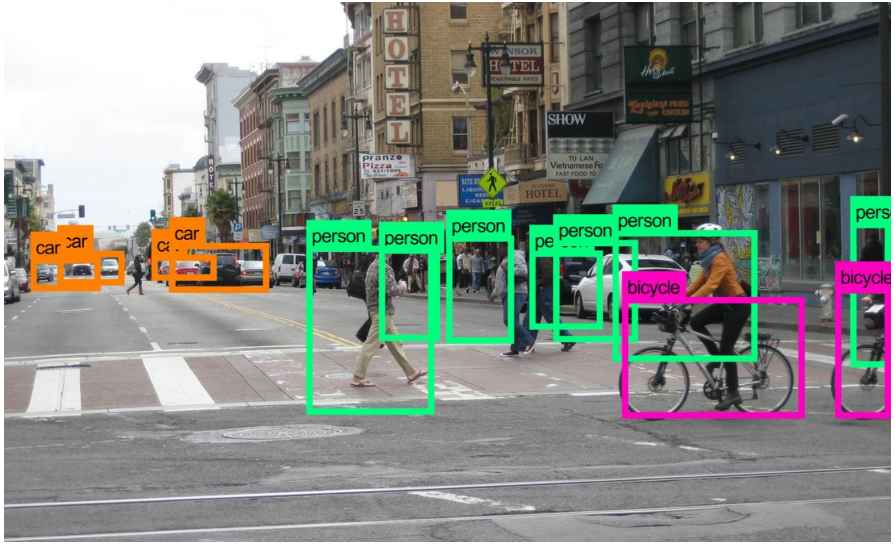
Identification of endangered animals using CV

Identification of Problem

The identification of objects is one of the most common applications in Computer Vision, which is being extensively applied for:

- Detection
- Classification



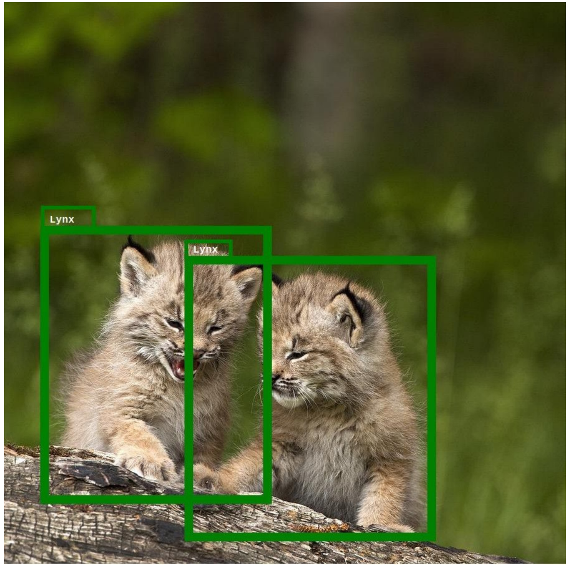
Application: Conservation of Endangered Animals

Challenges

- Need manual effort of experts
- Installation of Technologies (camera trap, drones)
- Difficult geographical regions

Modern CV techniques as a solution

- to increase the capacity to in different contexts



Understanding our Data (Platypus)

Mammal

Land & Water

Australian

Since it shares similar features to other animals, there are higher chances that it can be misclassified as other animals.

Research Challenges of This Problem

1. Land/Water environment

Source: https://content.techgig.com/technology/8-object-detection-challenges-data-scientists-should-be-aware-of/articleshow/89446600.cms

Research Challenges of This Problem (Cont'd)

1. Land/Water environment

2. Different shapes

Source: https://content.techgig.com/technology/8-object-detection-challenges-data-scientists-should-be-aware-of/articleshow/89446600.cms

Research Challenges of This Problem (Cont'd)

1. Land/Water environment

2. Different shapes

3. Nocturnal behaviour

Source: https://content.techgig.com/technology/8-object-detection-challenges-data-scientists-should-be-aware-of/articleshow/89446600.cms

Research Challenges of This Problem

1. Land/Water environment

2. Different shapes

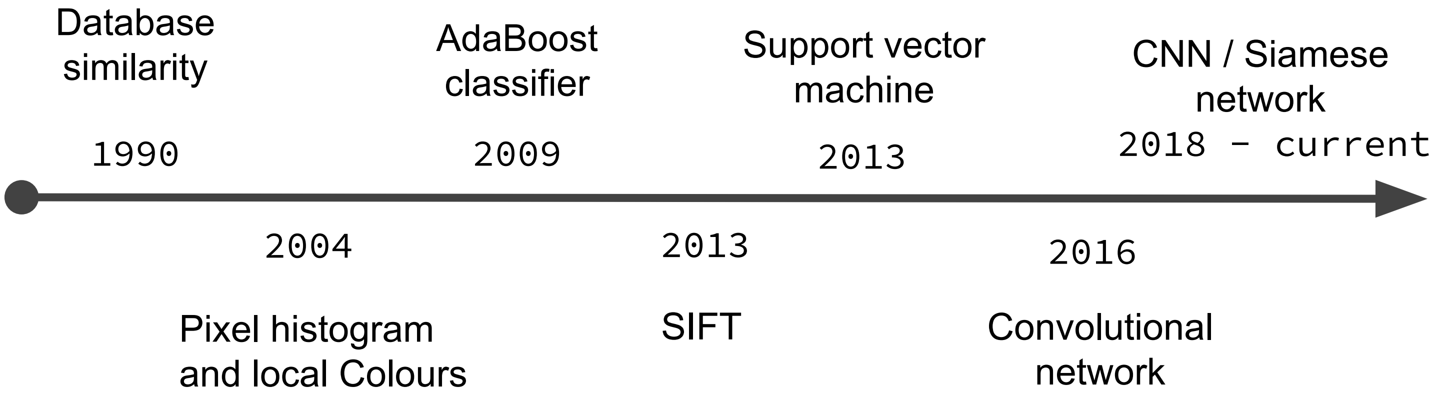
3. Nocturnal behaviour

4. Poses

Source: https://content.techgig.com/technology/8-object-detection-challenges-data-scientists-should-be-aware-of/articleshow/89446600.cms

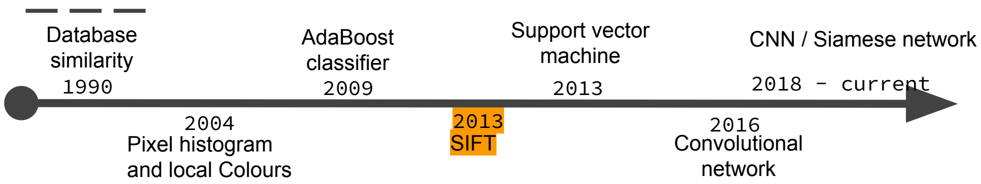
Literature Review

Past, present and future approaches using computer vision for animal identification from camera trap data.



Source: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13133

Past Research in Animal Identification



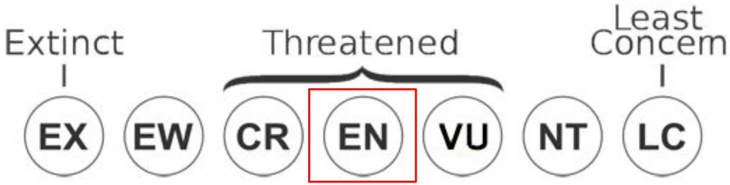
Endangered Animal: Manta Ray

Methodology: SIFT

Test size : 720

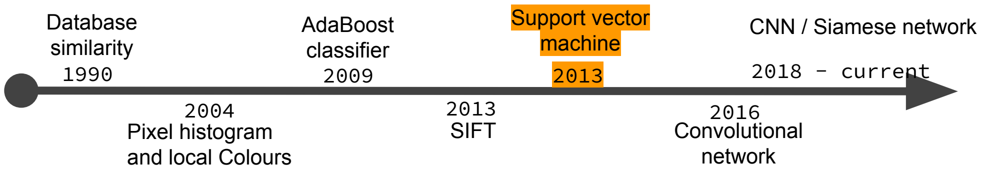
Num. classes: 265

Top-1 accuracy (%): 51.0



Source: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13133

Past Research in Animal Identification (cont'd)



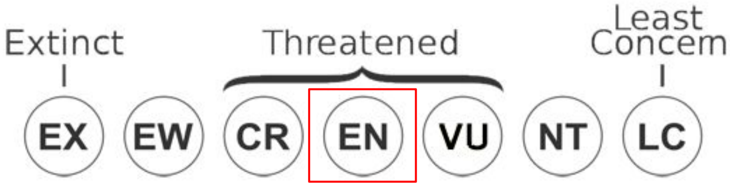
Endangered Animal: Support vector machine

Methodology: Support vector machine

Test size : 2,078

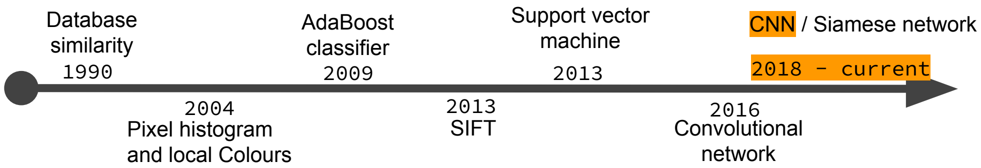
Num. classes: 276

Top-1 accuracy (%): 59.0



Source: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13133

Past Research in Animal Identification



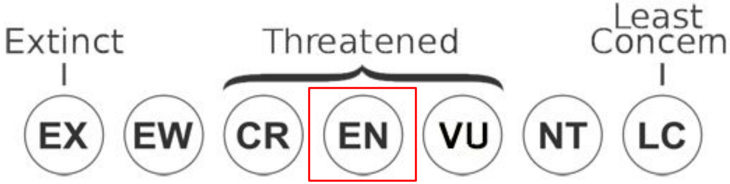
Endangered Animal: Chimpanzee (C-Tai)

