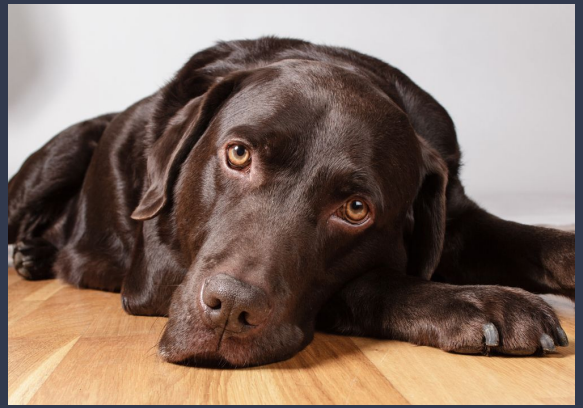
Who Let the Dogs Out?

Project 6: Brandon Griffin



Problem Statement

This project aims to automate the task of object detection, and specifically classify dogs.

Missing Dog Detection:
SF Animal Care & Control
Lost Missing Dogs San Francisco

What does an Atomic Physicist's dog do with bones?

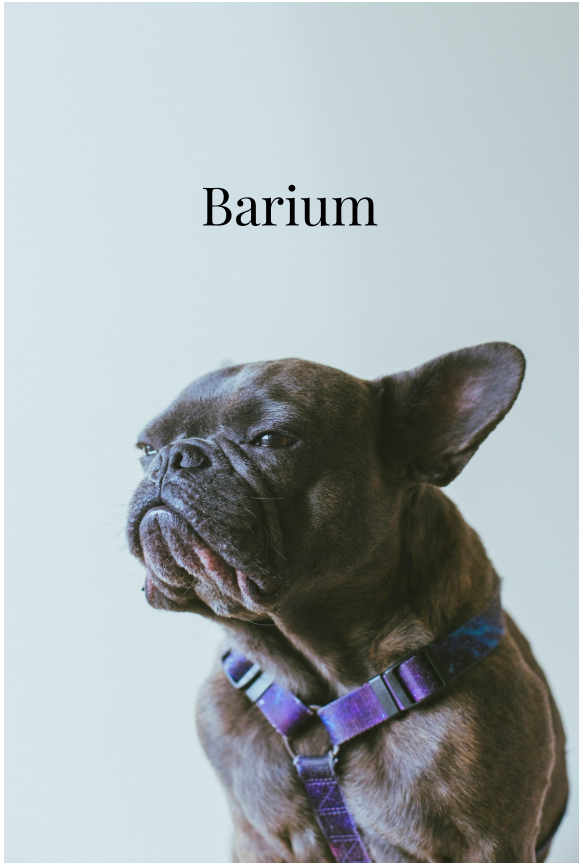


Photo by Channey on Unsplash

Is it a dog...?

@teenybiscuit/Twitter

Base Model

- ImageNet
  - 14 million images
  - 20,000 classes
- ILSVRC
  - 1 million images
  - 1,000 classes
  - 1,000 images
- VGG-16
  - Multiclass Target

VGG-16 Block	Name (Type)	Kernel Size	Nodes	Params #	Stride/Pool	Output (h x w x depth)
00-First	input1 (Input)	No Filter	None	0	None	(Batch, 224, 224, 3-RGB)
01-Block 01	conv1 (Conv2D)	(3 x 3)	64	1,792	(1 x 1)	(Batch, 224, 224, 64)
02-Block 01	conv2 (Conv2D)	(3 x 3)	64	36,928	(1 x 1)	(Batch, 224, 224, 64)
03-Block 01	pool1 (MaxPooling2D)	(2 x 2)	None	0	(2 x 2)	(Batch, 112, 112, 64)
04-Block 02	conv1 (Conv2D)	(3 x 3)	128	73,856	(1 x 1)	(Batch, 112, 112, 128)
05-Block 02	conv2 (Conv2D)	(3 x 3)	128	147,584	(1 x 1)	(Batch, 112, 112, 128)
06-Block 02	pool2 (MaxPooling2D)	(2 x 2)	None	0	(2 x 2)	(Batch, 56, 56, 128)
07-Block 03	conv1 (Conv2D)	(3 x 3)	256	295,168	(1 x 1)	(Batch, 56, 56, 256)
08-Block 03	conv2 (Conv2D)	(3 x 3)	256	590,080	(1 x 1)	(Batch, 56, 56, 256)
09-Block 03	conv3 (Conv2D)	(3 x 3)	256	590,080	(1 x 1)	(Batch, 56, 56, 256)
10-Block 03	pool3 (MaxPooling2D)	(2 x 2)	None	0	(2 x 2)	(Batch, 28, 28, 256)
11-Block 04	conv1 (Conv2D)	(3 x 3)	512	1,180,160	(1 x 1)	(Batch, 28, 28, 512)
12-Block 04	conv2 (Conv2D)	(3 x 3)	512	2,359,808	(1 x 1)	(Batch, 28, 28, 512)
13-Block 04	conv3 (Conv2D)	(3 x 3)	512	2,359,808	(1 x 1)	(Batch, 28, 28, 512)
14-Block 04	pool4 (MaxPooling2D)	(2 x 2)	None	0	(2 x 2)	(Batch, 14, 14, 512)
15-Block 05	conv1 (Conv2D)	(3 x 3)	512	2,359,808	(1 x 1)	(Batch, 14, 14, 512)
16-Block 05	conv2 (Conv2D)	(3 x 3)	512	2,359,808	(1 x 1)	(Batch, 14, 14, 512)
17-Block 05	conv3 (Conv2D)	(3 x 3)	512	2,359,808	(1 x 1)	(Batch, 14, 14, 512)
18-Block 05	pool5 (MaxPooling2D)	(2 x 2)	None	0	(2 x 2)	(Batch, 7, 7, 512)
19 4D --> 2D	flatten (Flatten)	No Filter	None	0	None	(Batch, 25,088)
20-Fully Connected	fcon1 (Dense)	No Filter	4,096	102,764,544	None	(Batch, 4,096)
21-Fully Connected	fcon2 (Dense)	No Filter	4,096	16,781,312	None	(Batch, 4,096)
22-Last Layer	Output (Dense)	No Filter	1,000	4,097,000	None	(Batch, 1,000)

