Whale Tail Recognition

Nicholas Williams, Boris Teodorovich, Austin Cheng

University of California, Santa Cruz

Abstract

The current state-of-the-art technique for identifying a whale from its *flukes* (tail) is for marine biologists to manually inspect thousands of photographs of whale tails comparing flukes' proportions, coloring, scars, birthmarks, bumps, and other unique markings. Our goal is to train a machine learning model to automatically identify a particular humpback whale from an image of its tail. We take inspiration from other image-based biometric identification tasks like facial recognition. Specifically, we apply techniques used in the FaceNet paper [1] for clustering similar images.

Dataset Cleaning and Pre-processing

The dataset consists of 25,000 humpback whale tail images and an associated id number that indicates a specific whale. There are 3,000 unique whale ids. Duplicate images as well as any photos that were not specifically of a whale tail were manually removed from the dataset. All images made grayscale.

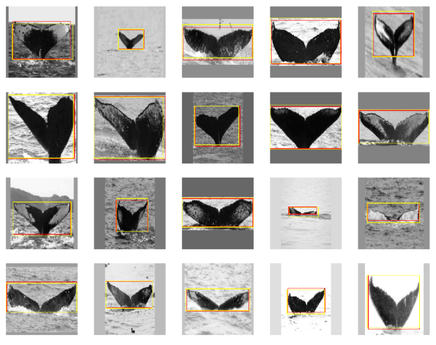

Figure 1: Bounding box model for data pre-processing

A bounding box model (fig. 1) was trained on the images, which were then cropped accordingly and resized to the 384x384 input size of the neural network. The aspect ratio for each image was chosen to be close to the average.

Approach

Model Overview

To associate a whale tail to a specific identity, our model generates an embedding vector $f(a) \in \mathbb{R}^{128}$ from an image a, then compares against a database of embeddings for known whales. Two images are identified as belonging to a particular whale if the squared euclidean distance between the two whale tail embeddings is within the margin threshold. The embeddings are generated by a neural network optimized on the *triplet loss* function.

Triplet Loss Function

The triplet loss function is defined on three inputs: an anchor image a, an image of the same identity p, and an image of a different identity n. The triplet loss L for a single triplet (a, p, n) is defined as

$$L(a, p, n) = max(\|f(a) - f(p)\|_2^2 - \|f(a) - f(n)\|_2^2 + margin, 0)$$

By minimizing this loss, we push the euclidean distance $d(a, p)$ to 0 and $d(a, n)$ to be greater than $d(a, p)$+margin, ensuring that embeddings from images with the same identity are closely clustered in the embedding space while embeddings of different identities are a maximum distance apart. We utilize *semi-hard* triplets satisfying

$$d(a, p) < d(a, n) < d(a, p) + margin$$

i.e. triplets where the negative is not closer to the anchor than the positive, but loss is still positive.

Figure 2: Whale Triplets

Neural Network Architecture

Our 5 inception module architecture inspired by GoogLeNet [2]. dynamically selects features of varying sizes to filter out. This works especially well for detecting objects of different scale and in different locations in an image. The last layer of our network performs l2 normalization which constrains the embeddings to live on a hypersphere of dimension 128.

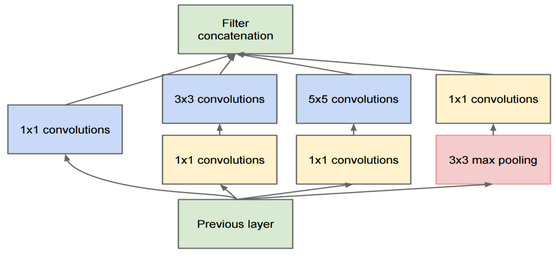

Figure 3: Inception Model [2]

Results

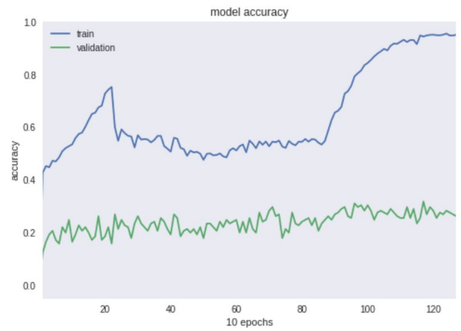

Figure 4: Accuracy

After training for 130 epochs our model achieved an accuracy of 27%. The low validation accuracy is likely caused by our small training dataset size. Our solution although not perfect, provides a massive time savings tool for marine biologists.

Challenges

The embeddings of the network collapsed quite easily depending on the specific optimization algorithm, learning rate, margin term, gradient clipping value, minibatch size, etc. This model requires significant attention to hyper-parameters.

Conclusion

Deep representation learning of faces for use in recognition is easily extendable to recognition of whales, and likely many more animals or objects with a unique marker.

References

[1] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015.

[2] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.