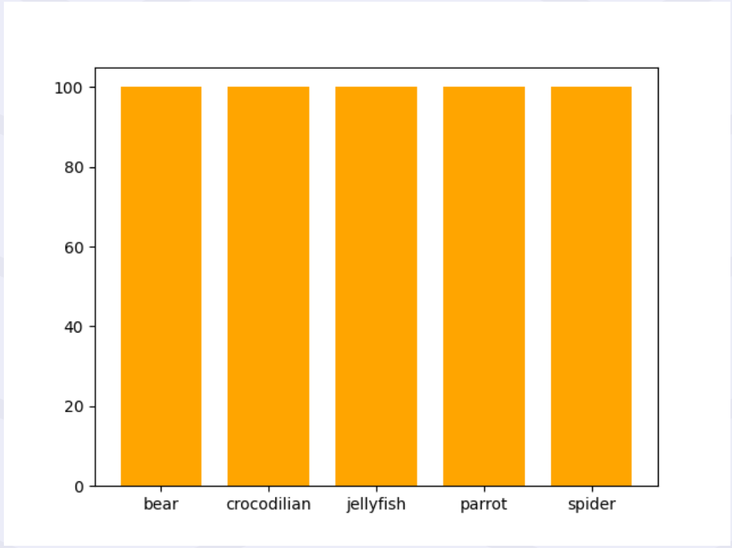
SketchZoo: Animal Image Retrieval with Siamese Neural Networks

Davide Brescia

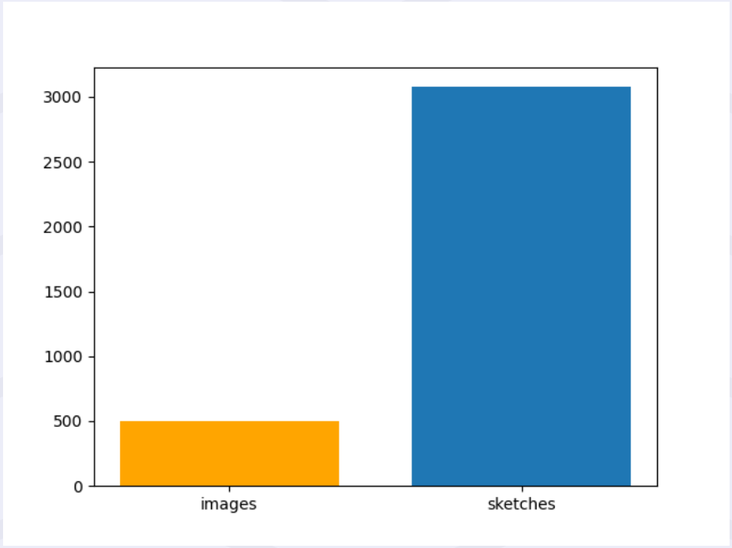
MINI-DATASET DISTRIBUTION

To ensure the initial functionality of the system, I opted to work with a small subset of the complete dataset. The subset consists of **5 Classes** (bear, crocodilian, jellyfish, parrot, spider) and comprises a total of **500 Images** and **3076 Sketches**.