Methodology: Convolutional network

Test size : 1,146

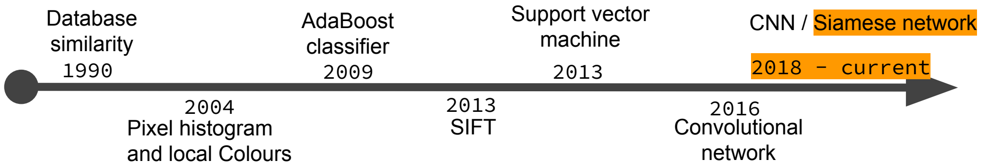
Num. classes: 286

Top-1 accuracy (%): 75.7



Source: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13133

Past Research in Animal Identification (cont'd)



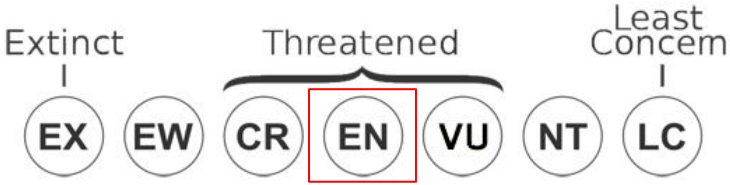
Endangered Animal: Golden Monkey

Methodology: Siamese network

Test size : 241 videos

Num. classes: 49

Top-1 accuracy (%): 75.8



Source: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13133

Traditional ML method: Bag of SIFT Feature Method

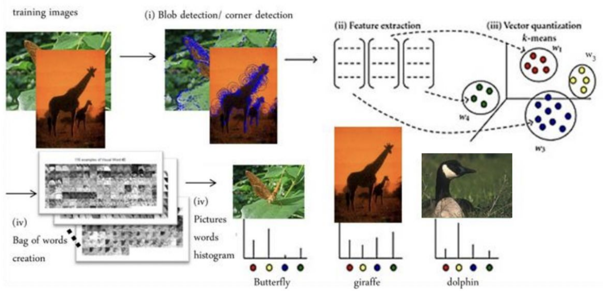


Figure 2. Animal recognition using BoF model training stages

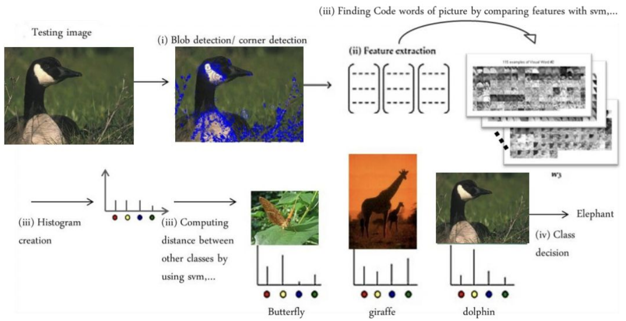


Figure 4. Animal recognition using BoF model testing stages

Evaluating classification strategies in Bag of SIFT Feature method for Animal Recognition

Leila Mansourian, Muhamad Taufik Abdullah, Lilli Nurliyana Abdullah and Azreen Azman

Department of Multimedia, Faculty of Computer Science and Information Technology,
University Putra Malaysia, 43400 UPM Serdang, Selangor, MALAYSIA
Corresponding addresses
l_mansourian@yahoo.com, {mta, liyana, azreenazman}@upm.edu.my

Abstract: These days automatic image annotation is an important topic and several efforts are made to solve the semantic gap problem which is still an open issue. Also, Content Based Image Retrieval (CBIR) cannot solve this problem. One of the efficient and effective models for solving the semantic gap and visual recognition and retrieval is Bag Of Feature (BoF) model which can quantize local visual features like SIFT perfectly. In this paper we investigated the potential usage of Bag of SIFT Feature in animal recognition. Also, we specified which classification method is better for animal pictures.

Keywords: Bag of Feature, SIFT feature, feature quantization, Content Based Image Retrieval (CBIR), image annotation, Support Vector Machines (SVM).

1. Introduction

In Content Based Image Retrieval (CBIR) [1], proposed in the early 1990s, images are automatically indexed by extracting their different low level features such as texture, color and shape. Semantic gap is a well-known problem among Content Based Image Retrieval (CBIR) systems. This

The rest of this paper is structured as follows. In Section 2, we review some related works in this area. Section 3 presents our experiment. A discussion about the experimental results and the usefulness of BoW model for animal recognition is presented in Section 4. The paper is concluded and some future works are suggested in Section 5.

2. Related works

At the starting point of BoF methodology we must identify local interest regions or points. Then we can extract features from these points, both of which described in the following section

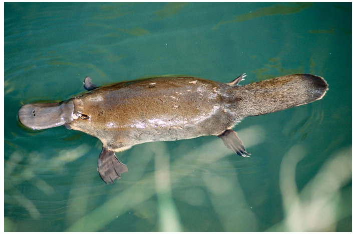
2.1 Interest Point Detection

There are several distinguished methods which are listed below [5]

(i) Harris-Laplace regions

In this method corners are detected by using Laplacian-of-Gaussian operator in scale-space.

Training Steps

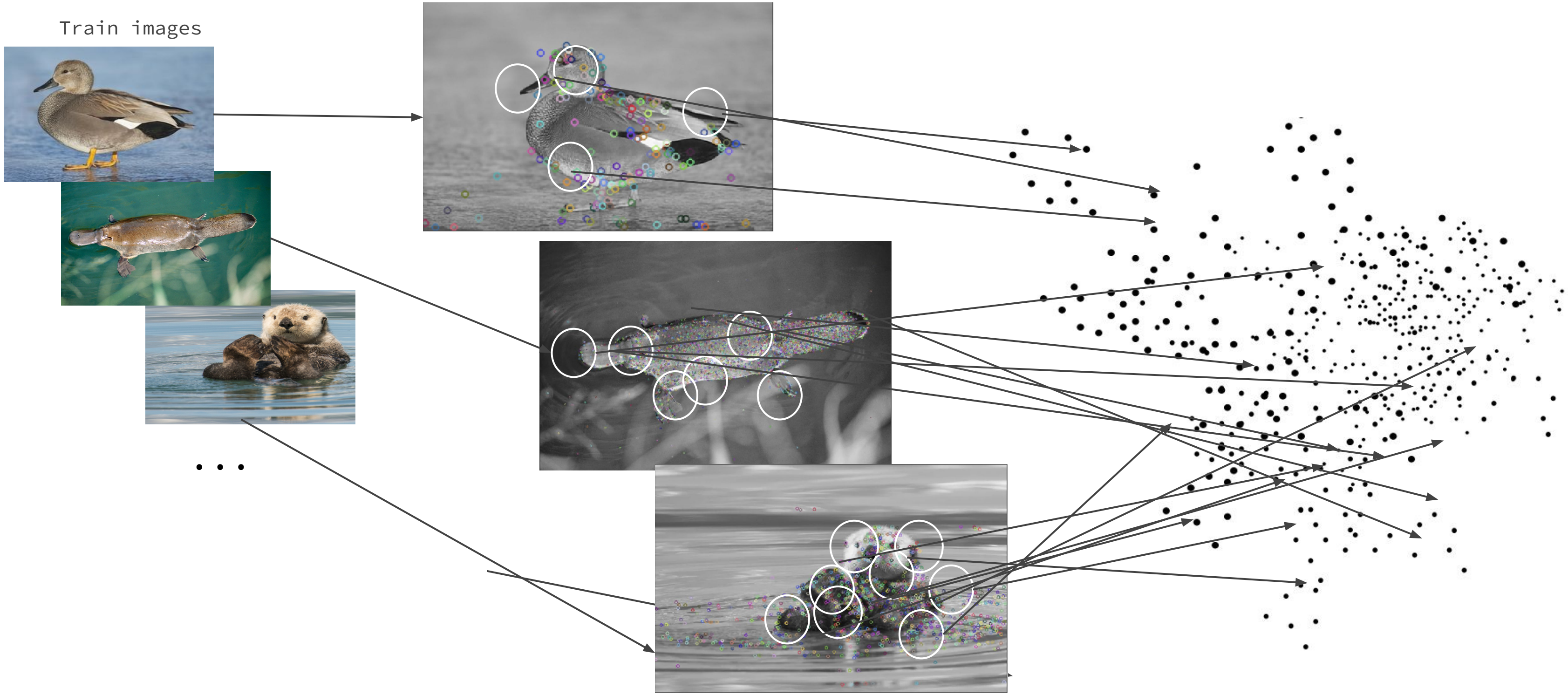
Input image 

Also for non-platypus images

  ...

