CS221 FALL 2023 PROJECT

Find Similar Dogs

SU KARA, ISHVI MATHAI, UMUT EREN

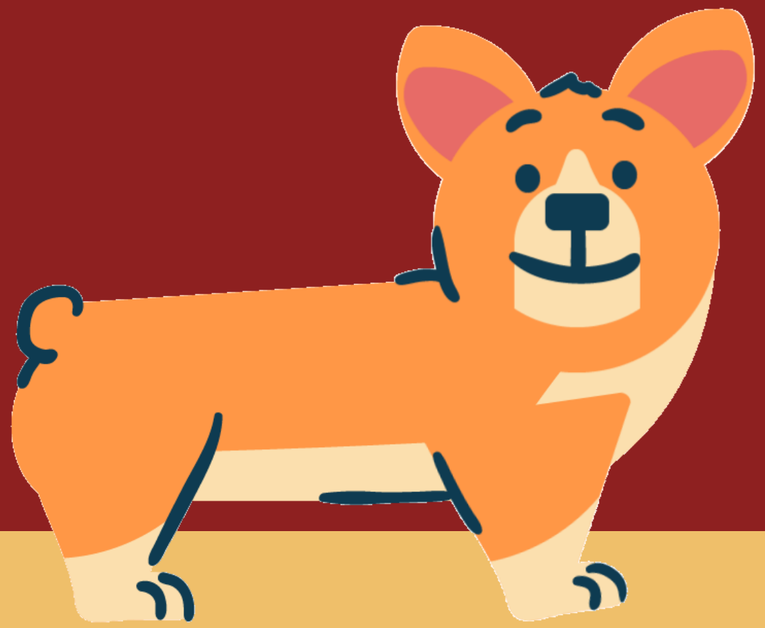
Introduction

- Visual Similarity Search is a useful technique in many domains
- Used by travel sites like Expedia to show similar hotels and destinations
- Used by shopping sites Amazon to show similar products
- Used by veterinarians to identify dog breeds

Literature Review

- So far, dog breed identification using AI has utilized transfer learning with pre-trained Convolutional Neural Networks (CNN). This approach is prominent in several studies that utilized the Stanford Dogs Dataset, which is a benchmark for assessing the performance of various AI models in breed classification.
- Our project, exploring the k-nearest neighbors (KNN) technique, diverges in its methodological approach.

Methods



Dataset

- The Stanford Dogs dataset contains 20,580 images of 120 breeds
- This dataset was built using images and annotations from ImageNet
- Each annotation XML file contains image width, height, and bounding boxes for one or more dogs in an image

Image Preprocessing

Cropped and resized each bounding box to generate 22,125 dog images with a width and height of 256x256

Google Colab

- Used T4 GPU that was 100x faster than CPU
- Developed with Python, Keras, and TensorFlow

Baseline Approach

- Random guessing strategy that picks 1 of the 120 breeds in a dataset and returns an image of that breed.
- Used Top-1 accuracy and Top-5 accuracy as evaluation metrics.

Transfer Learning

- Keras Applications offers pre-built models trained on ImageNet
- Used 8 different models to generate feature vectors for all images

K-Nearest Neighbors

- Used KNeighborsClassifier to find the 5 nearest neighbors
- Tested on 1,000 random images to find 5 similar dogs

Evaluation Metrics

Top-1 Accuracy

Top-1 accuracy is calculated as the proportion of instances where the breed of the most similar dog image matches the breed of the input image. This is measured over the total number of trials.

$$\text{Top-1 Accuracy} = \frac{\textit{Number of correct first predictions}}{\textit{Total number of predictions}}$$

Top-5 Accuracy

Top-5 accuracy is calculated as the proportion of instances where at least one of the 5 most similiar dog images' breeds matches the breed of the input image. This is measured over the total number of trials.

$$\text{Top-5 Accuracy} = \frac{\textit{Number of trials with the correct breed in top 5 predictions}}{\textit{Total number of trials}}$$