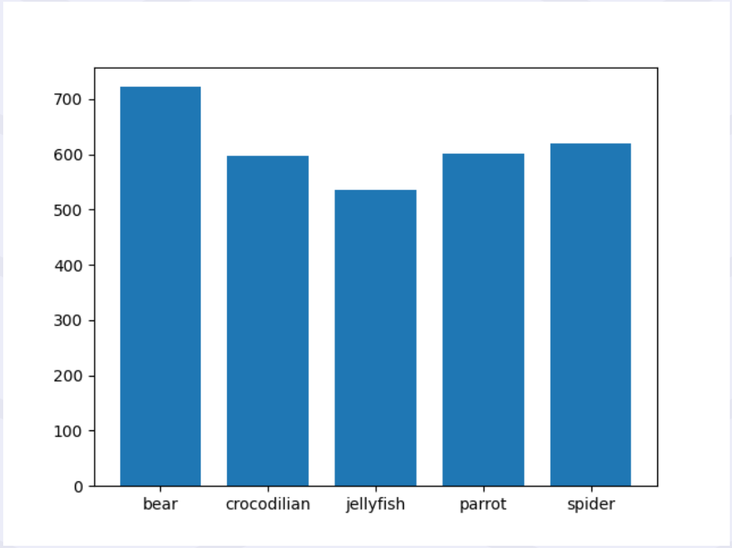
Images Distributions



Images vs Sketches



Sketches Distribution



Source: The Sketchy Database: Learning to Retrieve Badly Drawn Bunnies

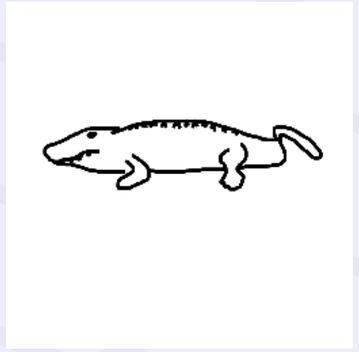
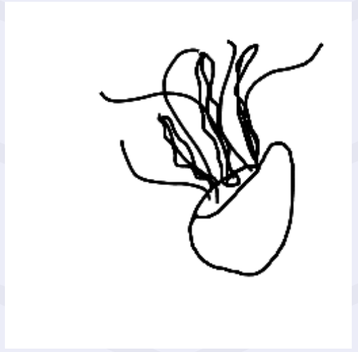
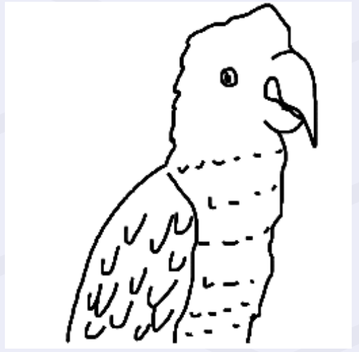
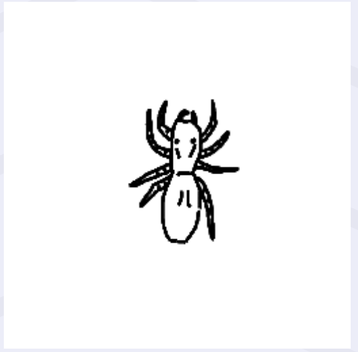
MINI-DATASET EXPLORATION