Training Step 1: Feature Extraction

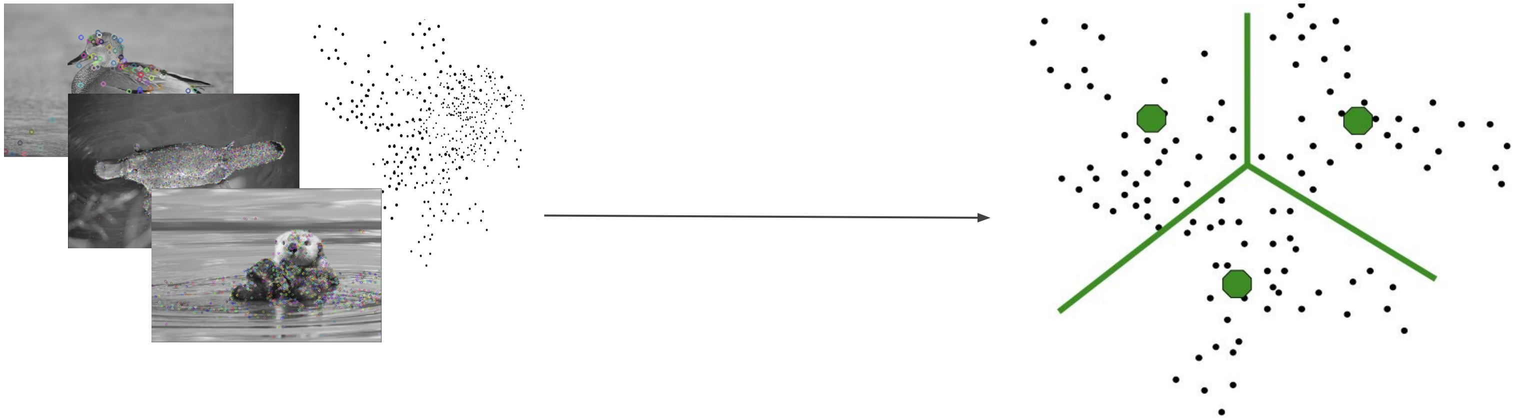
SIFT: get keypoints, descriptors



Training Step 2: Quantization of Feature Space

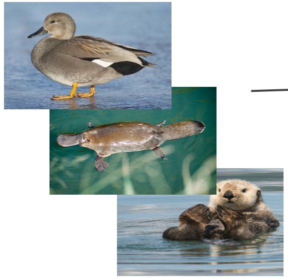
The centre of each cluster is used as a visual word by using **K-means**

SIFT: get keypoints, descriptors

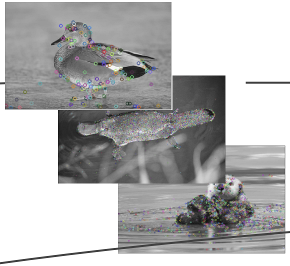


Training Step 3: Bag of words (BoW)

Train images

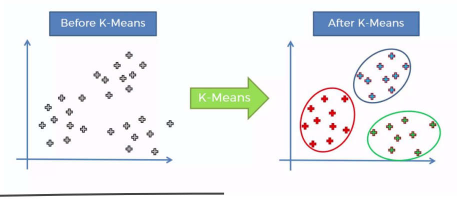


SIFT:
get keypoints, descriptors



Extract features locally

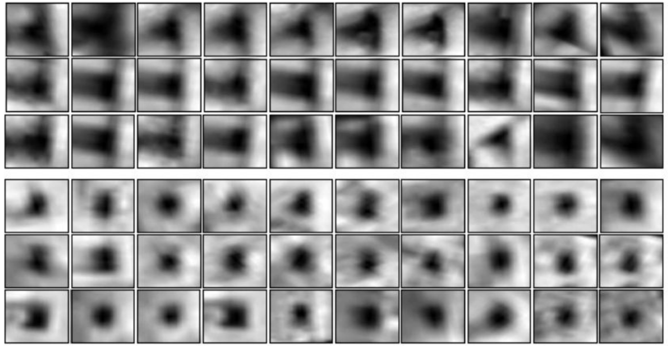
Cluster local features



Bag of words

c1

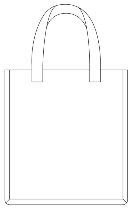
c2



Each group of patches belongs to the same visual word.

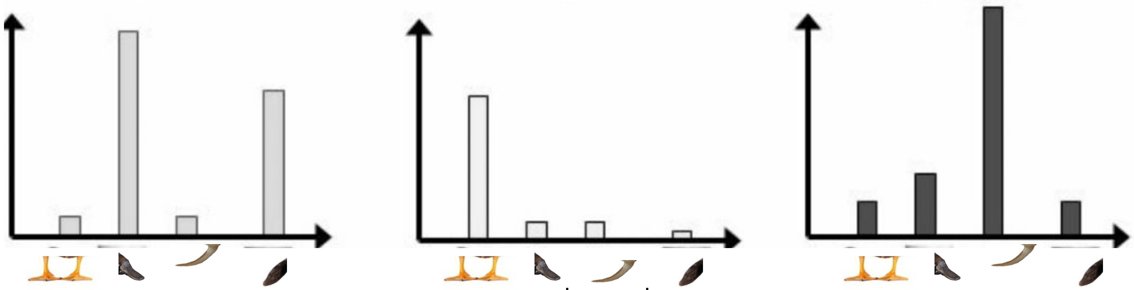
Bag of words (BoW)

Conceptual object



Histogram of visual words

Frequency

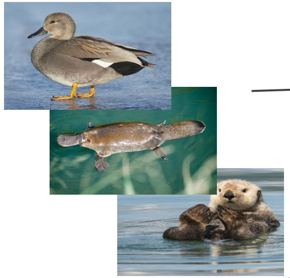


codewords

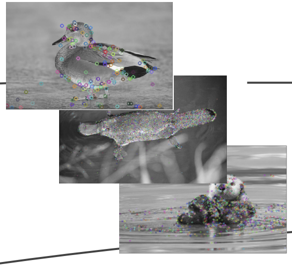
Source: https://towardsdatascience.com/bag-of-visual-words-in-a-nutshell-9ceea97ce0fb

Training Step 3: Bag of words (BoW)

Train images

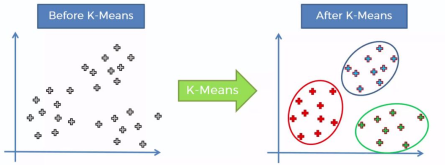


SIFT:
get keypoints, descriptors



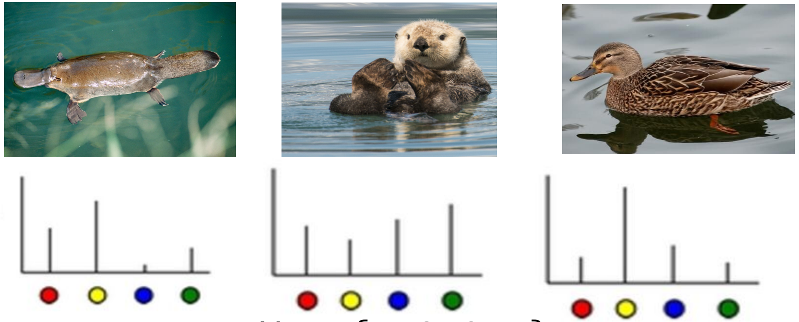
Extract features locally

Cluster local features



Bag of words

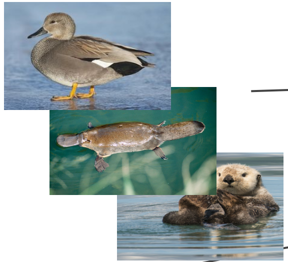
Representation using histogram



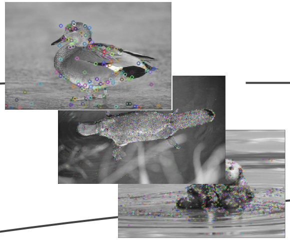
V : {4,3,2,7}

Training Step 4: Classification

Train images

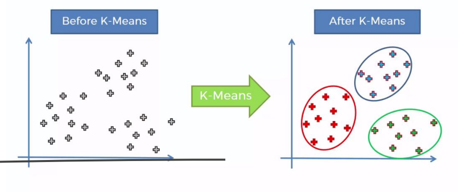


SIFT:
get keypoints, descriptors



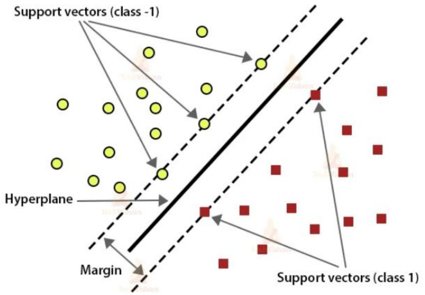
Extract features locally

Cluster local features



Bag of words

Classifier with SVM



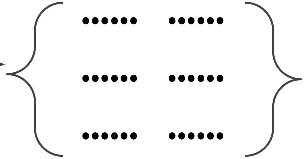
Output:
1 or 0

Prediction Steps

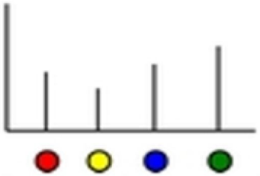
Test image

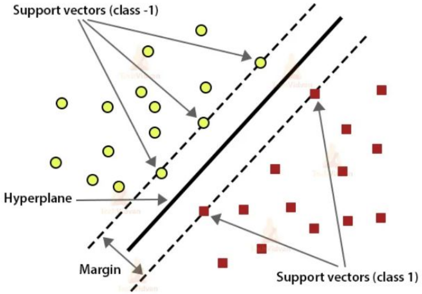
SIFT:
get keypoints, descriptors

Extract features locally



Histogram: getting frequency





Computing distance with
known classes using SVM

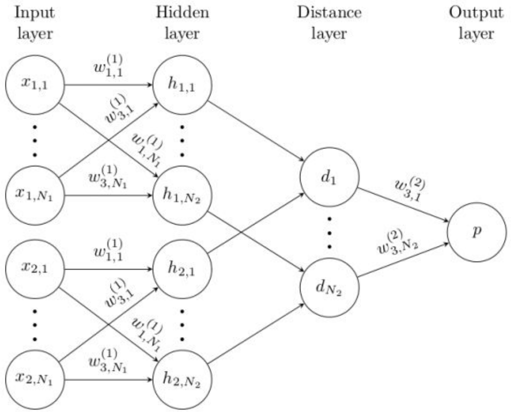
Platypus!

