Nathan Little

An Attempt to Identify Humpback Whales from Flukes Using Similarity Comparison Neural Networks.

Abstract

To track the recovery of the worlds humpback whale population, a fully autonomous system which can identify individuals from photographs of their flukes is necessary. This paper evaluates the cutting-edge technology involved in developing such systems by answering the question: What is the most effective similarity network for the identification of humpback whales from flukes? In the study described here, two similarity neural networks were developed and tested on an image data set in an attempt to produce metrics which could be used to analyse and compare them to produce actionable insights to assist in the development of similar solutions.

Figure 1: An example image from the dataset used in the study. (Happy whale, 2015)

Background

Identifying humpback whales from flukes was originally done by manually marking descriptive features such as nicks and scratches or the black and white patterns on the flukes. This was then developed into algorithmic methods such as describing the ratio of black to white in specific areas of the flukes. Naturally this task lends itself to a computer vision system as it can be modelled as an image classification task. However, the main problem is the class imbalance within the datasets as most whales have only a couple of images, which makes it difficult to train neural networks. Similarity networks learn differences between images rather than identifying features of classes. This makes them ideal for one-shot classification problems such as this one. A Kaggle competition (Happy whale, 2015) was made by the organisation happy whale, who also provided the dataset used and inspired many of the state-of-the-art solutions.

Methodology

The two comparison network architectures selected were Siamese and triplet loss networks. The Siamese network functions by passing a pair of input images through the same core network, before then computing the distance between the embeddings and passing the result through a function to produce a similarity score.

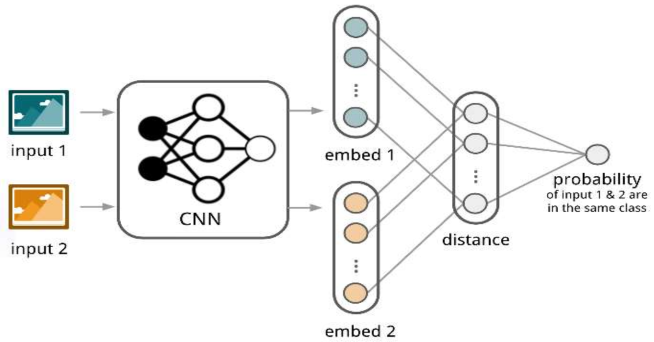

Figure 2: A diagram of the flow of information through a Siamese network (Weng , 2018)

Triplet loss networks function similarly by passing triplets of images, made of an anchor, positive and negative examples of a class, through the network to produce embeddings. Before using a loss function that aims to make the distance from the anchor to the positive less than the distance from the anchor to the negative.

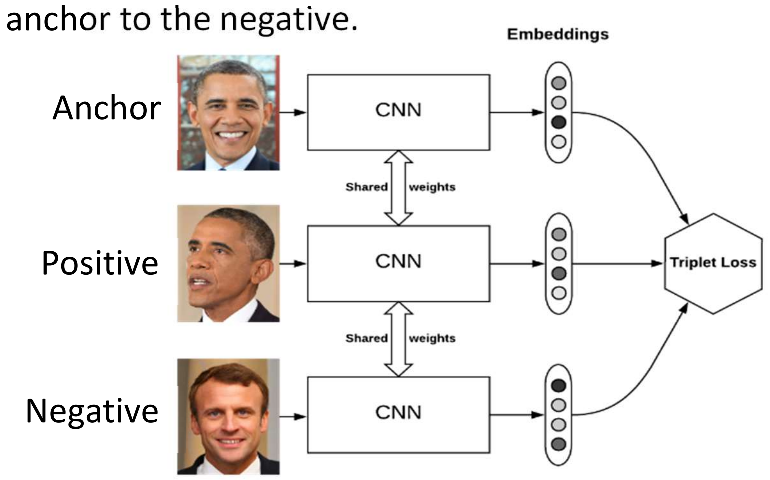

Figure 3: A diagram showing the structure of the triplet loss mechanism. (Moindrot, 2018)

Each Image was processed by converting to black and white and removing poor quality images based on a criterion centred around removing images with extreme resolution size/shape before standardising their dimensions. Both networks were tested using N-way one shot testing which involves passing the test image through the network with one image of each class before then ranking which each class based on the distance between the output embeddings, the shortest distance is the prediction.

Results

Unfortunately, both models failed to properly learn patterns from the data resulting in both models predicting the same outputs for every input. However, by comparing the methodology used in this study to others it was possible to estimate why this was the case. Most probably, the first problem was that the core network used was too simple, this was caused by a lack of computing power. Building more complex networks caused out-of-memory errors as well as increasing training time to be beyond what was appropriate given the projects timescale. The second issue was that the image pre-processing pipeline was insufficient as it did not remove enough irrelevant information which meant the identifying small patterns on the flukes were lost in a large amount of varying information between images.

Figure 4: An example of a high quality image (left) and a poor quality image (right) from the happywhale dataset

Possible solutions to both these problems were identified, the necessary computing power could be sourced through either improved local hardware or more likely, by utilising a cloud computing service. Either would allow more complex networks to be built and trained in time. The image processing pipeline could be improved by utilising a region of interest system to crop the images to the flukes or even an image masking solution to remove absolutely all irrelevant information. This would prevent images such as those seen above from hindering the networks learning capabilities.

The key conclusions of this research are that this task is complex and so requires a sufficiently complex core model and a fairly advanced image processing pipeline to give the system a chance of learning the relevant patterns within the data.

References

Happy whale, 2015. Our Mission. [Online] Available at: https://www.happywhale.com/about [Accessed 22 04 2020]

Moindrot, O., 2018. Triplet Loss and Online Triplet Mining in TensorFlow. [Online] Available at: https://omoindrot.github.io/triplet-loss [Accessed 28 April 2020].

Weng , L., 2018. Meta-Learning: Learning to Learn Fast. [Online] Available at: https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html [Accessed 28 April 2020].