Elephace: Elephant Re-identification Using Deep Learning

First Author
Institution1
Institution1 address
firstauthor@i1.org

Second Author
Institution2
First line of institution2 address
secondauthor@i2.org

Abstract

The goal of our project is to build vision model to re-identify elephants using their images. This model will be used by researchers to track elephant movement pattern in zoos and sanctuaries. This non-invasive way of recognizing elephants is better than capturing and attaching devices to them which requires human intervention. Manually identifying elephants from images is a time-consuming process and a good automated model is essential. We have built an elephant re-identification model using three deep learning methods and performed experiments to identify the best parameters for the model. Our primary dataset (aka zoo dataset) provided by Prof. Chusyd. We have built a classification model to classify an image into one of the known elephant categories. With ResNet50 as the backbone and our primary dataset, this model's top-1 accuracy was 99.71%. We aimed to recognize elephants with as few images as possible in our next approach, therefore we constructed a few shot learning model. Using just 3 images per class we obtained top-1 accuracy of 81% and top-3 accuracy of 92% on our primary dataset. The performance of few shot on more complex wild elephants dataset was not as good as zoo elephants so we moved to training Siamese network using triple loss (ala FaceNet). With the best hyperparameter choices, we got validation rate of 0.585 at false acceptance rate of 0.01.

1. Introduction

Re-identification of animals is necessary for biodiversity monitoring and ecological research projects. Biologists and anthropologists need animal tracking to monitor their health, behavior, group dynamics and variation of population over time [Deb, Debayan, et al.]. Biodiversity development and health information is highly valuable for assessing the environmental effect and necessary steps required to preserve the ecosystems.

Animal tracking can be performed by using a device that is attached to the animal. This technique requires human intervention and it is intrusive. It requires capturing animals which can be dangerous and also disruptive for an animal's wild habitat [Deb, Debayan, et al.]. Another method of tracking animals is by capturing the images and re-identifying animals from those images. The re-identification performed manually can be a time consuming process as multiple and vast numbers of images need to be analyzed. This re-identification requires domain knowledge about the particular species and also its prone to biases related to human judgment [Schneider, Stefan, et al.].

Automated re-identification of animals provides a better solution to animal tracking without involving disruptive techniques and human errors. There has been ongoing research to extend human face recognition and identification techniques to animal identification using deep learning. There are many challenges while building a state of the art animal re-identification and classification model. These challenges arise due to similar looking animals and very small distinctive features like body size, scars and marks, coloring, etc [Körschens, et al.]. Apart from this, these features could also change over time. For example in the case of elephants, it could lose a tusk and a hole in the ear may become a rip [Körschens, et al.]. Other challenges are variations in image captured such as occlusion, varying viewpoints, different poses and angles of the animal in an image that affect the detection of distinctive features. For example in the case of elephants the feature might not be clearly visible in images because of mud on their bodies or the angle and movement [Körschens, et al.].

In this project, we have build a model for elephant re-identification system using bounding box detection architecture for bounding box predictions followed by multiple deep learning network architectures for elephant re-identification.

2. Background and related work

In the context of animal re-identification, various works have used custom feature engineering and classical computer vision techniques. Recently there has been some work on using the advanced and already mature human face

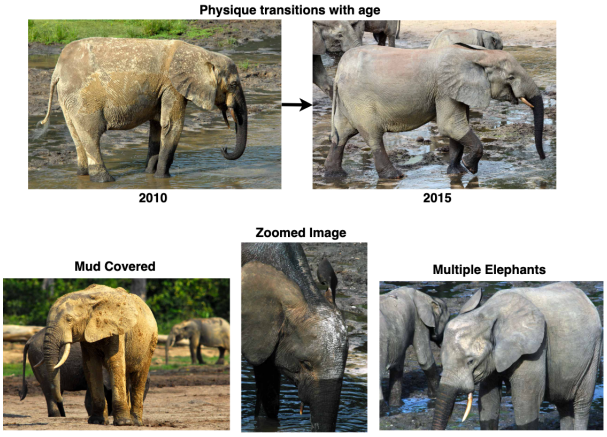

Figure 1. Clockwise: Elephant Physique transitions with age, Elephant covered in mud, zoomed image, multiple animals in an image