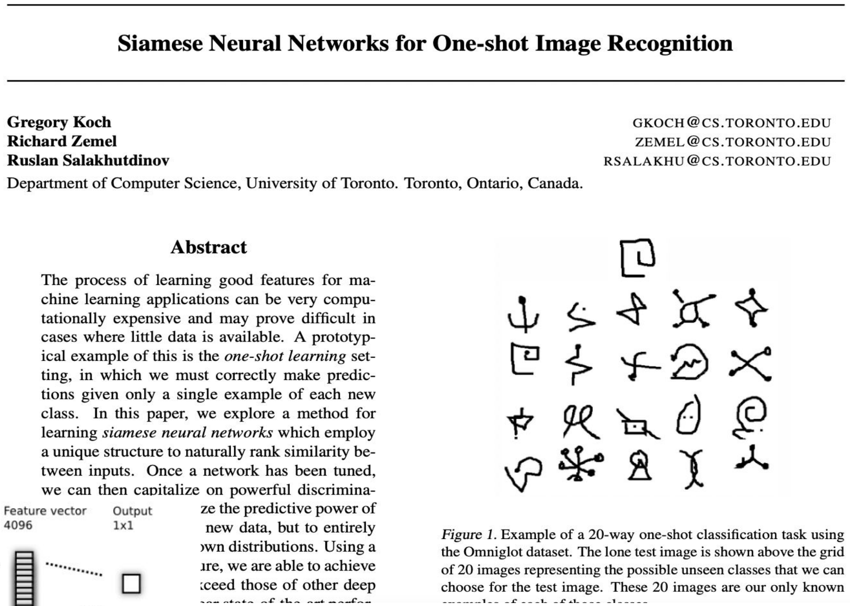
Modern ML method: Siamese Network

Siamese Neural Network (SNN) is a class of neural network architectures that contain two or more identical sub-networks.

Source: https://builtin.com/machine-learning/siamese-network

Siamese Neural Networks for One-shot Image Recognition





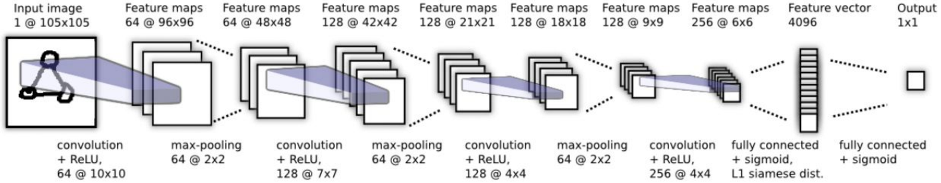
Siamese Neural Networks for One-shot Image Recognition

Gregory Koch GKOCH@CS.TORONTO.EDU
Richard Zemel ZEMEL@CS.TORONTO.EDU
Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Department of Computer Science, University of Toronto. Toronto, Ontario, Canada.

Abstract

The process of learning good features for machine learning applications can be very computationally expensive and may prove difficult in cases where little data is available. A prototypical example of this is the *one-shot learning* setting, in which we must correctly make predictions given only a single example of each new class. In this paper, we explore a method for learning *siamese neural networks* which employ a unique structure to naturally rank similarity between inputs. Once a network has been tuned, we can then capitalize on powerful discrimina-
ze the predictive power of
new data, but to entirely
›wn distributions. Using a
ıre, we are able to achieve
ceed those of other deep

Figure 1. Example of a 20-way one-shot classification task using the Omniglot dataset. The lone test image is shown above the grid of 20 images representing the possible unseen classes that we can choose for the test image. These 20 images are our only known

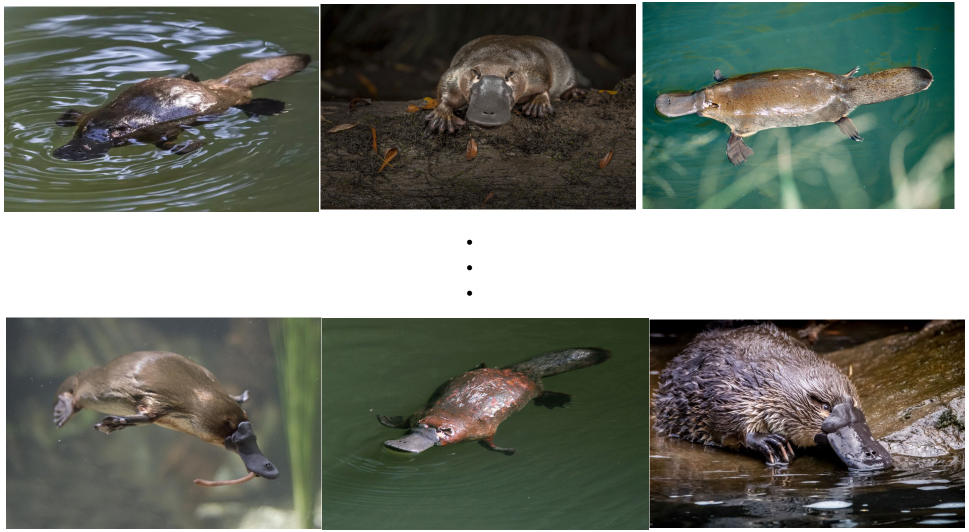


Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Training data

Not Platypus

Platypus

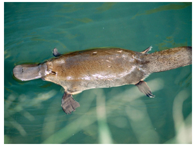


Training data

Positive Samples

( ,  ,1)

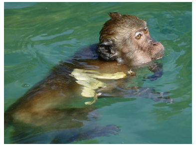
( ,  ,1)

( ,  ,1)

Negative Samples

( ,  ,0)

( ,  ,0)

( ,  ,0)

Training Step: Pairwise Inputs

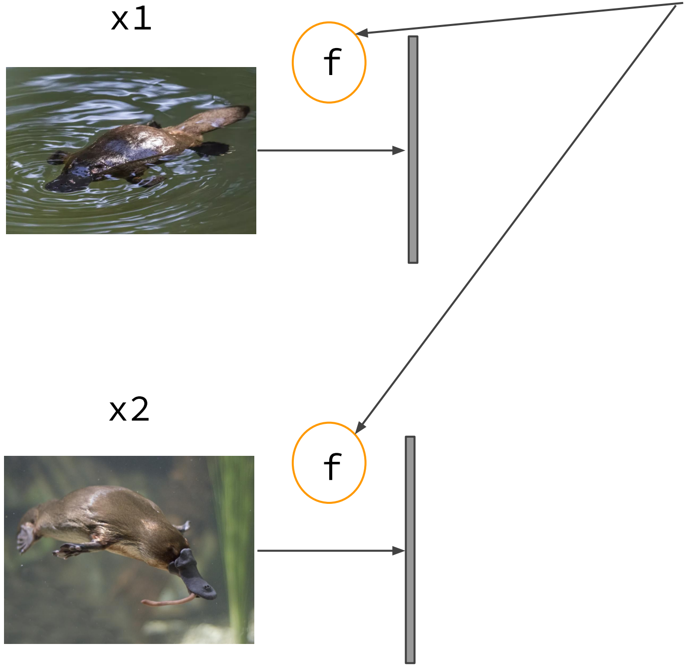
Input 1: x1

Input 2: x2

Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

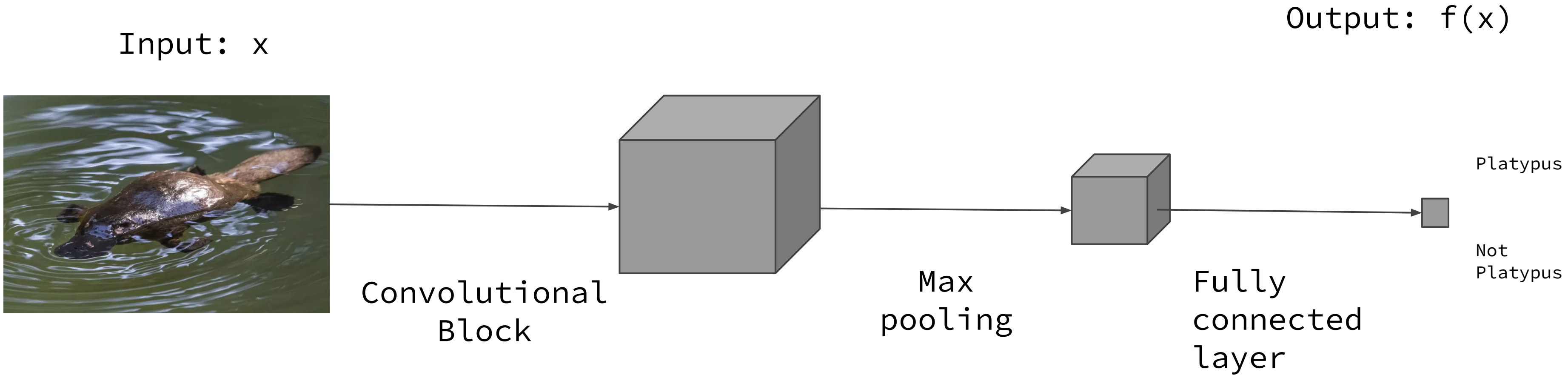
Training Step: Feature Extraction with CNN

put thru same convolutional neural network



Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Convolutional Neural Network (CNN) for Feature Extraction



Training Step (cont'd)

put thru convolutional neural network and get individual feature vectors

x1

f

h1 = f(x1)

x2

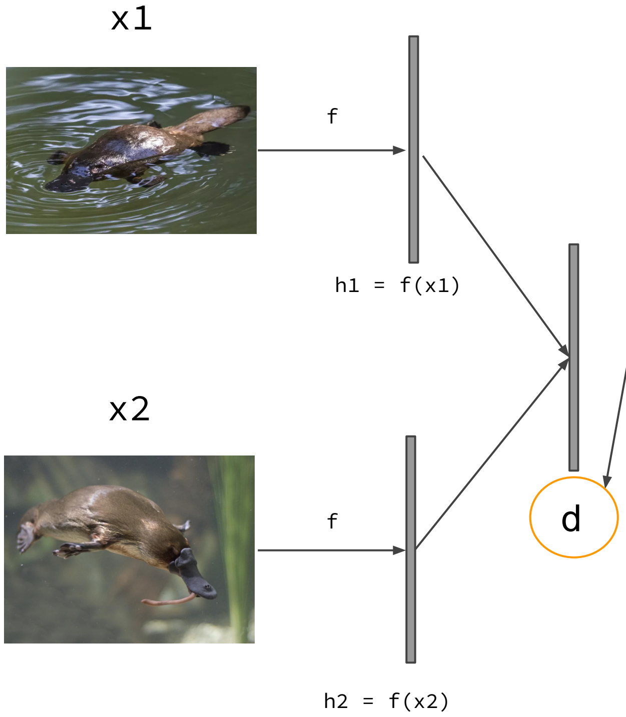
f

h2 = f(x2)

Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Training Step (cont'd)

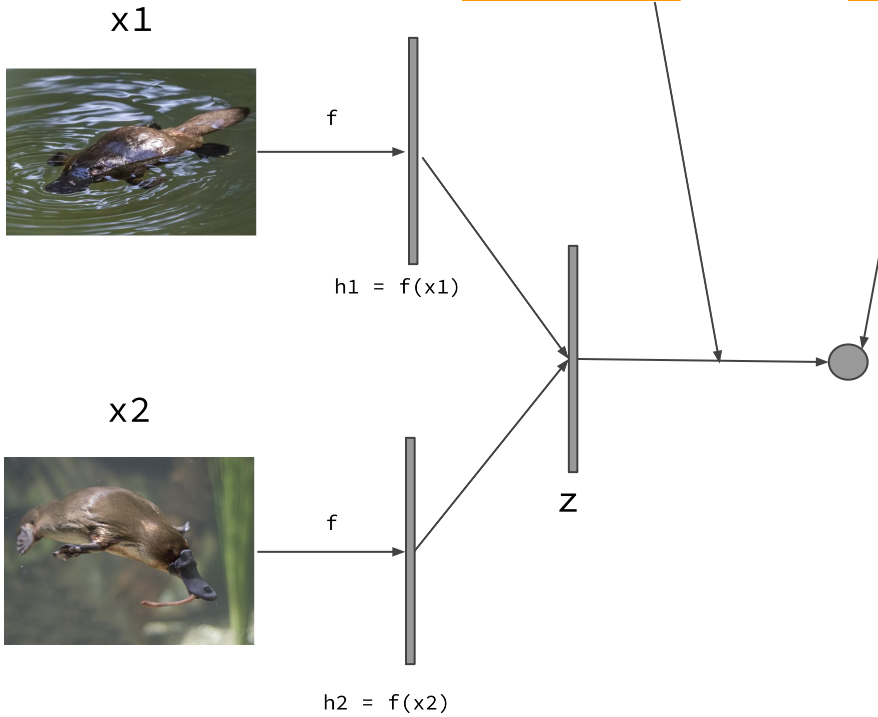
calculate distance between feature vectors



Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Training Step (cont'd)

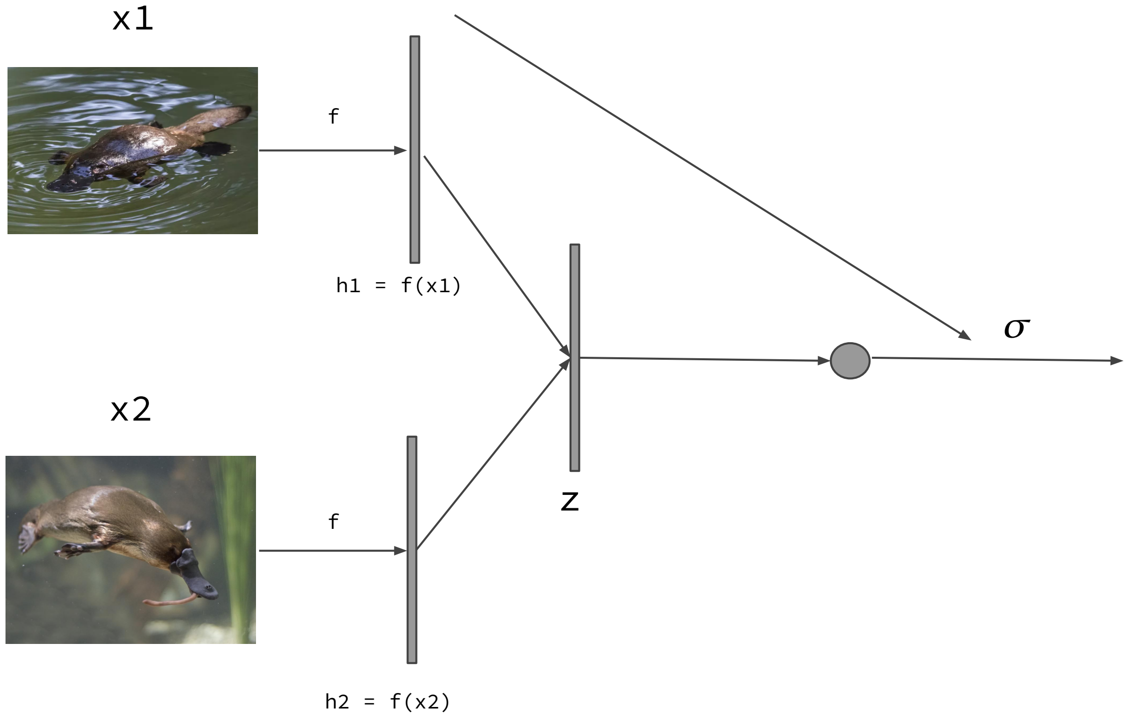
pass thru dense layer and get a scalar from it



Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Training Step (cont'd)

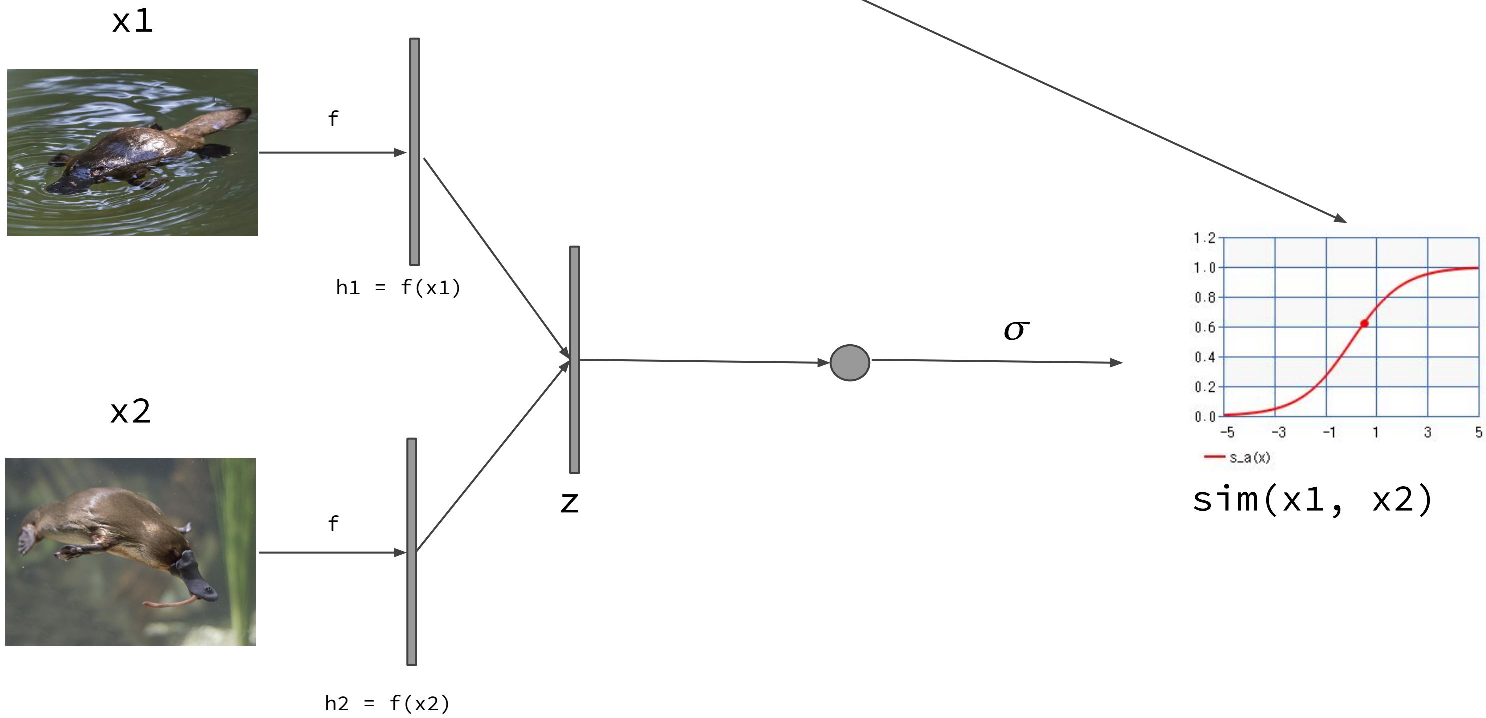
Use sigmoid as the activation function



Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Training Step (cont'd)