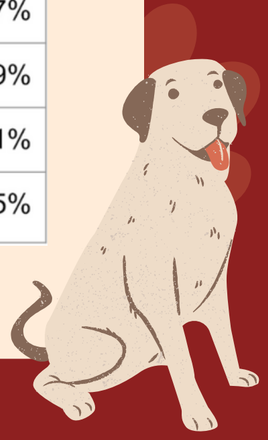
Main Approach

Pre-trained Models

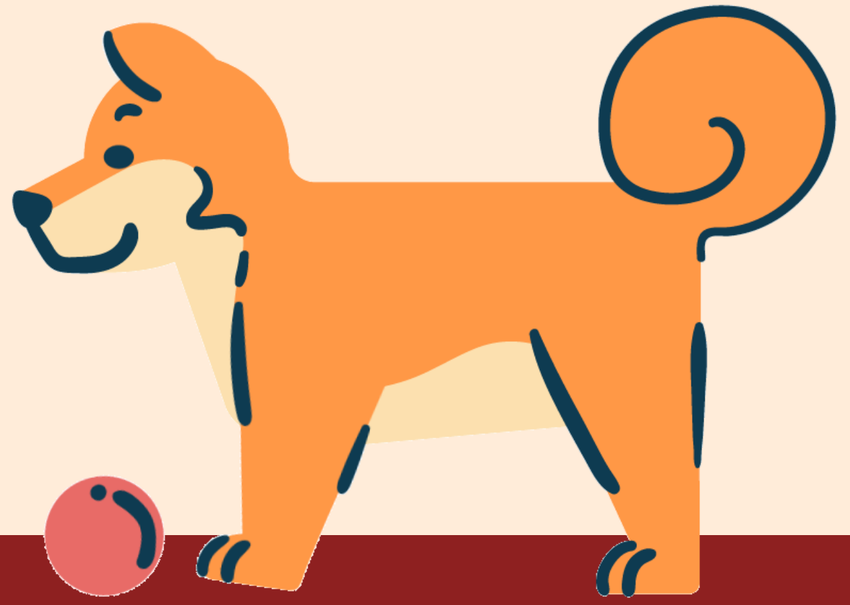
Model	Size (MB)	Parameters (M)	Depth	Feature Vector
NASNetLarge	343	88.9	533	4032
InceptionResNetV2	215	55.9	449	1536
Xception	88	22.9	81	2048
ResNet50V2	98	25.6	103	2048
DenseNet201	80	20.2	402	1920
VGG16	528	138.4	16	512
ConvNeXtBase	339	88.5	-	1024
EfficientNetV2M	220	54.4	-	1280

Results

Model	Top-1 Accuracy	Top-5 Accuracy
NASNetLarge	78%	90%
InceptionResNetV2	75%	89%
Xception	74%	92%
ResNet50V2	67%	89%
DenseNet201	66%	87%
VGG16	27%	49%
ConvNeXtBase	21%	41%
EfficientNetV2M	1%	5%



Fine-Tuning



Fine-Tuning Approach

- Divided the cropped images (22,125) into 3 datasets - training, validation and test - with the ratio of 80:10:10.
- First, checked accuracy with no training on the test set.
- Initial training by freezing all layers and adding a Dense fully connected layer, followed by a softmax layer for classification of the breed into 120 different categories, as specified in the dataset.
- Final fine-tuning phase involved unfreezing the last 10-20% of layers in the model but kept the first 90% frozen. Tested which number in the range of 10-20% performed the best, and recompiled the model.
- Chose to do the initial training for 5 epochs and fine-tuning for 2 epochs to avoid overfitting, which is seen by a drop in validation accuracy and an increase in validation loss.

Fine-Tuning Results

Model	Training	Top-1 Accuracy	Top-5 Accuracy
NASNetLarge	No Training	74%	90%
NASNetLarge	Initial Training (freeze=533, lr=1e-4, epochs=5)	78%	88%
NASNetLarge	Initial Training and Fine-Tuning (freeze=480, lr=1e-5, epochs=2)	82%	90%

Error Analysis

When does our model struggle?

