Imperial College London

Department of Earth Science and Engineering

MSc in Applied Computational Science and Engineering

Independent Research Project
Project Plan

Evaluating Deep Learning in Leopard Individual Identification: A Novel Approach Using Triplet Networks

by

David Colomer Matachana

Email: dc1823@ic.ac.uk
GitHub username: acse-dc1823
Repository: https://github.com/ese-msc-2023/irp-dc1823

Supervisors:

Dr Lluis Guasch
Dr. Sanjay Gubbi

June 14, 2024

Abstract

Accurately identifying individual leopards across multiple camera trap images is critical for population monitoring and ecological studies. This paper introduces a novel Deep Learning framework using Triplet Networks to distinguish between individual leopards based on their unique spot patterns. The objective is to evaluate whether this framework can effectively handle the complexity of animal pattern recognition. Unlike conventional methods that primarily rely on explicit feature extraction methods, this approach uses a Triplet Convolutional Neural Network architecture to implicitly learn discriminative features from leopard coat patterns. This approach has seen extensive success in facial recognition, demonstrating its potential for other open-set learning examples.

This research not only contributes to the Computer Vision field but also offers a powerful tool for biologists aiming to study and protect leopard populations more effectively. It serves as a stepping stone for applying the power of deep learning in Capture-Recapture studies for other patterned species.

1 Introduction

Possibly the most critical task in conservation consists of being able to identify individual animals over time. Through Capture-Recapture techniques, we can evaluate their fitness and track their demographics (Wearn et al., 2017). For example, in the case of leopards, monitoring their populations in an ever-growing urban landscape in India is the focus of intensive research (Gubbi et al., 2020), (Gubbi et al., 2021). Traditionally, this process has been carried out using invasive, telemetry-based methods; however, these techniques raise survival and stress concerns (McMahon et al., 2005), which pose significant issues when studying endangered species. With the advent of automatic camera traps, Photographic-Capture-Recapture (PCR) has gained popularity among researchers. This method relies on clear, distinct markings on the animals to facilitate individual identification. In earlier applications, the pattern matching was done visually by researchers (Karanth et al., 1998). This approach becomes extremely cumbersome for large datasets with thousands of images due to the need to attempt to match every pattern with every identified individual in the dataset. Moreover, it can lead to identification errors, notably the duplication of individuals, as it was seen in Verschueren et al., 2022 that humans identified up to 22% more cheetah individuals than the actual population.

Given these limitations, it immediately became obvious that there was a need for semi-automatic pattern matching through Computer Vision (CV) algorithms. Today, we have numerous such programs. The state-of-the-art programs include Hotspotter (Crall et al., 2013) and Wild-ID (Bolger et al., 2012). These algorithms are based on Scale Invariant Feature Transform (SIFT) (Lowe, 2004) for explicit pattern extraction, and then use various distance-based or Machine Learning pattern matching algorithms. These algorithms offer varying results according to the species. We have found no studies on our target species, although, in a species with similar rosettes, the Jaguar, the accuracy reported is over 70% (Nipko et al., 2020). In other patterned species, accuracy reported ranges from 36% (Morrison et al., 2016) to close to 100% (Burgstaller et al., 2021). Additionally, these systems require manual intervention to extract a Region of Interest (ROI) where the animal is located. This manual ROI extraction is not only time-consuming for large datasets but also introduces a layer of subjectivity and potential inconsistency.

Despite the efficiency of these algorithms, they were developed before recent advances in Deep Learning and may not harness the full potential of current technological capabilities. These

methods can potentially achieve higher accuracy in individual animal identification due to their ability to learn complex patterns and features from large datasets implicitly. There have been some recent attempts to apply Deep Learning to this problem, notably to elephants (Körschens et al., 2018), Giant Pandas (Hou et al., 2020), and even our own target species, leopards (Pucci et al., 2020). While they report high accuracies in re-identification, they have all built a traditional closed-set classifier, which means that their applicability outside of the training population is very limited. For example, Hou et al., 2020 report that their accuracy on a different population to the training descend from 95% to 21%.

This paper thus attempts to build an end-to-end Deep Learning solution to the PCR problem on an open set, with a focus on leopards. We will evaluate whether Deep Learning for implicit pattern extraction of leopard rosettes is a valid method, and compare to existing SIFT implementations in terms of accuracy, efficiency, and scalability.