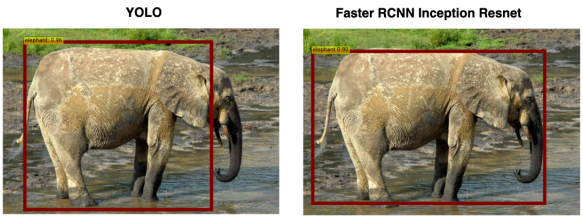

Figure 2. Bounding Box Detected by Yolov4 and Faster RCNN Inception ResNet V2

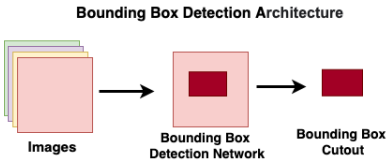

Figure 3. Elephant Bounding Box Detection

recognition techniques e.g. FaceNet, DeepFace for animal recognition and identification tasks.

[Korschens et al] used a pre-trained YOLO network for bounding box predictions to automatically locate an elephant's heads in images using pre pre-trained YOLO network.Used a human in the loop approach to select the best bounding box and then the bounding box is cropped and fed into a pre-trained ResNet 50 for feature extraction from the middle layer followed by pooling. PCA is used to reduce the number of features and then SVM is used for classification. To improve the accuracy, results of multiple images of the same elephant are aggregated by averaging the class wise confidence. The Top 1 accuracy for classification is 56% and 74% with two image aggregation. For the same task but using traditional methods [Ardovini et al][Schneider review] - performed re-identification of elephants using multi-curve matching technique and achieved top-1 accuracy for 75%.

For primate re-identification [Debayan Deb et al] implemented a variation of a sphereface deep recognition to build embeddings of faces which then can be compared using similarity functions. Instead of using an object detection network to identify the faces, they manually annotated landmarks to align the faces. They achieved a closed set accuracy of 75.82% on the chimpanzee dataset.

For identification of giant pandas, [Jin Hou et al] used significant data augmentation and trained a VGGnet for classification tasks. They were able to accurately identify 90% of panda individuals.

For the difficult task of Nyala identification, [Zyl et al] implemented an end to end system consisting of Faster R-CNN model for bounding box detection and then siamese network using ResNet-152 backbone for image embedding. Their top1 accuracy for zebra dataset is 74.1% and top-10 accuracy of 85% for Nyala animal identification.

[Freytag et al] uses matrix logarithm transformation on bilinear pooling to increase the discriminability of features in Alexnet and VGGfaces networks. Using these techniques, they achieved the state of the art for chimpanzee faces classification. For a similar task, Gorilla faces classification [Brust et al] uses YOLO and R-CNN for bounding box detention along with AlexNet for feature extraction upon which they used SVM for class label prediction.

3. Methods

3.1. Elephant Detection and Extraction

In order to accurately identify elephant in an image, it is essential to isolate it from surrounding which might contain other elephants, animals or people. For this reason, a pre-trained object detection model is used to obtain bounding boxes for all the elephants in the image. If there are multiple elephants detected, then the elephant with the bounding box of largest area is considered to be the subject and extracted.

Yolov4 [**?**], Faster R-CNN Inception ResNet V2 1024x1024 [**?**], and SSD MobileNet v2 320x320 [**?**] pre-trained models were tried to detect and get bounding boxes for the elephants. The Yolov4 pretrained model is obtained from [**?**] and Faster R-CNN Inception ResNet V2 1024x1024 and SSD MobileNet v2 320x320 are obtained from Tensorflow Object detection zoo [**?**]. It is found that for these datasets, bounding boxes obtained from Yolov4 often times did not include the head of the elephant see fig:**??**. For this reason, Faster R-CNN and SSD MobileNet are used to extract the elephant for the next phase.

To re-identify an elephant given some of its images we consider three approaches - classification, few-shot recognition, and Siamese network for similarity measurements.

3.2. Classification

When the problem of re-identification is posed as a classification, the model is trained with labelled images of elephants. Both the train and validation sets will consist of disjoint images of the same elephant and the model has to learn the discriminative features and use them to predict the label of elephant images from the validation set. Our classification model consists of a pre-trained CNN as a backbone to provide the image embeddings and a couple of fully connected layers with softmax to classify the images into one of the elephants from the train set. Cross-entropy loss along with L2 regularization for the fully connected layers is used for training the model.