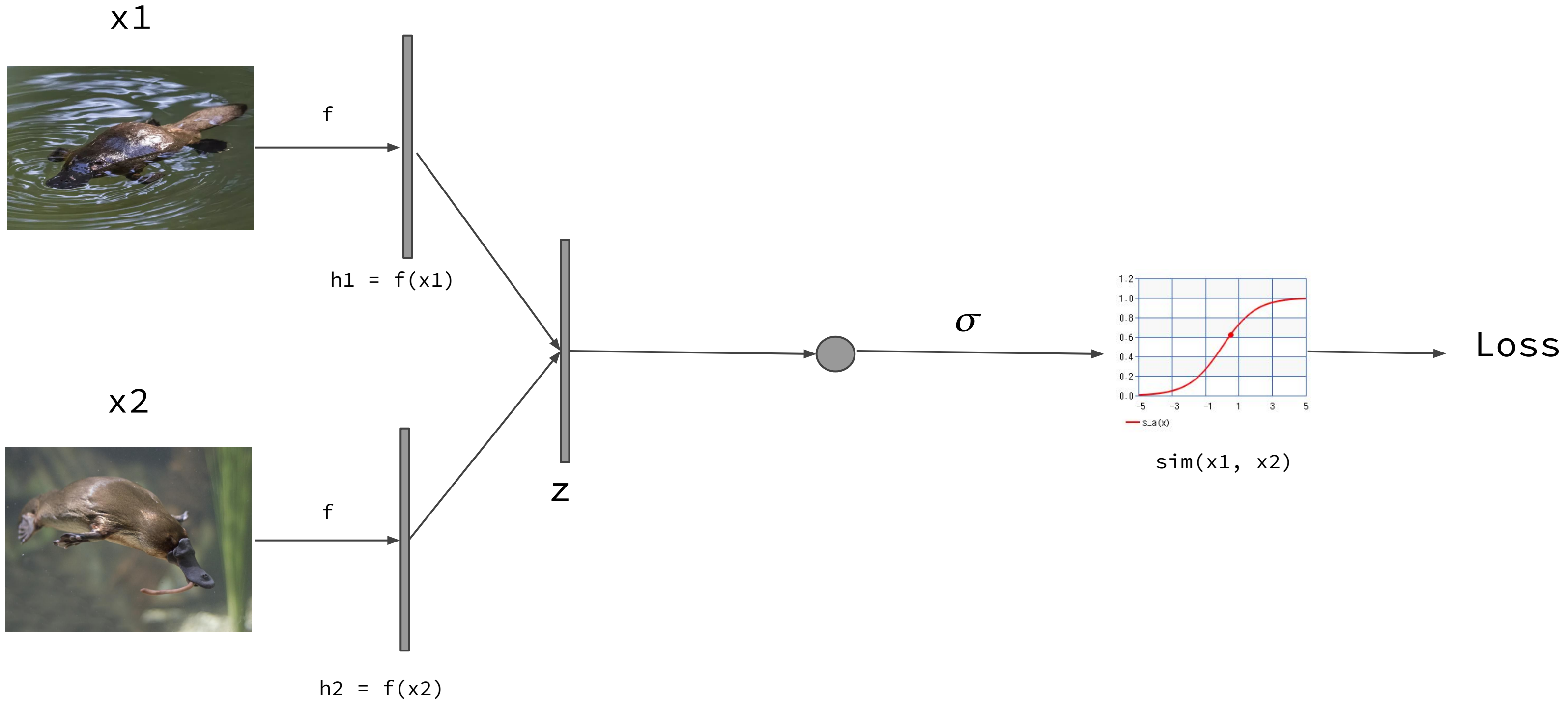
Similarity of inputs: x1, x2



Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Training Step (cont'd)

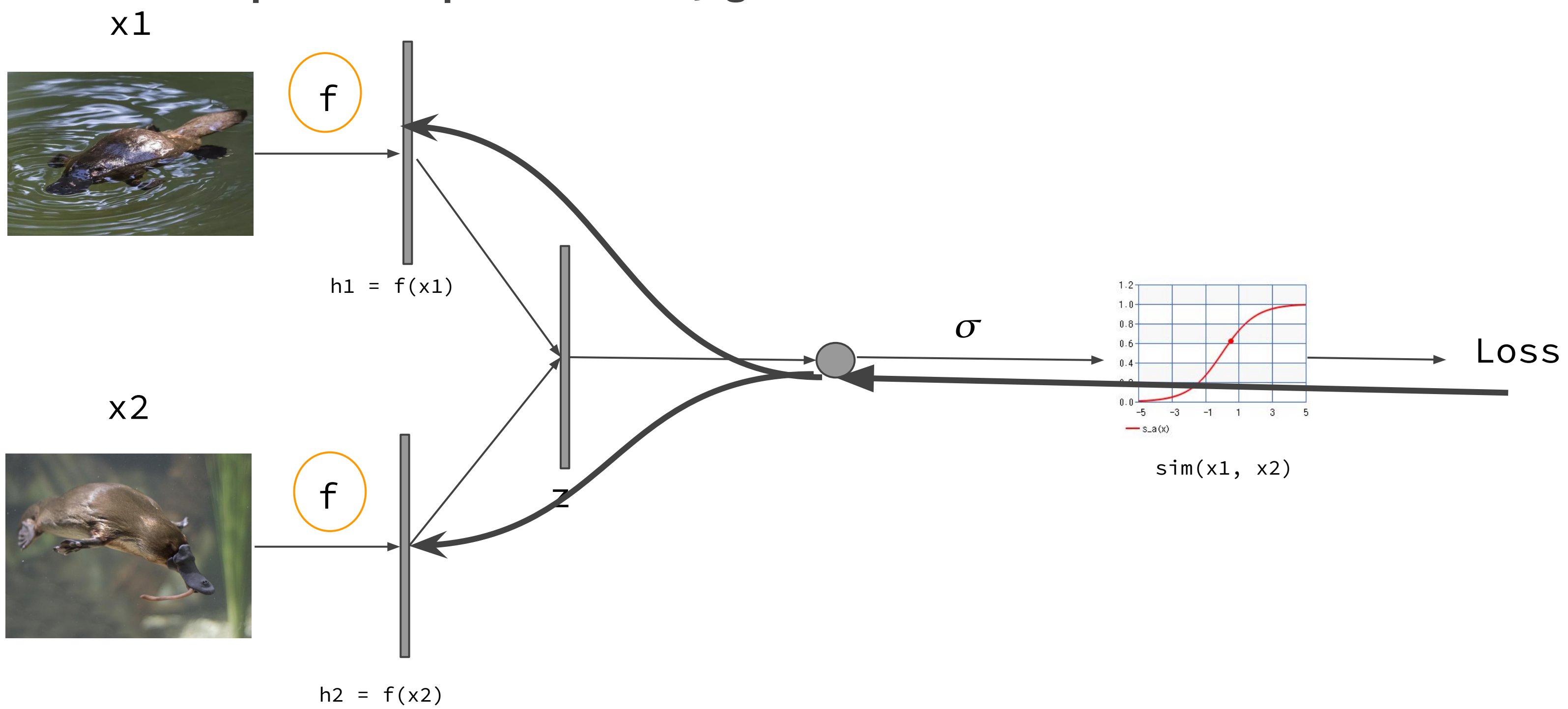
Backpropagation using loss function



Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

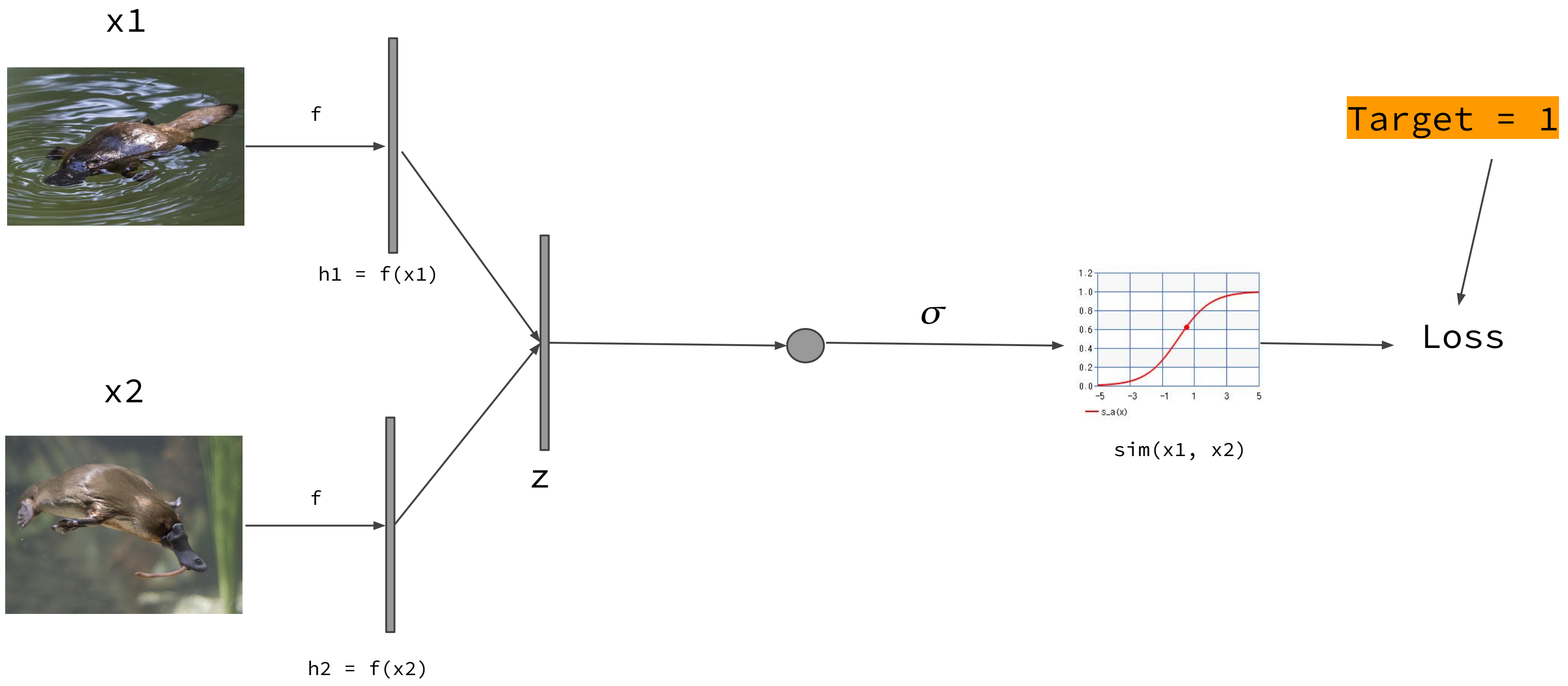
Training Step (cont'd)