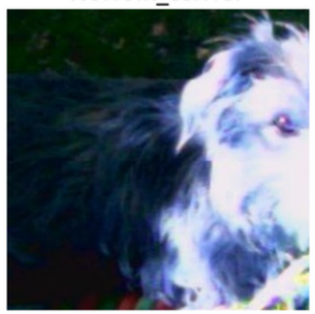
1) When the input dog breed looks very similar to another breed

Input Breed:
Old_English_sheepdog

Similar Breed #1:
Tibetan_terrier

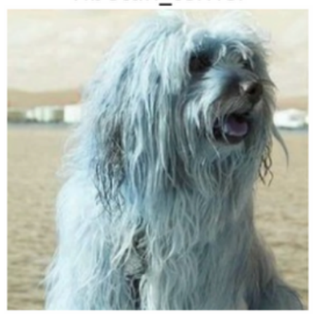
Similar Breed #2:
Tibetan_terrier

Similar Breed #3:
Norfolk_terrier



Similar Breed #4:
Australian_terrier

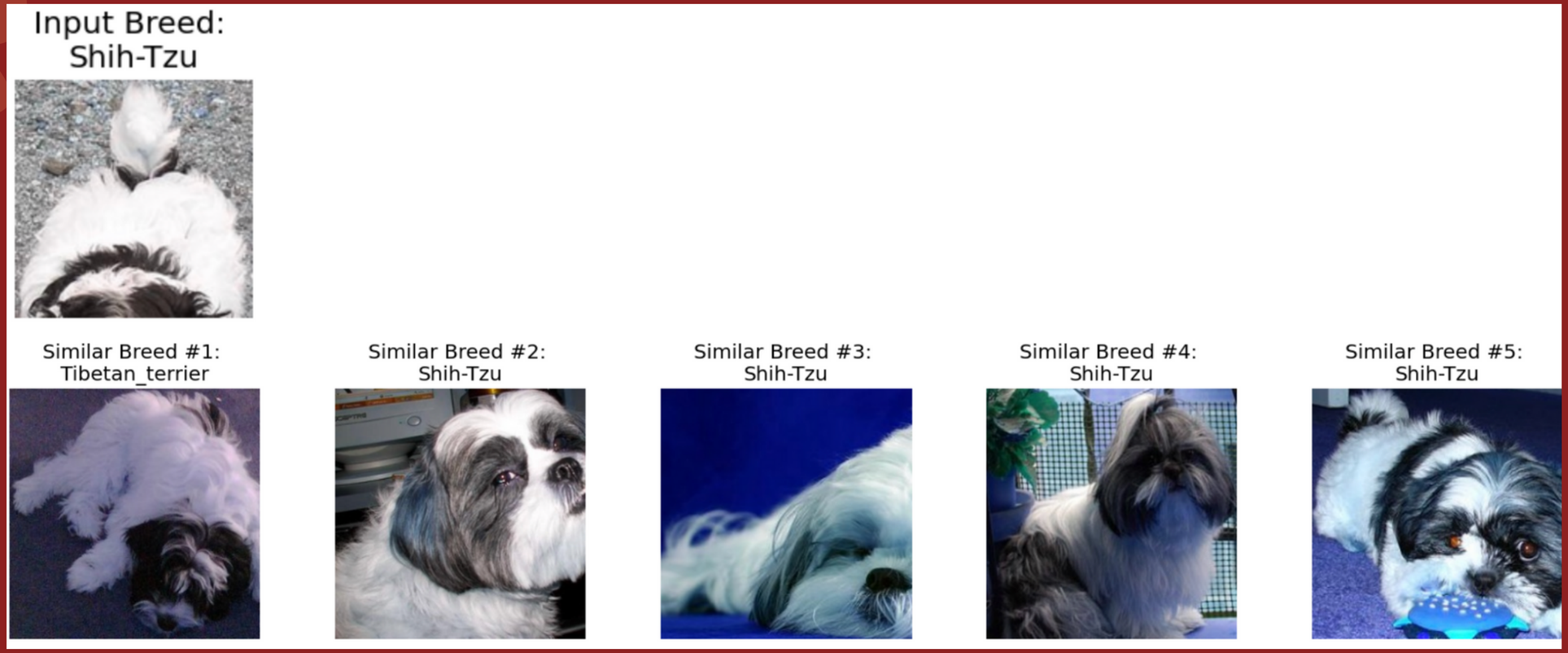
Similar Breed #5:
Tibetan_terrier



Error Analysis

When does our model struggle?

2) When the cropping algorithm is unable to picture the entire dog



Future Work

- Exploring vision transformers - currently considered the oracle method for finding similar dog images. The best performing ViT currently, ViT-NeT, has a Top-1 accuracy of 93.6%. Will attempt to reproduce these results and experiment with fine-tuning on this pre-trained model.
- Developing an app for veterinarians to use to identify dog breeds and prescribe medications accordingly.