dataset with a train/val/test split of 80/10/10 and extracted the baseline results
3. Thereafter, we tested the model to measure the performance based on the True Positive Rate, True Negative Rate and False Positive rate to determine the model's reliability.
4. Finally, we increased the training dataset size by annotating more images and re-tested the model to determine if model performance could be improved based on the metrics measured in the previous step

Baseline Results The initial training of the YOLO-v8s model was conducted on 74 images, with validation and testing on 9 images each. The baseline results showcased a promising mean average precision (mAP) of 0.956 at an IOU of 0.5. However, the model's performance showed some degradation when the IOU ranged from 0.5 to 0.95, indicating areas for improvement in localization accuracy. The model's overall accuracy stood at 94%, as depicted in Figures 5 and 6.

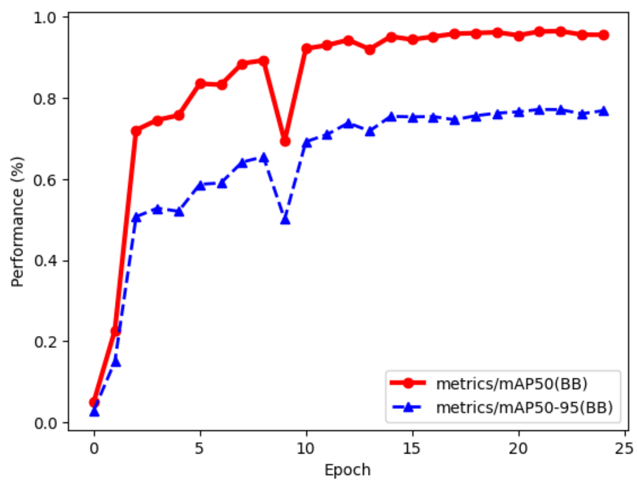

Fig. 5. mAP on training data with IOU threshold of 0.5 to 0.95

Model Testing The model was subjected to rigorous testing using three distinct datasets:

1. **Test set 1:** The test set was derived from using images that were not included in the training process, but that were taken in a similar context to the ones included in the training set, e.g., the pictures were taken at the same time of day as those in the training dataset. This was done to test the TPR of the model. In addition, we ensured that each image was derived from each of the clusters determined in the EDA process, while also taking into consideration images with the different pixel intensity scores.
2. **Test set 2:** This test set was focused on using images that were different from the training set and included features such as buildings or images that did not have crocodiles at all, to test the TNR of the model.
3. **Test set 3:** Finally, we tested the model on data completely different from our dataset. This dataset consisted of crocodiles photographed in India using a drone to determine if our model could be applied beyond our scope in real world scenarios to detect free roaming crocodiles [3].

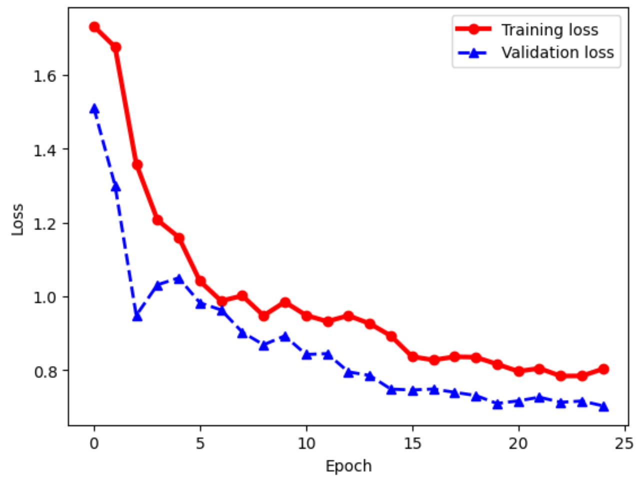

Fig. 6. YOLO-v8 training and validation loss.

The results, summarised in Table 4, revealed that while the model performed well on Test sets 1 and 2, its performance on Test set 3 was sub-optimal. This was evident from the model's tendency to misidentify rocks as crocodiles, as shown in Figures 7 and 8.