- Some Images -





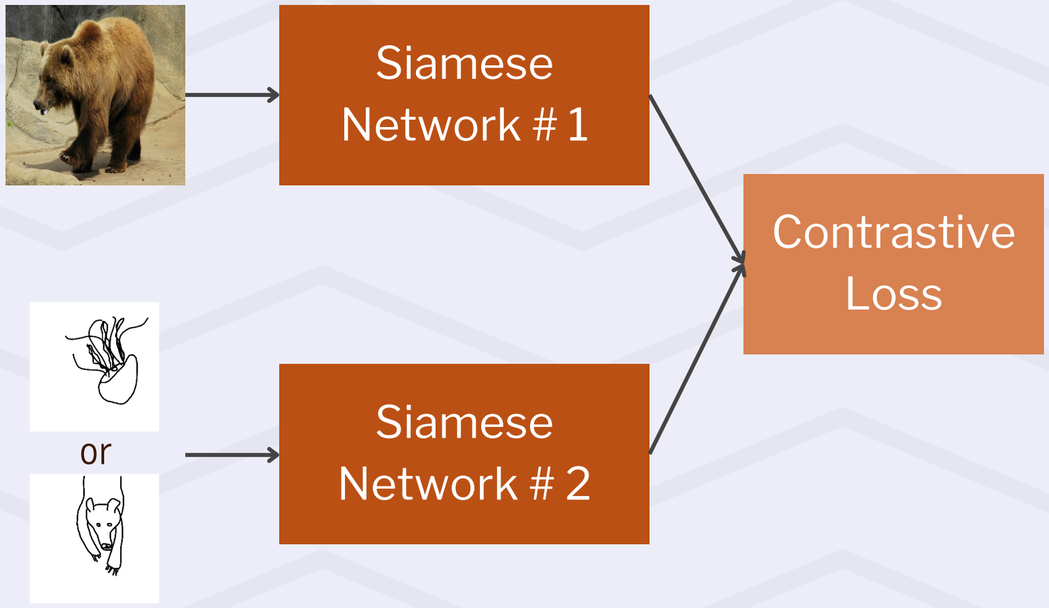
- Some Sketches -

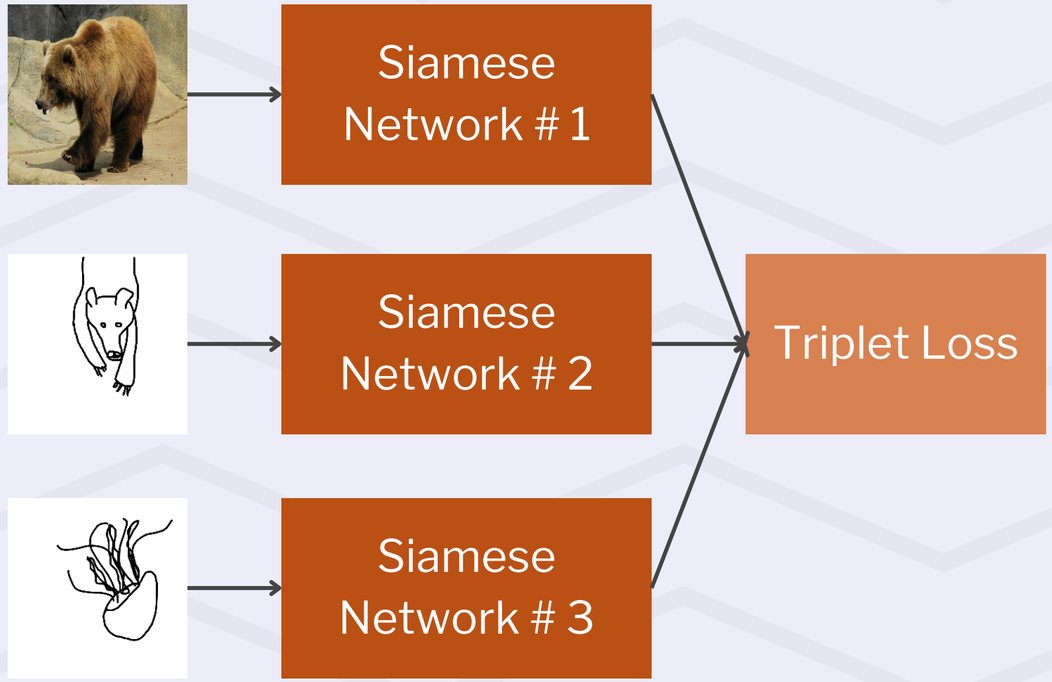






Source: The Sketchy Database: Learning to Retrieve Badly Drawn Bunnies

PROBLEM APPROACH

To address this issue, the **Siamese Neural Network** was employed, which involves the use of **two** (in the case of contrastive loss) **or three** (in the case of triplet loss) **neural networks** with shared weights.
The embeddings are then computed and compared to obtain the loss.