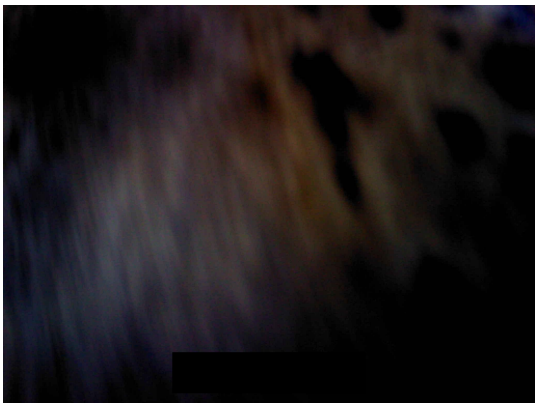
2 Methodology

2.1 Data Collection

Good and diverse data is paramount to building a good model. There has been extensive research showing that Neural Networks need a large dataset to learn features robustly (Gütter et al., 2022). In our case, this becomes even more crucial, as fixed camera trap images of leopards are extremely variable in pose, lighting, etc.

To train the model, the Nature Conservation Foundation have provided us with 8900 tagged images of 600+ individual leopards. While the majority of images are of high quality, all the images were inspected manually, of which 612 were removed given the impossibility of identifying an individual:

(a) Optimal image



(b) Unusable image

Figure 1: Comparison of optimal and unusable images

The remaining images have a skewed distribution in the number of images per leopard flank, with a mean of just 6.4 images per flank:

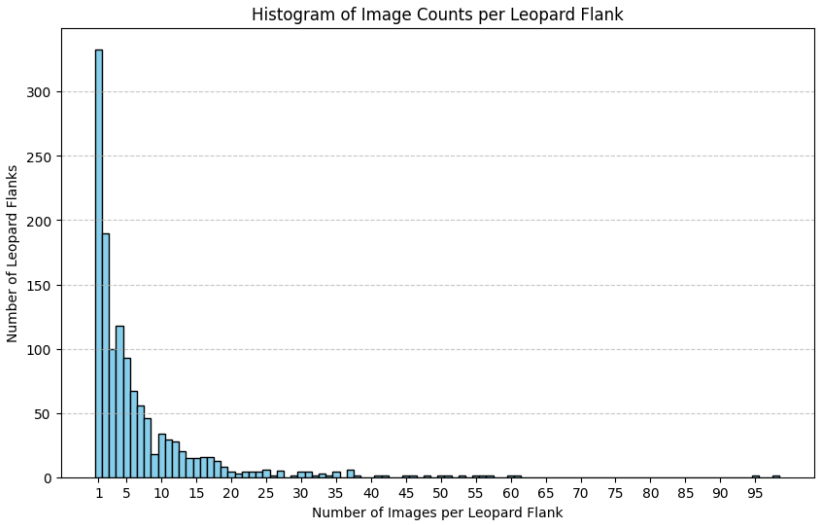


Figure 2: Histogram of the number of images per flank.

2.2 Preprocessing

Differing from Hotspotter and Wild-ID, in which the user has to manually identify the ROI, we'll attempt to use current advances in CV for automatic detection extraction of bounding boxes, using a pre-trained YOLO network for this (Redmon et al., 2015).

Furthermore, as seen in de Lorm et al., 2023, filtering out the background results in a performance increase in previous re-identification algorithms. We will thus use further existing CV algorithms for this, such as "rembg" (Gatis, 2020).

Edge-detection techniques as a preprocessing step have been associated with a performance increase for other classifiers (Chengeta et al., 2019). Hence, we can use Edge-detection to homogenize lighting condition variations across images and isolate the patterns that we want the model to focus on.