Update the parameters by gradient descent



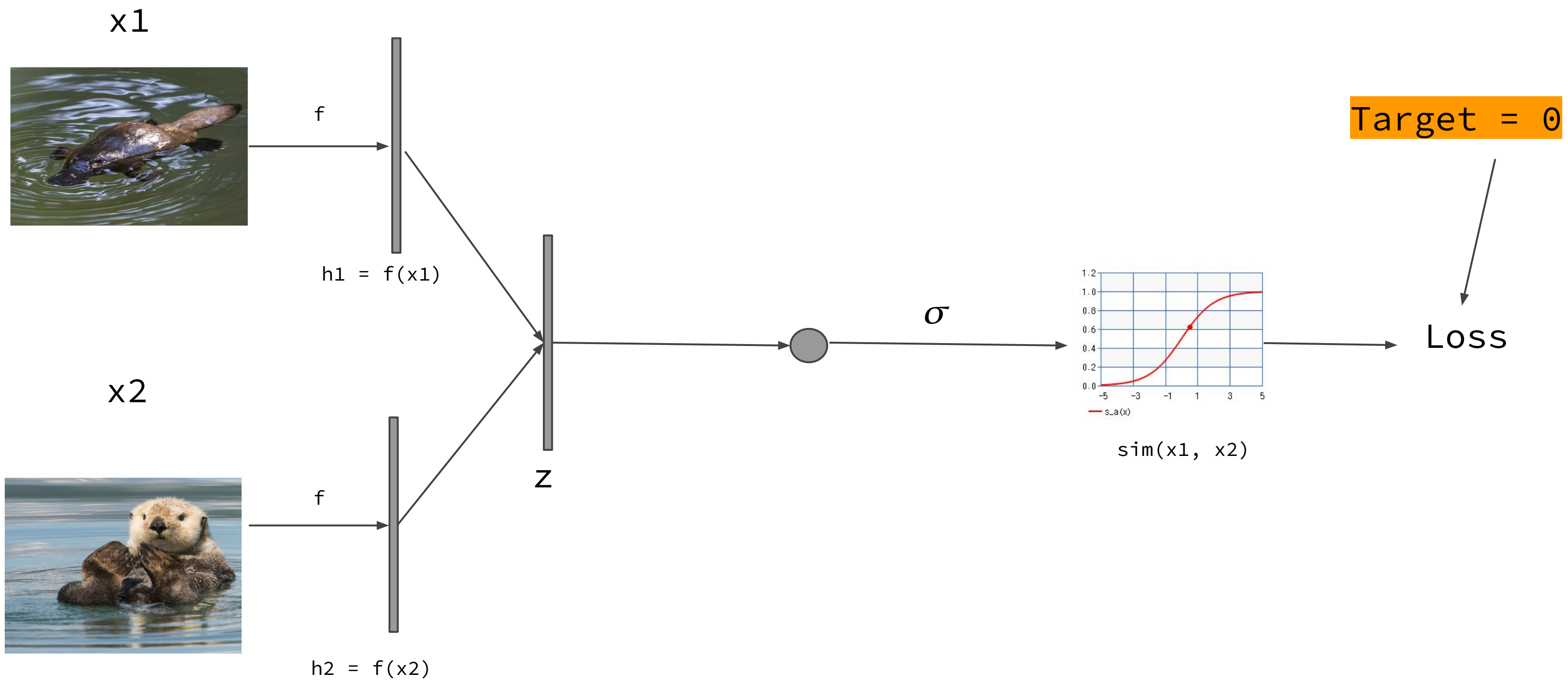
Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Training Step: Positive Case (cont'd)



Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Training Step: Negative Case (cont'd)



Source: https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

Prediction Step

Query

?

Sim = 0.2

Not Platypus
(Duck)

Prediction Step (Cont'd)

Query

?

Sim = 0.2

Sim = 0.6

Not Platypus
(Duck)

Not Platypus
(Otter)

Prediction Step (Cont'd)

Query

?

Sim = 0.2

Sim = 0.6

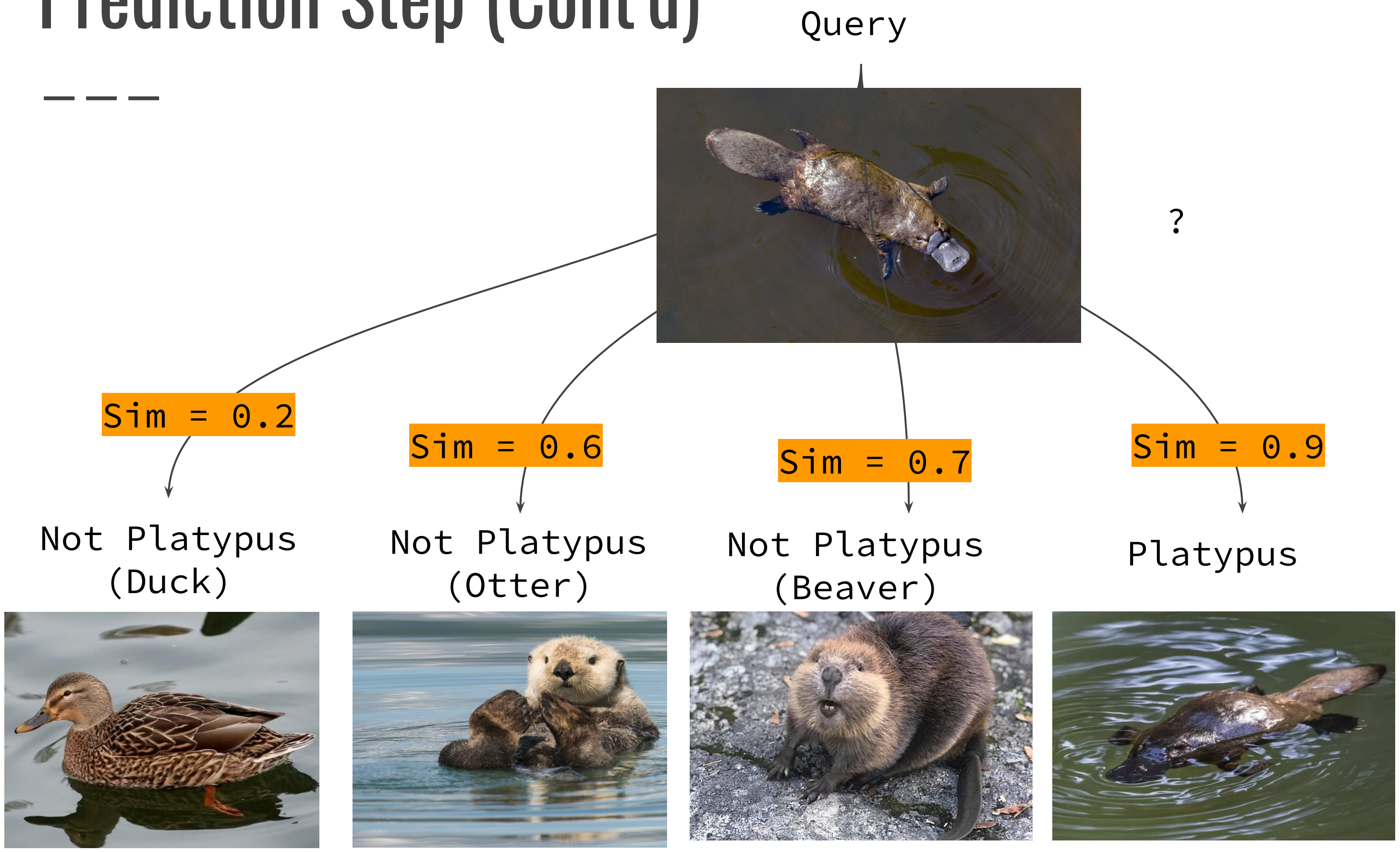
Sim = 0.7

Not Platypus (Duck)

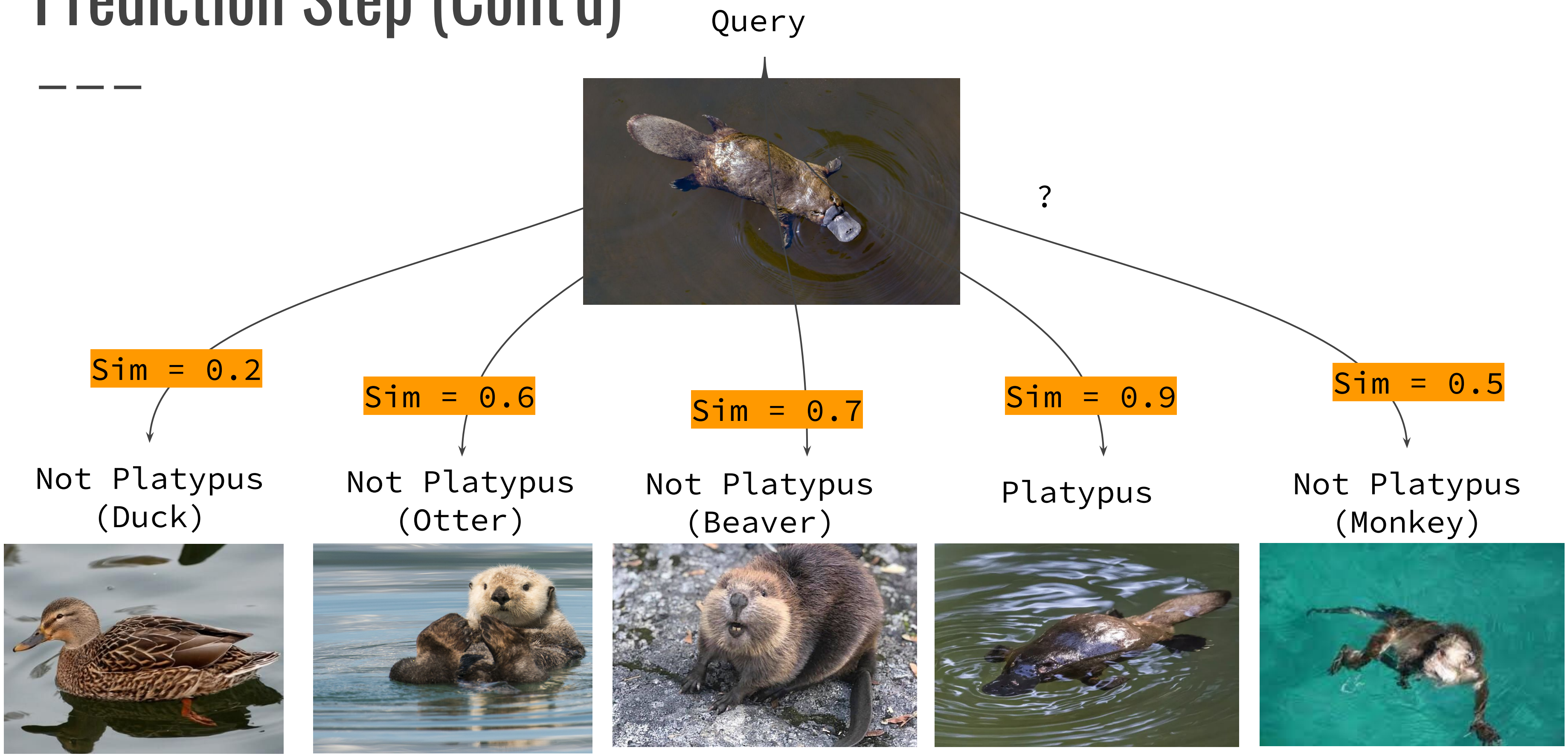
Not Platypus (Otter)