Table 4. Experimentation Results

	Test set 1	Test set 2	Test set 3
TPR	89%	43%	33%
TNR	47%	60%	14%
FPR	53%	40%	86%

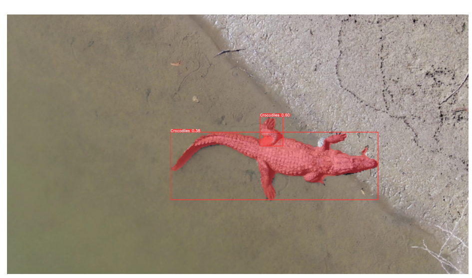

Fig. 7. Example of predicted images from Test set 3 showing false positives

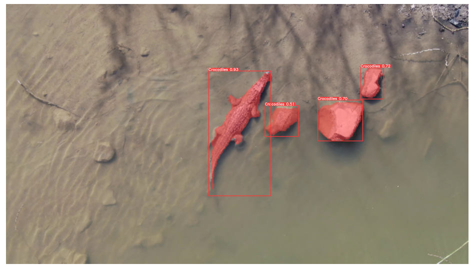

Fig. 8. Example of predicted image from Test set 3 the incorrect detection of a rock as a crocodile

Model Enhancement To address the identified shortcomings, we expanded our training dataset by incorporating images from Test sets 2 and 3. The new images sourced from test sets 2 and 3 were added to the original training dataset, with the strict stipulation that they only form part of the training set. However, any images from these two test sets that the model had not previously encountered were subsequently added to the test dataset. The augmented training dataset therafter consisted of 144 images, with the model achieving an accuracy of 91% post-training, with an increase in TPR performance by 3% for Test set 2 and 9% for Test set 3. However, any significant model performance improvement could be achieved using the following methods:

1. Increasing the diversity of images within our training dataset: This can be carried out by calculating the similarity within our entire dataset and choosing to annotate images that have the lowest level of similarity across the board, thus ensuring more variability within out dataset. Metrics such as the Structural Similarity Index Measure (SSIM) can be used to determine this score. Also, including images from Test set 3 in our training dataset would further increase this diversity. This would also eliminate redundancy within the images and reduce the bias in the model.
2. Modifications to the training dataset such as cropping, distorting and/or rotating the images may be necessary to create more variety in the training dataset.
3. We could also investigate the feature extraction process being carried by the model using tools such as Grad-CAM, which visualizes the CNN feature extraction process using a heat map of the RGB scores within the image to show the prominent features the CNN model detects. The results of this would provide guidance on what images would best be included in the training dataset, or inform our technique of image annotation.
4. Furthermore, experimentation with different learning rates, initial seed values, and alternating between hyper-parameter optimisation techniques to fine-tune the YOLO-v8 model could be explored in future studies to improve the model performance for the detection of crocodiles.

5.2 Phase II: Crocodile Re-Identification

Our proposed solution leveraged “Pose-estimation" methods to discern unique physiological features of crocodiles. Drawing inspiration from the research by [6], we aimed to use specific physiological measurements, such as the snout-neck length and snout-hind leg lengths, as unique identifiers for individual re-identification. We began our analysis with five candidate crocodile individuals. For each individual crocodile, multiple images were acquired, capturing the crocodile in various poses and assigning key points as shown in Figure 9.

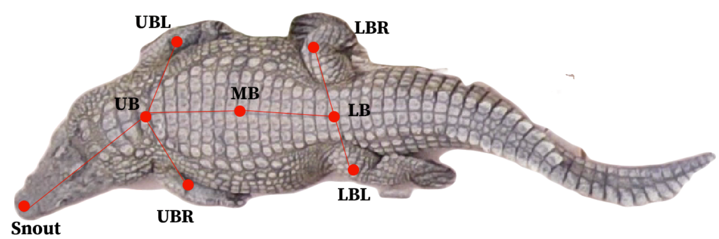

Fig. 9. Crocodile key point pose.

The subsequent steps taken in this phase included:

1. **Grouping Key points:** To ensure consistency in the analysis, key points were grouped based on the crocodile individual they represented.
2. **Interpolation of Key points:** We ensured a consistent number of key points across all crocodiles by interpolating key points in the case where some crocodiles were cut out of the image and only a partial part of the individual was visible.
3. **Computation of Distance Ratios:** We computed a pairwise distance matrix for each crocodile and formed a distance ratio matrix.
4. **Dimensionality Reduction and Visualisation:** PCA was applied to the crocodile feature vectors to visualise the differentiation between individual crocodiles. Thereby testing whether individual differentiation and thus identification was possible.

Computation of Distance Ratios

Pairwise Distance Computation: Given a set of key points K for a crocodile, where each key point k_i was represented as a coordinate pair (x_i, y_i), the pairwise distance between any two key points k_i and k_j was calculated using the Euclidean distance formula:

$$d(k_i, k_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{5}$$

This formula was applied to every pair of key points, resulting in a distance matrix D where the element $D_{i,j}$ represented the distance between key points k_i and k_j.