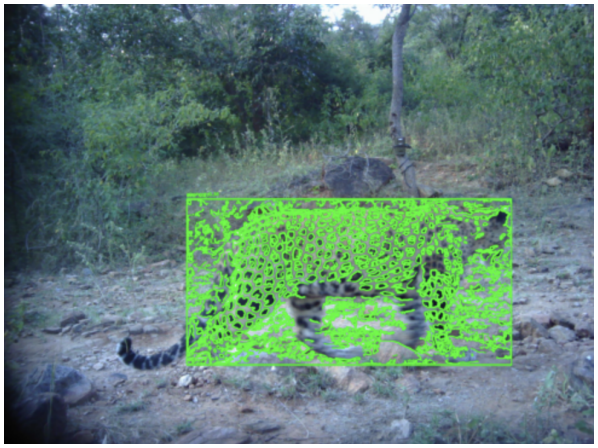


Figure 3: Canny-Edge-Detection applied to a YOLO bounding box around a Leopard. Background not removed

2.3 Triplet Convolutional Neural Network

With sufficiently preprocessed training data, we can feed it into our network. Open-set identification tasks, in which the "classes" seen in the test and application phases are not the same as those seen in training pose significant issues. Training traditional classifiers for this task is unsuitable, as the model learns to discriminate solely between classes seen in training and cannot extrapolate to new individuals (Hou et al., 2020).

Given this, we can take inspiration from the parallels of this task with facial recognition models, and design a Triplet Network (Schroff et al., 2015, Sankaranarayanan et al., 2016). In such an architecture, instead of outputting a class label, the network learns a feature space where embeddings of images are directly compared. The network is designed to process three images at a time: an anchor image, a positive image (i.e., another image of the same leopard), and a negative image (i.e., an image of a different leopard).

The Triplet CNN consists of three identical sub-networks sharing the same weights. Each sub-network outputs an embedding vector for its input image, and the similarity between these vectors is measured using a distance function.

The training objective is to minimize the distance between the anchor and the positive while maximizing the distance between the anchor and the negative through the triplet loss (Schroff et al., 2015):

$$L = \max(d(a, p) - d(a, n) + \alpha, 0)$$

where $d(x, y)$ is the distance between the embeddings of images x and y, a, p, and n represent the anchor, positive, and negative images, respectively, and α is a hyperparameter that defines how much the negative example should be farther away from the anchor compared to the positive. Other relevant losses that have succeeded in facial recognition will also be examined (Khan et al., 2024).

By not only learning to minimize the distance of intra-class instances but also maximizing the distance of inter-class instances, this architecture is particularly suited for open-set discriminative learning of similar classes. Other recent architectures such as CLIP are gaining traction in open-set learning, but it seems that Triplet Networks are uncontested winners in discriminating between similar classes (Radford et al., 2021, Bhat et al., 2023).

Considering our limited dataset size, we intend to implement transfer learning using pre-trained networks designed for facial recognition, known for their robust discriminative capabilities of highly similar classes (Liu et al., 2017). These networks, primarily optimized for human facial features, may encounter generalization challenges when adapted to the completely differing leopard rosettes. Thus, a detailed empirical evaluation will have to be carried out to find the best architecture.

2.4 Evaluation and Application

2.4.1 Validation set loss

To evaluate the model, we apply the loss to the validation set of unseen leopards, differing from previous attempts at classifying individual animals (Hou et al., 2020, Körschens et al., 2018).

With this, we can ensure that the network is learning to discriminate between individuals in general, and not just learning the representations of the leopard coats seen in training.

2.4.2 Application to new leopards

Firstly, during the application phase, each image is processed through one branch of the Triplet Network to generate the rich embeddings.

Given the open-set nature of the problem, whenever we apply the model to new data, we will have to have a decision-making algorithm that, given the patterns extracted through the network, matches the most similar ones together. Given that the one-vs-many approach will be inefficient for large datasets, using nearest-neighbours algorithms will be used here. At this point, we will use common metrics in the field such as top-rank accuracy (Nipko et al., 2020, Grabham et al., 2024) to compare our implementation with that of Hotspotter.

3 Expected Deliverable

The expected deliverable will be an end-to-end Deep Learning model able to classify any given dataset of leopard images into individuals. It will overcome current limitations of Hotspotter and Wild-ID, such as their manual ROI selection. However, the main objective of the project is to explore whether Deep Learning can be used as an effective solution to leopard individual identification, comparing the attained accuracy against SIFT-based Hotspotter.