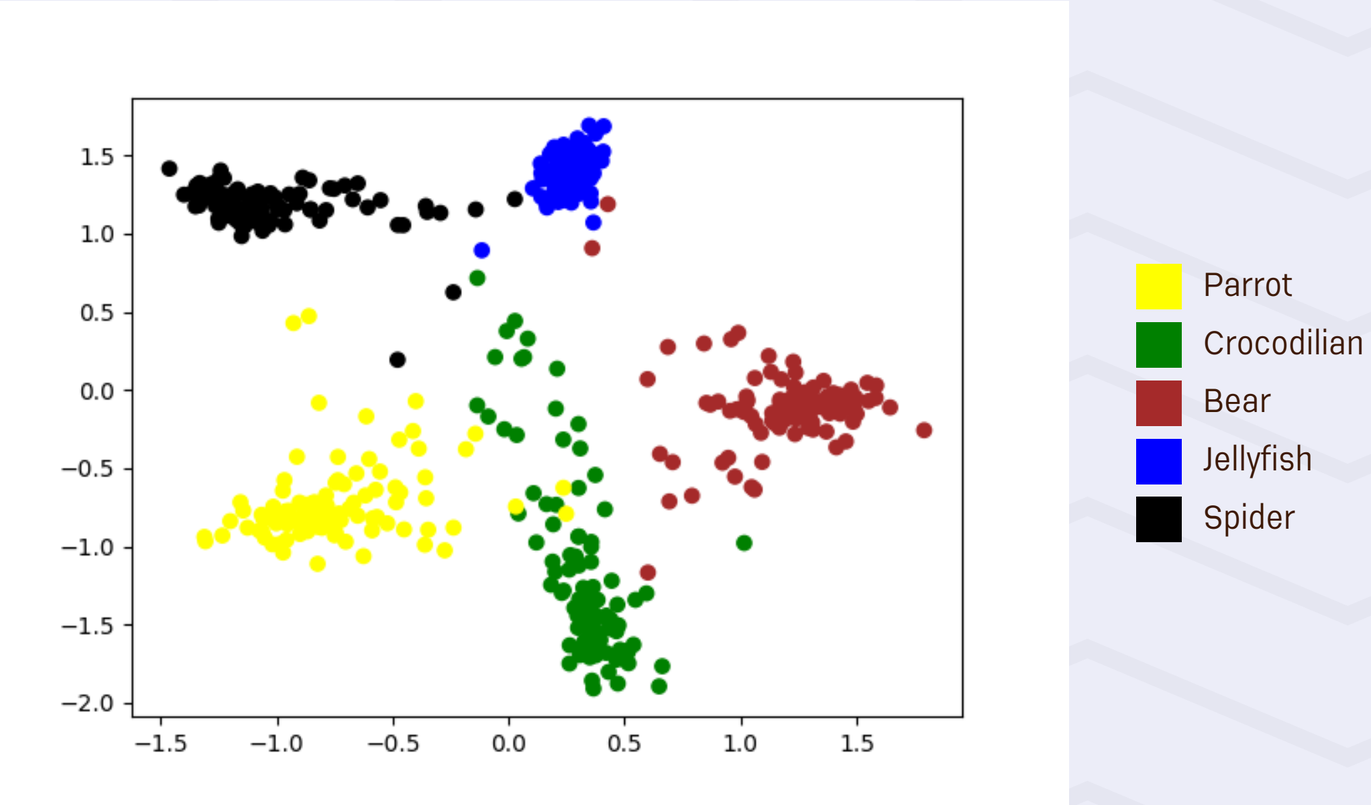
EXPERIMENT 1 – DETAILS

Name: Experiment 1

Backbone: ResNet18

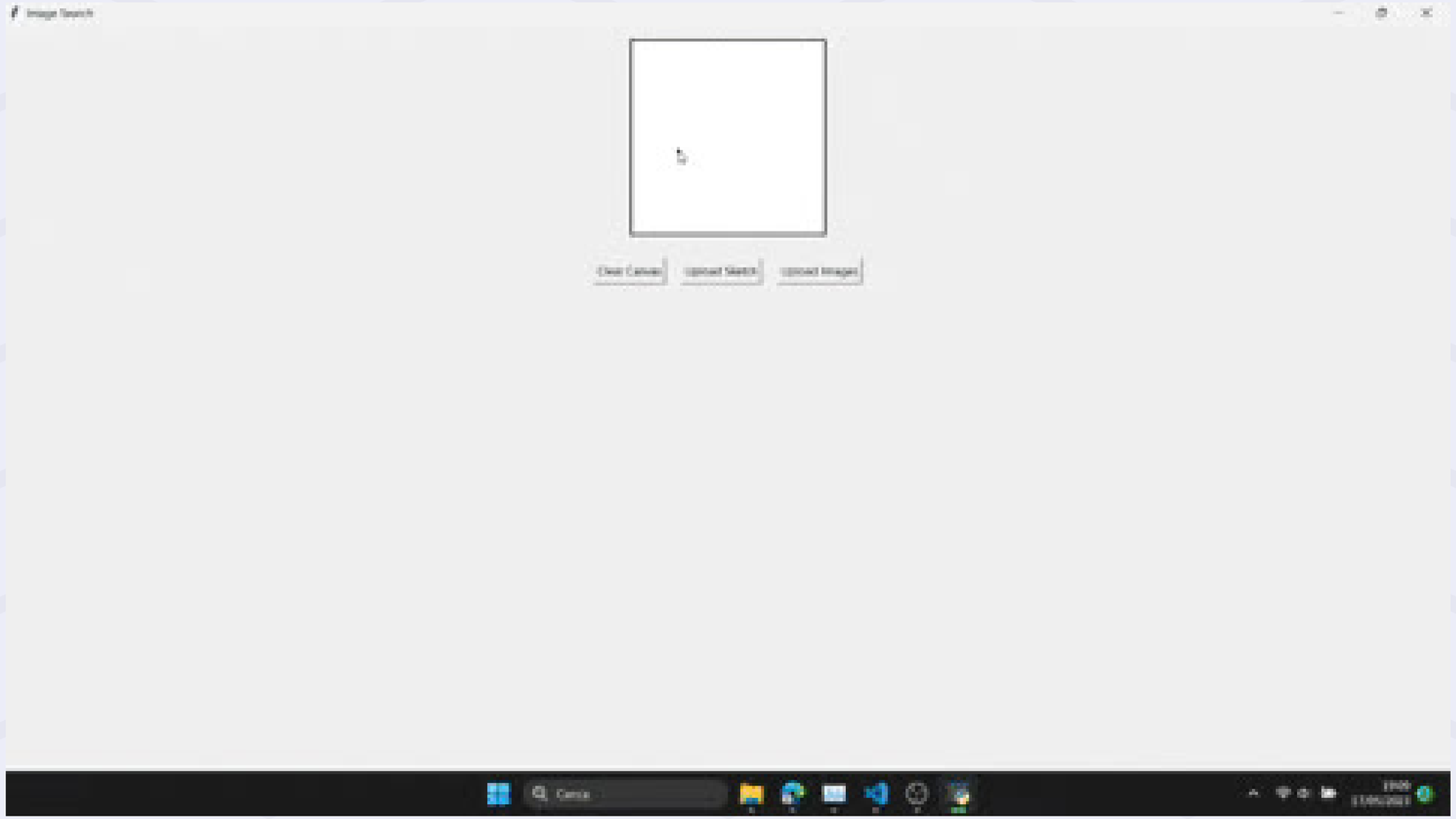
Loss: Constrastive Loss

Embedding Size: 2



EXPERIMENT 1 – RESULTS

As we can see the model works and has some … **interesting results**!