Distance Ratios: The aim of computing distance ratios was to capture the relative spatial relationships between the key points. For a given key point k_i, the ratio of its distance to another key point k_j relative to the sum of its distances to all key points was given by:

$$r(k_i, k_j) = \frac{d(k_i, k_j)}{\sum_{n=1}^{N} d(k_i, k_n)} \tag{6}$$

Where N was the total number of key points. This computation resulted in a distance ratio matrix R where the element $R_{i,j}$ was the distance ratio of k_i to k_j. This matrix captured the spatial relationships of key points independent of their absolute positions or scales.

Principal Component Analysis (PCA)

Mathematical Explanation: Given a data matrix X with zero mean (each feature has been centered around zero), the covariance matrix is:

$$\text{Cov}(X) = \frac{X^T X}{N - 1} \tag{7}$$

Where N is the number of data points. The next step was to compute the eigenvectors and eigenvalues of this covariance matrix. The eigenvectors represent the directions of maximum variance, and the eigenvalues signify the magnitude of the variance in these directions.

PCA transformed the original data based on these eigenvectors. The first principal component is the direction of the highest variance, the second principal component (orthogonal to the first) captures the highest variance that hasn't been captured by the first one, and so on.

Data Reduction: For visualisation, we retained the first three principal components. The transformed data was then visualised on a 2D plane, where each axis represented one of the two principal components. The results are presented in Figures 10.

PCA 1 vs PCA 2

Crocodile individuals CR_3 (triangle) and CR_5 (diamond) in Figure 10 are

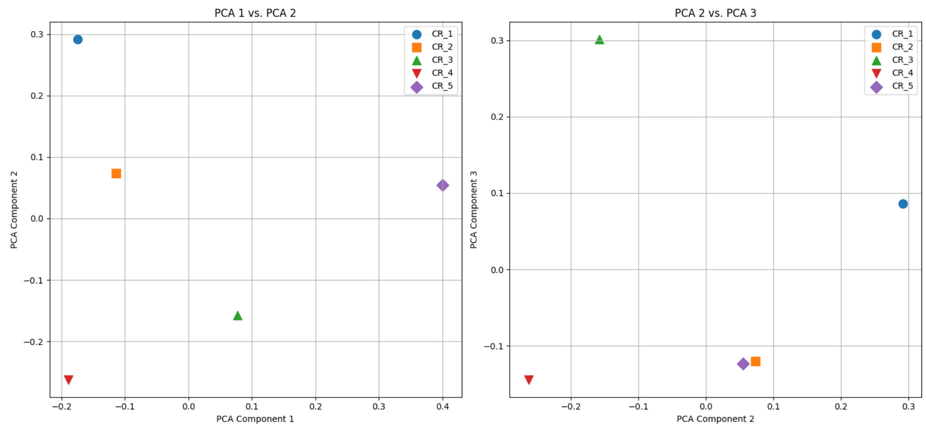

Fig. 10. PCA results showing the variation between individuals

distinctly separated from the other individuals in the PCA 1 direction, suggesting that their keypoint ratios have unique characteristics not shared by the other crocodiles, whilst CR_1 (circle), CR_2 (square) and CR_4 (inverted triangle) being more closely grouped together in the direction of highest variances (PCA 1). We do see good separation in the PCA 2 direction for CR_1, CR_2 and CR_4 in the PCA 2 direction.

PCA 2 vs PCA 3

CR_5 (circle) and CR_2 (square) are quite close in the PCA 2 direction, but we have already observed that they are easily separable in the PCA 1 direction. Overall the PCA visualisation indicated that the key points provide some differentiation between crocodile individuals. Certain crocodiles appeared closer in the feature space, indicating potential overlaps or similarities in their key point structures. The analysis suggested that while key points offer some ability to tell crocodiles apart, overlaps or similarities between certain individuals might exist, thus necessitating validation within the workflow.

Recommendations for Improvement The results from Phase II of our study, as visualised in Figure 10, underscored the potential of using key points to differentiate between individual crocodiles. However, the observed overlaps and similarities between certain individuals in the PCA visualisation highlight the need for a more refined approach.

Given the preliminary findings, we recommend the development of a dedicated machine-learning model that can accurately predict the key points on crocodiles identified in the images. This would not only enhance the precision of the identification process but also streamline the workflow outlined in Phase II.

Moreover, to ensure the reliability and robustness of the re-identification process, it is imperative to introduce a validation step. This step would involve cross-checking the predicted key points against a set of manually annotated key points to assess the accuracy of the predictions. Such a validation mechanism would provide insights into the model's performance and areas of improvement, ensuring that the re-identification process is both accurate and reliable.