In facial recognition implementations, CNN's are seen to struggle with pose variations (Singh et al., 2018). We postulate that this will be the major challenge to overcome for our model, as fixed camera traps result in completely varying leopard poses:

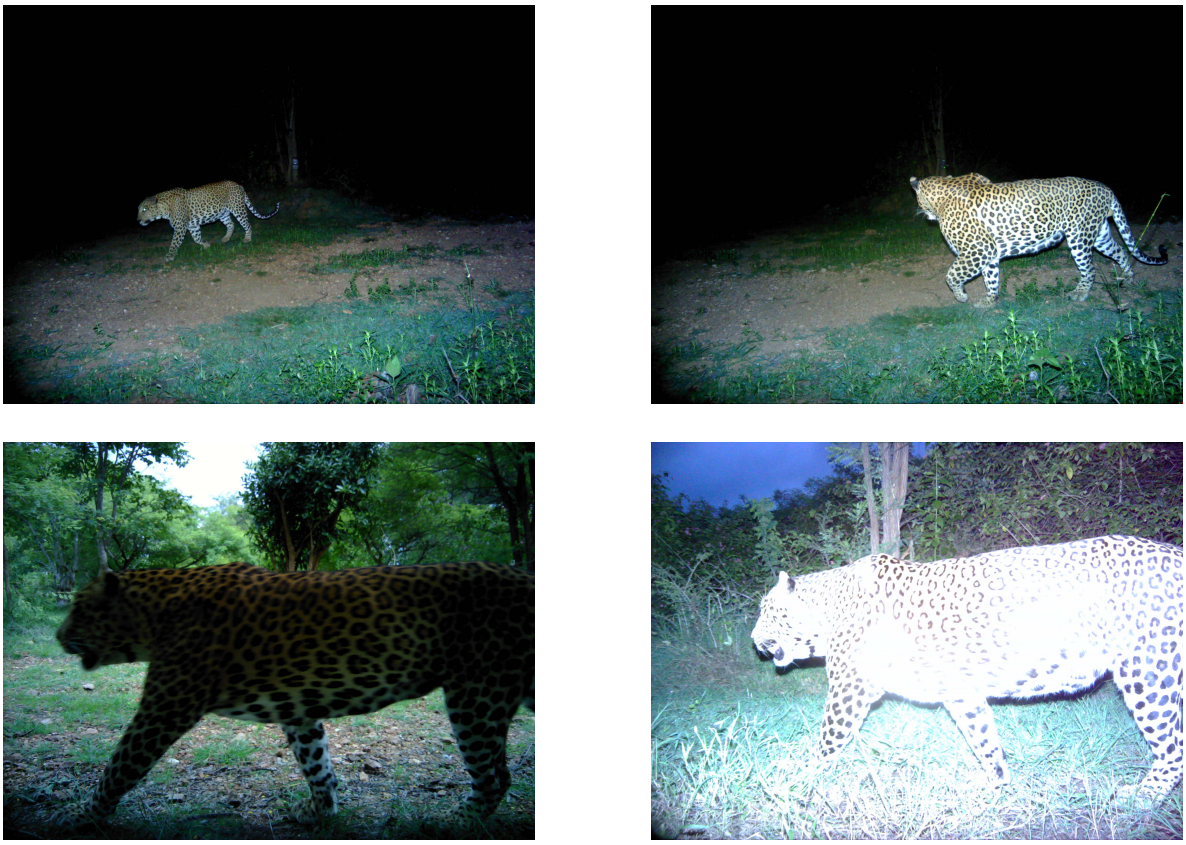


Figure 4: Comparison of 4 images of the same leopard among widely different poses, distances from the camera, and lighting conditions

4 Future Plan

The Gantt chart below outlines the plan for the whole project. One will notice that the work currently done is extensive, yet intangible.

Figure 5: Gantt Chart of work to be done

References

Bhat, A., et al. (2023). Face recognition in the age of clip billion image datasets. *ArXiv*. https://doi.org/10.48550/arXiv.2301.07315

Bolger, D. T., et al. (2012). A computer-assisted system for photographic mark-recapture analysis. *Methods in Ecology and Evolution*. https://doi.org/10.1111/j.2041-210X.2012.00212.x

Burgstaller et al. (2021). The green toad example: A comparison of pattern recognition software. *North-Western Journal of Zoology*, *17*, e211506.

Chengeta, K., et al. (2019). *Image preprocessing techniques for facial expression recognition with canny and kirsch edge detectors* (N. T. Nguyen, R. Chbeir, E. Exposito, P. Aniorté, & B. Trawiński, Eds.). Springer International Publishing.