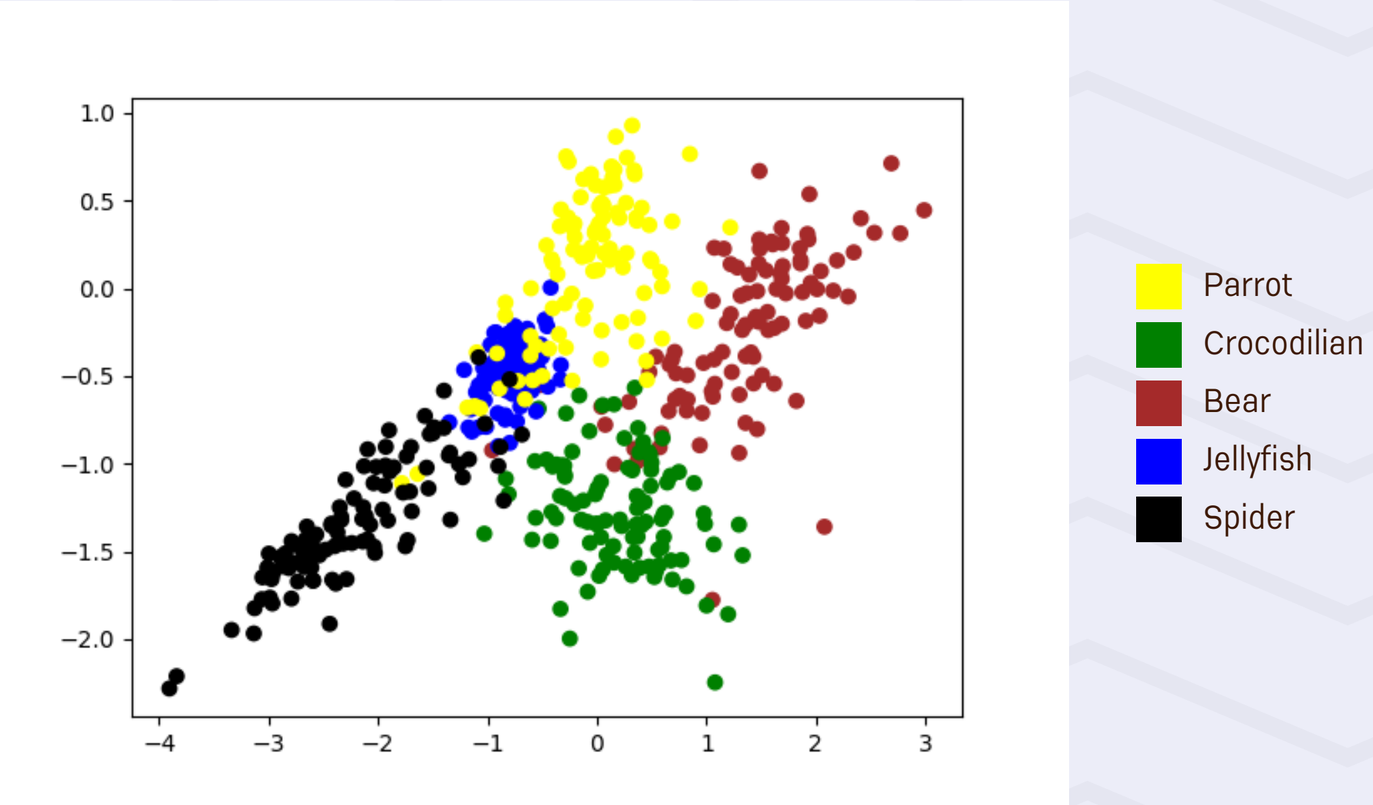
EXPERIMENT 2 - DETAILS

Name: Experiment 2

Backbone: ResNet18

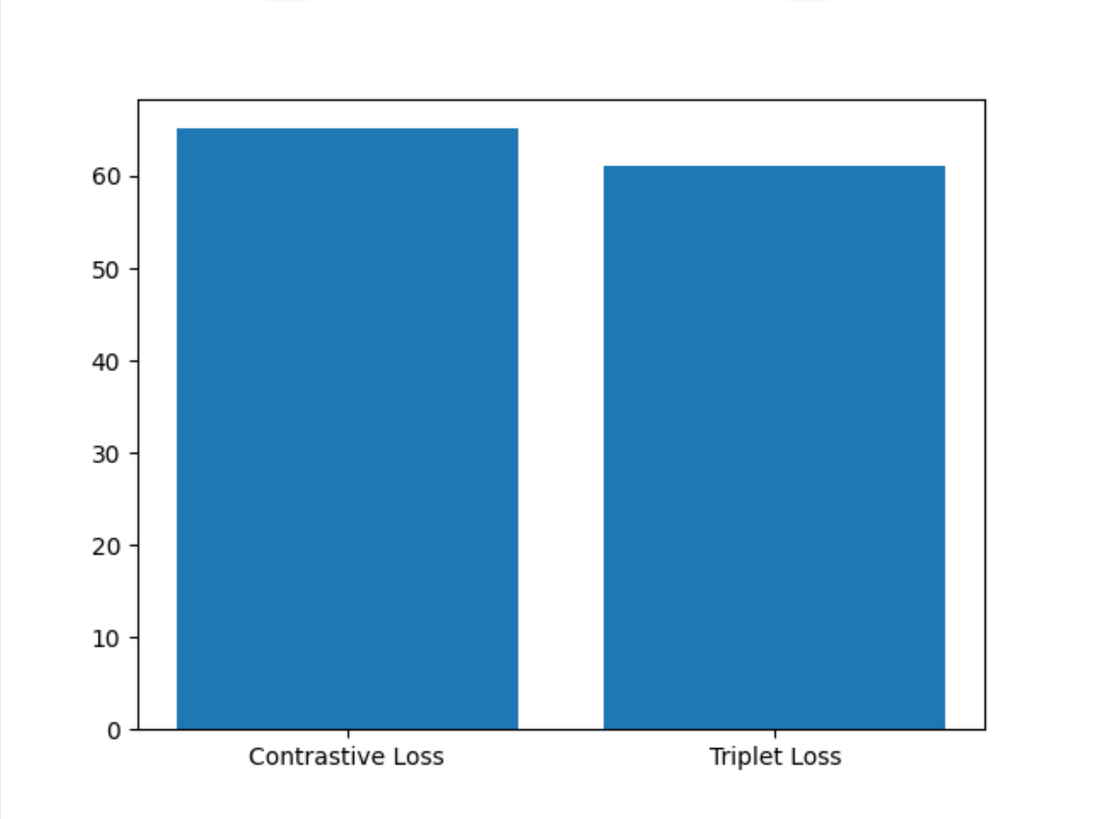
Loss: Triplet Loss

Embedding Size: 2



TRIPLET LOSS VS CONTRASTIVE LOSS

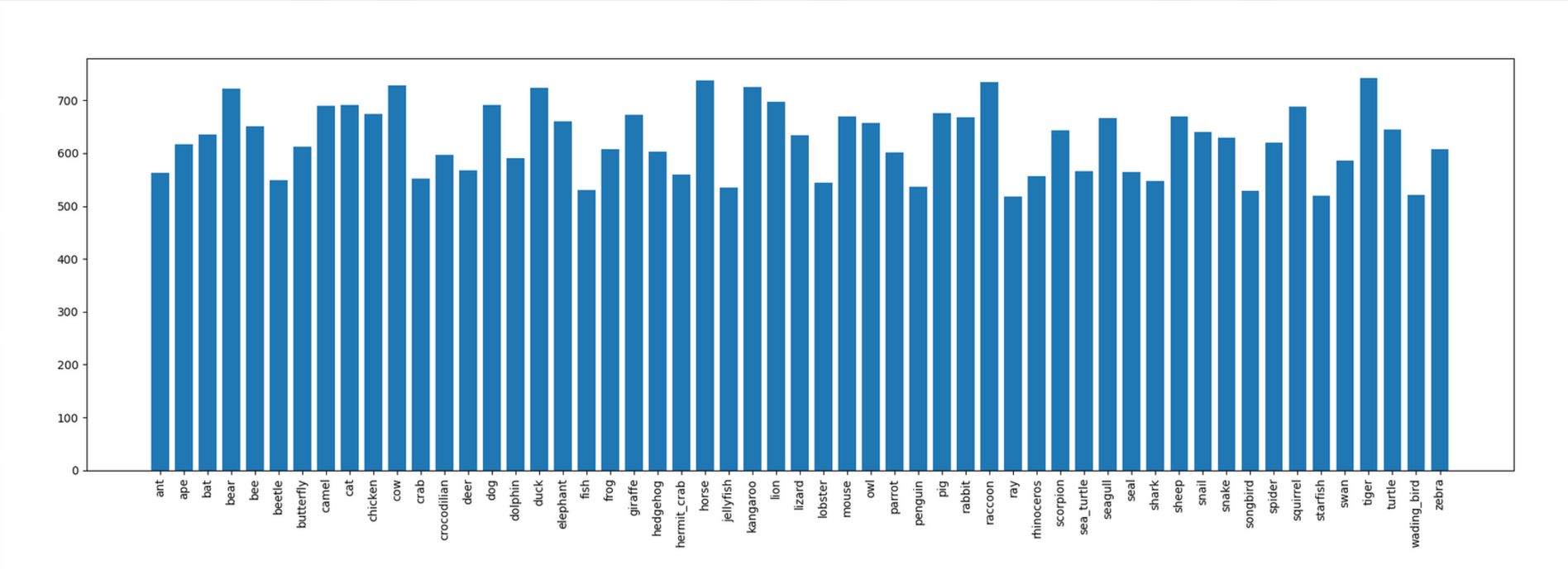
To conduct the evaluation, the **K–Precision** or **P@K** metric was employed, which quantifies the number of accurately identified classes among a set of K images (typically K=12) when applied to unseen images and sketches.



FULL-DATASET - DETAILS

It is now time to utilize a **larger subset**. Specifically, **55 classes** belonging to the animal kingdom domain have been selected, with **each class associated with 100 images** and a variable number of sketches, totaling **34,366 sketches**.