3.3. Few Shot Learning

When there are only a few images available per elephant, the train set used for classification becomes imbalanced and learning becomes difficult. To solve this problem few shot learning paradigm is utilized, where image embeddings from a pre-trained CNN of known images are compared with the embeddings of unknown images using a suitable similarity metric [**?**].

Using a pre-trained CNN, the image embeddings are obtained.

Support set, query set

query set image embedding, dot product, with mean embedding of the support set

normalized embeddings

softmax over the dot product results of all comparisons

finetune - initial value of W - entropy regularization - cosine similarity instead of dot product

$$w_{S_k} = \sum_{s \in S_k} \frac{\|f(x_s)\|}{|S|}$$

$$z_{S_k}^{(q)} = \frac{w_{S_k}^T \|f(x_q)\|}{\|w_{S_k}\|} + b_{S_k}$$

$$\hat{y}^{(q)} = softmax(z^{(q)})$$

$$\mathcal{L} = \sum_j \sum_i (y_i^{(j)} log(\hat{y}_i^{(j)}) + \hat{y}_i^{(j)} log(\hat{y}_i^{(j)}))$$

where $y_i^{(j)}$ and $\hat{y}_i^{(j)}$ is the i^{th} component of the true label and predicted labels of j^{th} example respectively.

The ideas above are from [**?**], [**?**], [**?**]

3.4. Siamese Network

Facenet paper

triplet loss

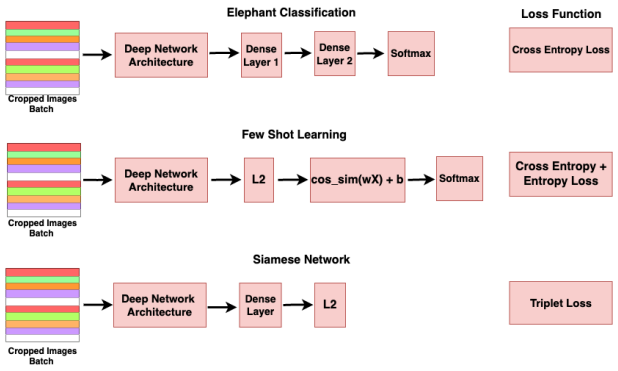

Figure 4. Elephant re-identification architectures

$$\mathcal{L} = \sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]$$

batchall

batch hard

hard positive: $x_i^P = \text{argmax}_{x_i^p} \|f(x_i^a) - f(x_i^p)\|_2^2$

hard negative: $x_i^N = \text{argmin}_{x_i^n} \|f(x_i^a) - f(x_i^n)\|_2^2$

batch partial hard

4. Experiments and Results

Number of experiments were performed on all three model. The Table1 shows parameters tried for each of the models.

Classification	**Few Shot**	**Siamese**
RestNet50, InceptionV3	RestNet50, InceptionV3	RestNet50, InceptionV3
Image Augmentation	Image Augmentation	Image Augmentation
Finetuning	Finetuning	Embedding size
	Support set size	Triplet loss margin
		Triplet loss strategy
		Optimizer: Adam, SGD
		Aspect Ratio

Table 1. Model Experiments

4.1. Elephant Classification

For elephant classification model, we used two backbone models of ResNet50 and InceptionV3. We found that both the model performed well on zoo elephants dataset with top-1 accuracies of 0.9971 and 0.979 respectively. This model did not perform well on wild dataset with top-1 accuracies