Not Platypus (Beaver)

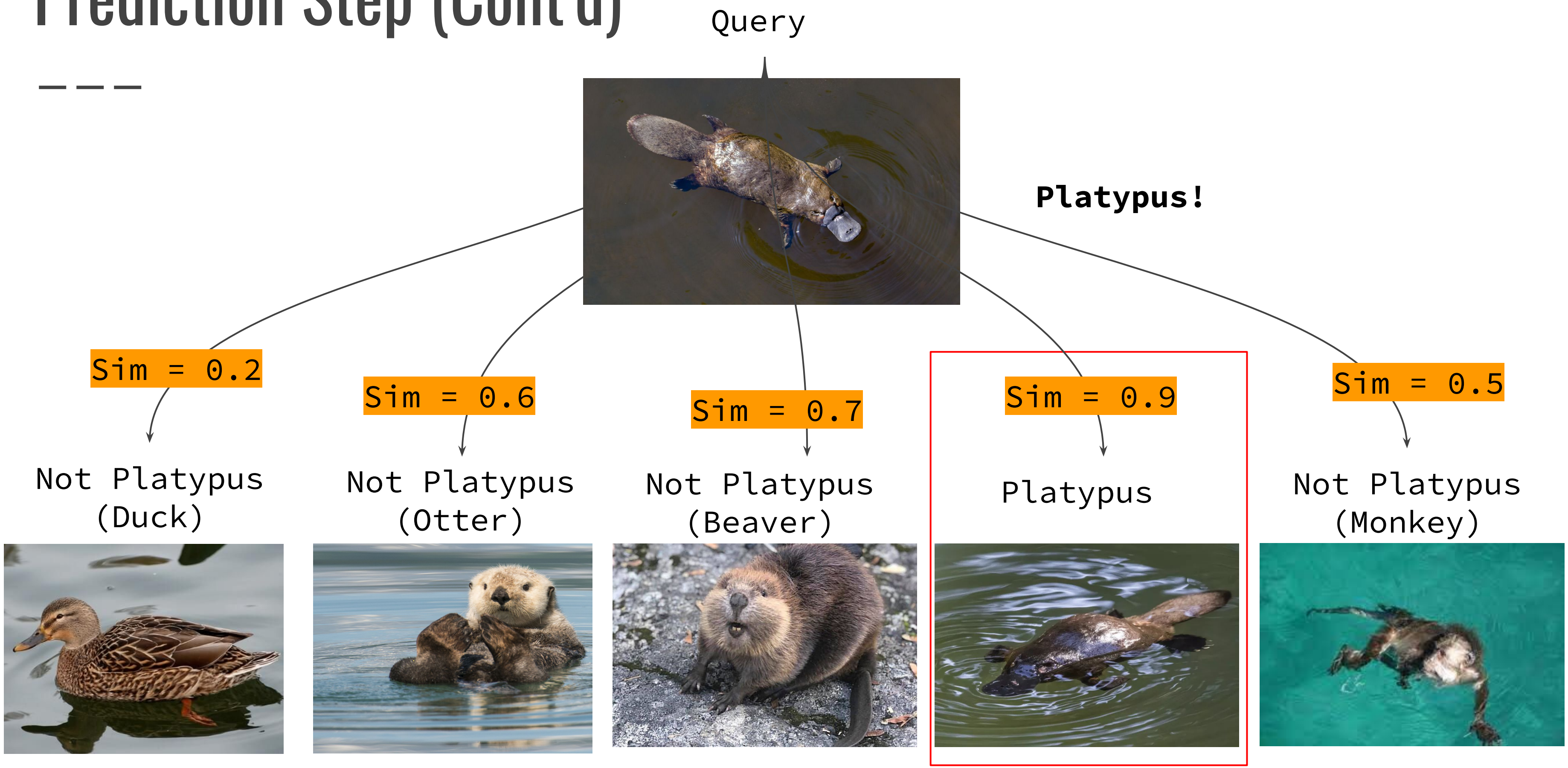
Prediction Step (Cont'd)



Prediction Step (Cont'd)



Prediction Step (Cont'd)

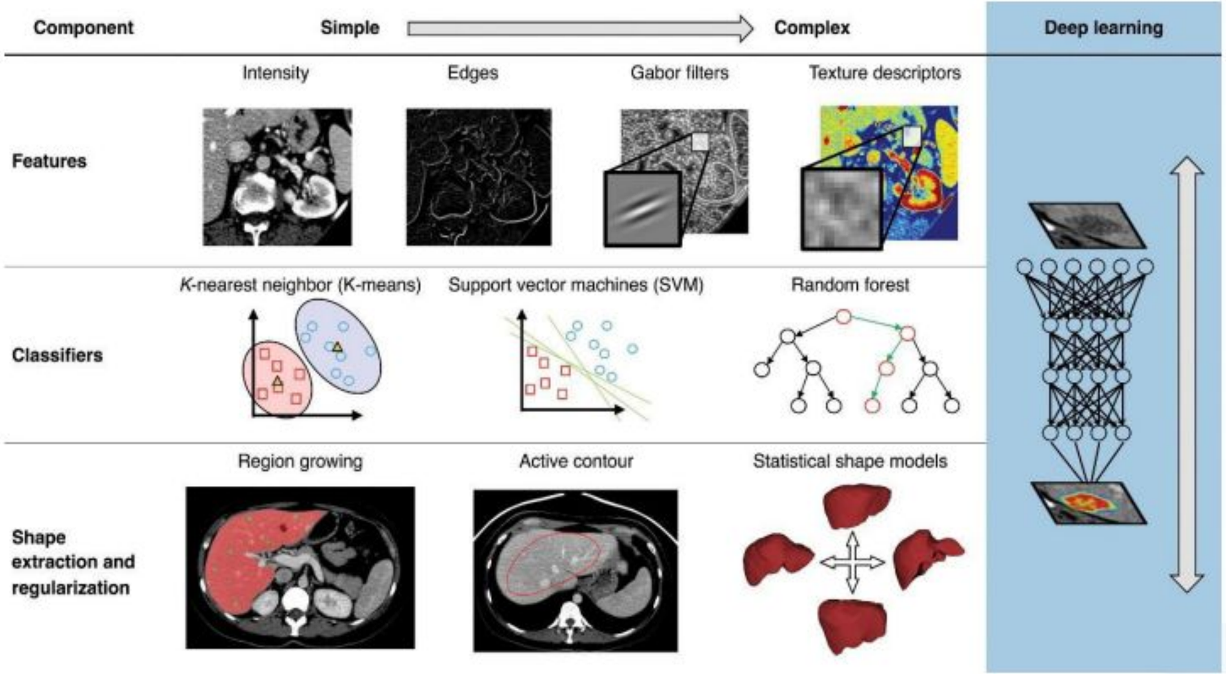


Discussion

We have seen the main properties of each model. Now, we compare the performance of them in different aspects.

- Feature extraction
- Classifier

Traditional ML vs DL



Chartrand et al. RadioGraphics 2017

Discussion: Features

Traditional method:

- SIFT is scale and orientation invariant and it can get representative patterns
- K-means is easy to implement
- Hard to define a k value for k-clusters
- Clustering is not robust with outliers
- More images are necessary to get a well performance

Modern method:

- CNN automatically detects the important features without any human supervision.
- Deep learning techniques for feature extraction are robust to scale, occlusion, deformation, rotation
- Hard to understand the blackbox. Unable to visualize the features.

Source: https://towardsdatascience.com/a-friendly-introduction-to-siamese-networks-85ab17522942

Discussion: Classifiers

Traditional method:

- SVM is effective in high dimension space
- SVM is good for binary classification
- SVM requires more training time when number of data is high

Modern method:

- Siamese Networks work well in high dimension space
- More robust to class imbalance
- Long training time

Source: https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107

Discussion: Accuracy (%)

Year	Animal	Methodology	Top-1 accuracy (%)
2013	Manta Ray	SIFT	51.0
2013	Chimpanzee (C-Zoo)	Support vector machine	84.0
2013	Chimpanzee (C-Tai)	Support vector machine	68.8
2018	Elephant	Support vector machine	59.0
2018	Chimpanzee	Siamese network	93.8
2018	Lemur	Siamese network	90.4
2018	Golden Monkey	Siamese network (videos)	75.8

Source: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13133

Prefer Deep Learning When:

- Have a lot of computing power (CPU, GPU, TPU, etc.) to allow intensive model training and good app performance.
- Uncertainty about the positive feature-engineering outcome
- Only high-performance devices are allowed to be deployed

Traditional Method When:

- Inadequate storage and processing power.
- A less expensive solution is desired.
- Want to be able to deploy on a variety of hardware.

Source: https://roboticsbiz.com/deep-learning-vs-traditional-image-processing-a-comparison/

Thank You!