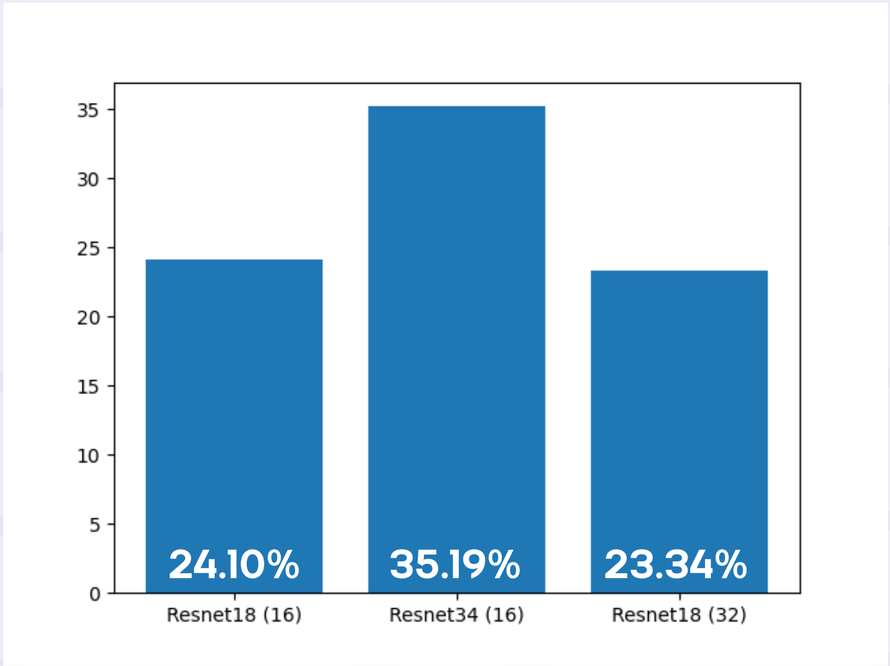
- Sketches Distribution -



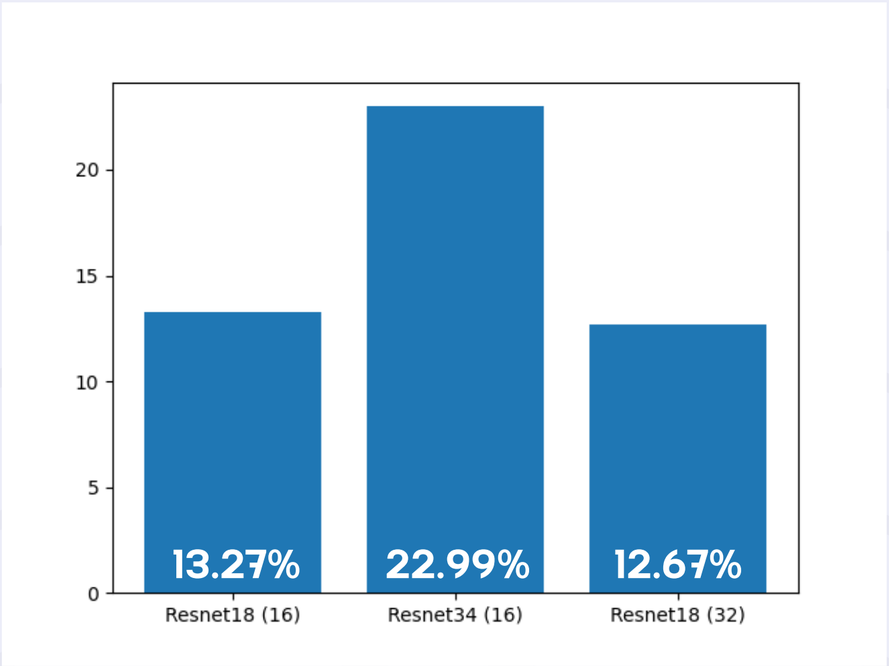
FULL-DATASET - RESULTS COMPARISON

I conducted additional experiments by using more **complex models** and **increasing the embedding size**. As we can see from the significant increase in the lower graphs, transitioning from **ResNet18 to ResNet34 yields noticeable improvements**.

- K@12 Contrastive Loss -



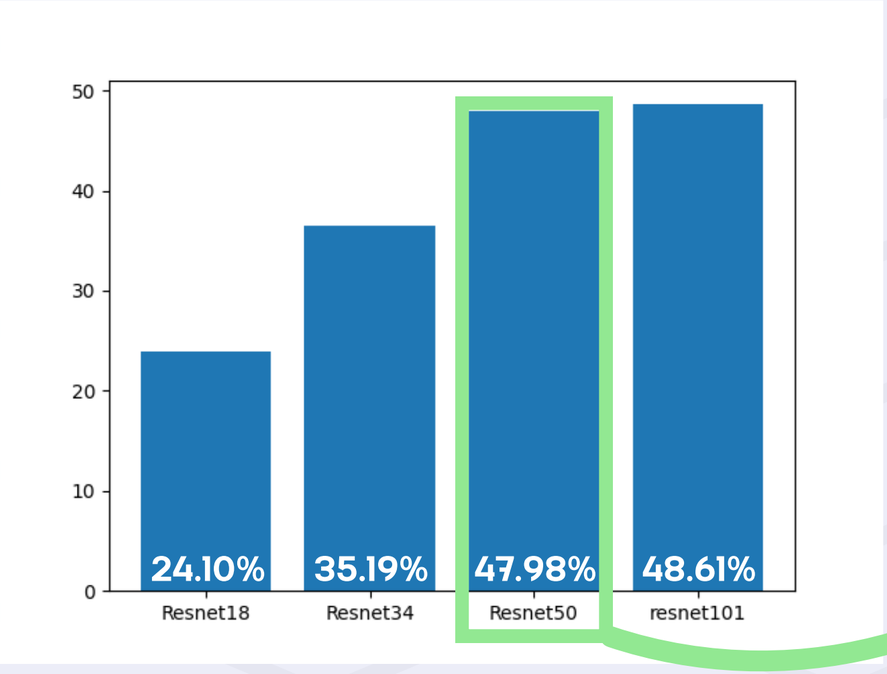
- K@12 Triplet Loss -



FULL-DATASET - MODEL IMPROVEMENT

Since the **Contrastive Loss** was leading to better results, I decided to focus solely on that and train **ResNet50** and **Resnet101** as well to see if the achieved results continued to improve. **Not surprisingly, the results are better**.

- Models Comparison in Contrastive Loss -



In the end, **I chose to implement this model**, which, despite having a lower K@12, has a significantly **shorter inference time**.

LIVE DEMO

THANKS FOR YOUR ATTENTION

Davide Brescia