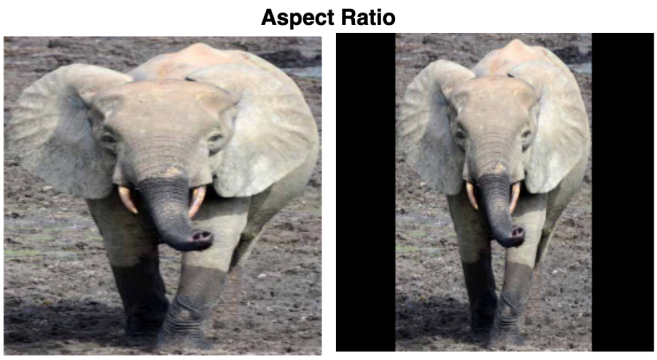

Figure 5. Image Padding: To preserve Aspect Ratio

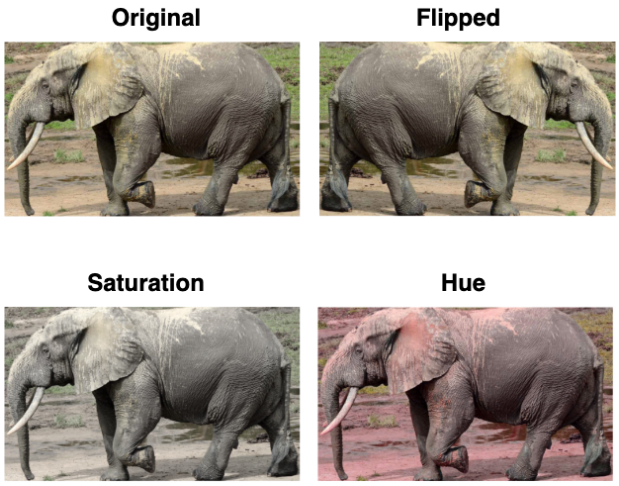

Figure 6. Image augmentations performed are right-left flip, saturation and hue changes

of 0.21 and 0.16 respectively. The top-1, 3, and 5 accuracies for both zoo and wild dataset are shown in Table 2

Classification	Few Shot	Siamese
Multi-column		
X	X	X
X	X	X

Table 2. Elephant Classifier Zoo and Wild Dataset Accuracies

4.2. Few Shot Learning

4.3. Siamese Network

5. Discussion

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. This style guide now has several important modifications (for example, you are no longer warned against the use of sticky tape to attach your artwork to the paper), so all authors should read this new version.

6. Conclusion

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. This style guide now has several important modifications (for example, you are no longer warned against the use of sticky tape to attach your artwork to the paper), so all authors should read this new version.

References

[1] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020). 2

[2] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015). 2

[3] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016. 2

[4] Deb, Debayan, et al. "Face recognition: Primates in the wild." 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, 2018.

[5] Korschens, Matthias, Bjorn Barz, and Joachim Denzler. "Towards automatic identification of elephants in the wild." arXiv preprint arXiv:1812.04418 (2018).

[6] Freytag, Alexander, et al. "Chimpanzee faces in the wild: Log-euclidean CNNs for predicting identities and attributes of primates." German Conference on Pattern Recognition. Springer, Cham, 2016.

[7] Brust, Clemens-Alexander, et al. "Towards automated visual monitoring of individual gorillas in the wild." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017.

[8] Van Zyl, Terence L., Matthew Woolway, and B. Engelbrecht. "Unique animal identification using deep transfer learning for data fusion in siamese networks." 2020 IEEE 23rd International Conference on Information Fusion (FUSION). IEEE, 2020.

[9] Hou, Jin, et al. "Identification of animal individuals using deep learning: A case study of giant panda." Biological Conservation 242 (2020): 108414.

[10] Korschens, Matthias, and Joachim Denzler. "Elpephants: A fine-grained dataset for elephant re-identification." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.

[11] Schneider, Stefan, Graham W. Taylor, and Stefan Kremer. "Deep learning object detection methods for ecological camera trap data." 2018 15th Conference on computer and robot vision (CRV). IEEE, 2018.

[12] Schneider, Stefan, et al. "Past, present and future approaches using computer vision for animal re-identification from camera trap data." Methods in Ecology and Evolution 10.4 (2019): 461-470.

[13] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[14] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." ICML deep learning workshop. Vol. 2. 2015.

[15] Tensorflow Object Detection Zoo: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md

[16] https://github.com/sicara/tf2-yolov4 2

[17] Chen, Liu, Kira, Wang, & Huang. A Closer Look at Few-shot Classification. In ICLR, 2019. 2

[18] Dhillon, Chaudhari, Ravichandran, & Soatto. A baseline for few-shot image classification. In ICLR, 2020. 3

[19] Chen, Wang, Liu, Xu, & Darrell. A New Meta-Baseline for Few-Shot Learning. arXiv, 2020. 3

[20] Few-Shot Learning Lectures, Shusen Wang: https://www.youtube.com/watch?v=U6uFOIURcD0 3

3