Crall, J. P., et al. (2013). Hotspotter - patterned species instance recognition. *IEEE*. https://ieeexplore.ieee.org/document/6475023

de Lorm, T. A., et al. (2023). Optimizing the automated recognition of individual animals to support population monitoring. *Ecology and Evolution*. https://doi.org/10.1002/ece3.10260

Gatis, D. (2020). rembg: A tool to remove background from images and videos.

Grabham, A. A., et al. (2024). Evaluating the performance of semiautomated photographic identification programs for leopard seals. *Wildlife Society Bulletin*. https://doi.org/10.1002/wsb.1520

Gubbi, S., et al. (2020). Every hill has its leopard: Patterns of space use by leopards (Panthera pardus) in a mixed use landscape in india. *PeerJ*, *8*, e10072. https://doi.org/10.7717/peerj.10072

Gubbi, S., et al. (2021). Variation in leopard density and abundance: Multi-year study in cauvery wildlife sanctuary. *Nature Conservation Foundation*. https://www.ncf-india.org/western-ghats/variation-in-leopard-density-and-abundance-multi-year-study-in-cauvery-wildlife-sanctuary

Gütter, J., et al. (2022). Impact of training set size on the ability of deep neural networks to deal with omission noise. *Frontiers in Remote Sensing*, *3*. https://doi.org/10.3389/frsen.2022.932431

Hou, J., et al. (2020). Identification of animal individuals using deep learning: A case study of giant panda. *Biological Conservation*. https://doi.org/10.1016/j.biocon.2020.108414

Karanth, K. U., et al. (1998). Estimation of tiger densities in india using photographic captures and recaptures. *Ecology*, *79*(12), 2852–2862. https://doi.org/10.1890/0012-9658(1998)079[2852:EOTDII]2.0.CO;2

Khan, Z., et al. (2024). Improvised contrastive loss for improved face recognition in open-set nature. *Pattern Recognition letters*. https://doi.org/10.1016/j.patrec.2024.03.004

Körschens, M., et al. (2018). Towards automatic identification of elephants in the wild. *ArXiv*. https://doi.org/10.48550/arXiv.1812.04418

Liu, W., et al. (2017). Sphereface: Deep hypersphere embedding for face recognition. *ArXiv*. https://doi.org/10.48550/arXiv.1704.08063

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

McMahon, C., et al. (2005). Handling intensity and the short- and long-term survival of elephant seals: Addressing and quantifying research effects on wild animals. *AMBIO: A Journal of the Human Environment*, *34*(6), 426–429. https://doi.org/10.1579/0044-7447-34.6.426

Morrison, T. A., et al. (2016). Individual identification of the endangered wyoming toad Anaxyrus baxteri and implications for monitoring species recovery. *Journal of Herpetology*, *50*(1), 44–49. https://doi.org/10.1670/14-155

Nipko, R. B., et al. (2020). Identifying individual jaguars and ocelots via pattern-recognition software: Comparing hotspotter and wild-id. *Wildlife Society Bulletin*. https://doi.org/10.1002/wsb.1086

Pucci, R., et al. (2020). Whoami: An automatic tool for visual recognition of tiger and leopard individuals in the wild. *ArXiv*. https://doi.org/10.48550/arXiv.2006.09962

Radford, A., et al. (2021). Learning transferable visual models from natural language supervision. *Proceedings of Machine Learning Research*. https://doi.org/10.48550/arXiv.2103.00020

Redmon, J., et al. (2015). You only look once: Unified, real-time object detection. *CoRR*, *abs/1506.02640*. http://arxiv.org/abs/1506.02640

Sankaranarayanan, S., et al. (2016). Triplet similarity embedding for face verification. *ArXiv*. https://doi.org/10.48550/arXiv.1602.03418

Schroff, F., et al. (2015). Facenet: A unified embedding for face recognition and clustering, 815–823. https://doi.org/10.1109/CVPR.2015.7298682

Singh, S., et al. (2018). Techniques and challenges of face recognition: A critical review [8th International Conference on Advances in Computing Communications (ICACC-2018)]. *Procedia Computer Science*, *143*, 536–543. https://doi.org/https://doi.org/10.1016/j.procs.2018.10.427

Verschueren, S., et al. (2022). Reducing identification errors of african carnivores from photographs through computer-assisted workflow. *Mammal Research*. https://doi.org/10.1007/s13364-022-00657-z

Wearn et al. (2017). Camera-trapping for conservation: A guide to best-practices. https://doi.org/10.13140/RG.2.2.23409.17767