In conclusion, while the current approach shows promise, there is ample room for improvement. By integrating a machine learning model for key point prediction and incorporating a validation step, we can pave the way for a more robust and reliable crocodile re-identification system, which could have significant implications for wildlife conservation and research.

6 Conclusions

The intersection of drone technology and Machine Learning has opened up a plethora of opportunities in the realm of wildlife research and conservation. This study stands as a testament to the potential of harnessing these technologies to discern unique features of wildlife, with a specific focus on crocodiles. Through the integration of the YOLO-v8s model for crocodile detection and the innovative pose estimation method for individual re-identification, we have laid the groundwork for a promising solution that can revolutionise wildlife monitoring.

Our findings from Phase I underscored the importance of data diversity in training datasets. While the model showcased potential in terms of True Positive Rate (TPR) and True Negative Rate (TNR) scores, it also highlighted areas of improvement. The introduction of more variety in the training data, coupled with strategies like data diversification, data augmentation and model fine-tuning, could significantly improve the model's reliability and robustness in real-world scenarios.

Phase II delved deeper into the intricate task of re-identifying individual crocodiles. We propose the approach of using key point estimation for individual identification. Traditional approaches to wildlife differentiation often lean heavily on high-resolution imagery, aiming to capture minute details to distinguish one subject from another. However, this method, while effective, often falters in real-world conservation efforts where high resolution imagery is not always available.

Looking ahead, there's immense potential in integrating the developed system with established software platforms like QGIS. Such integration would not only streamline the implementation process but also make the solution more accessible and user-friendly for conservationists and researchers.

Acknowledgments This research was conducted as a collaboration between the Computer Science Department, the Department of Para-clinical Sciences, the Faculty of Veterinary Sciences & Departments of Zoology and Entomology, and Animals Sciences at the University of Pretoria as part of the MIT in Big Data Science course. We would therefore like to express our deepest gratitude to the Department of Para-clinical Sciences, Faculty of Veterinary Sciences, Departments of Zoology and Entomology, and Animals Sciences for posing the research question that led to the development of this system. Their invaluable contribution to collecting and managing the data was instrumental to the success of this project.

Bibliography

[1] F. Ciaglia, F. S. Zuppichini, P. Guerrie, M. McQuade, and J. Solawetz. Roboflow 100: A rich, multi-domain object detection benchmark. *arXiv preprint arXiv:2211.13523*, 2022.

[2] S. Dargan and M. Kumar. A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. *Expert Systems with Applications*, 143:113114, 2020.

[3] B. Desai, A. Patel, V. Patel, S. Shah, M. S. Raval, and R. Ghosal. Identification of free-ranging mugger crocodiles by applying deep learning methods on uav imagery. *Ecological Informatics*, 72:101874, 2022.

[4] G. V. Horn, O. M. Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie. The inaturalist species classification and detection dataset, 2018.

[5] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018.

[6] H. A. Myburgh. *Aspects of monitoring wild and captive Nile crocodile (Crocodylus niloticus) populations in southern Africa.* PhD thesis, 2021.

[7] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1): 1, Dec. 2015. ISSN 2196-1115. `https://doi.org/10.1186/s40537-014-0007-7`. URL `https://journalofbigdata.springeropen.com/articles/10.1186/s40537-014-0007-7`.

[8] A. Z. Perez, C. Bone, and G. Stenhouse. Simulating multi-scale movement decision-making and learning in a large carnivore using agent-based modelling. *Ecological Modelling*, 452:109568, 2021.

[9] D. Schofield, A. Nagrani, A. Zisserman, M. Hayashi, T. Matsuzawa, D. Biro, and S. Carvalho. Chimpanzee face recognition from videos in the wild using deep learning. *Science advances*, 5(9):eaaw0736, 2019.

[10] L. Wang, C. Leckie, K. Ramamohanarao, and J. Bezdek. Automatically Determining the Number of Clusters in Unlabeled Data Sets. *IEEE Transactions on Knowledge and Data Engineering*, 21(3):335–350, Mar. 2009. ISSN 1041-4347. `https://doi.org/10.1109/TKDE.2008.158`. URL `http://ieeexplore.ieee.org/document/4759628/`.

[11] A. B. Woolcock, S. Cotton, and A. J. Cotton. Effectiveness of using drones and convolutional neural networks to monitor aquatic megafauna. *African Journal of Ecology*, 60(3):544–556, 2022.