Visual Genome

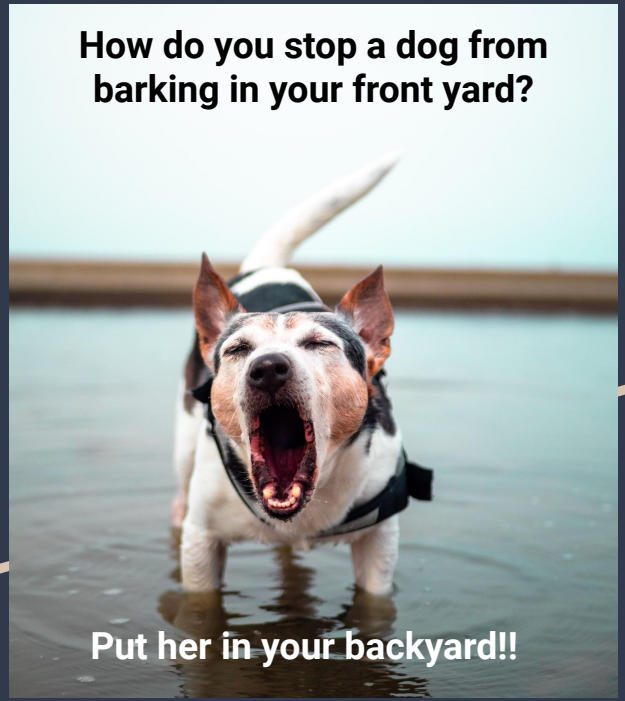
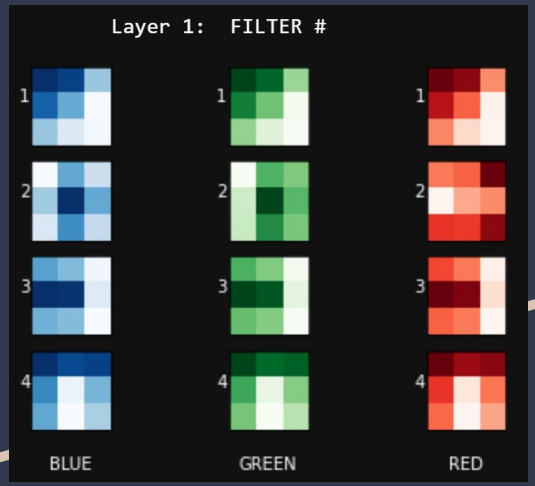


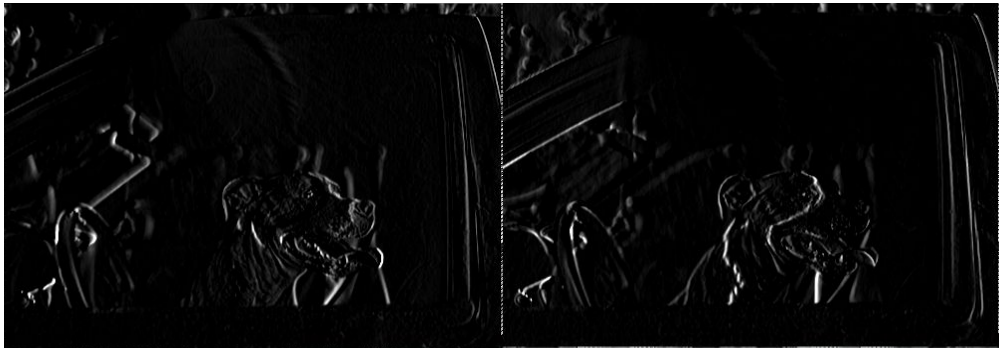
Photo by Angelo CARNIATO on Unsplash

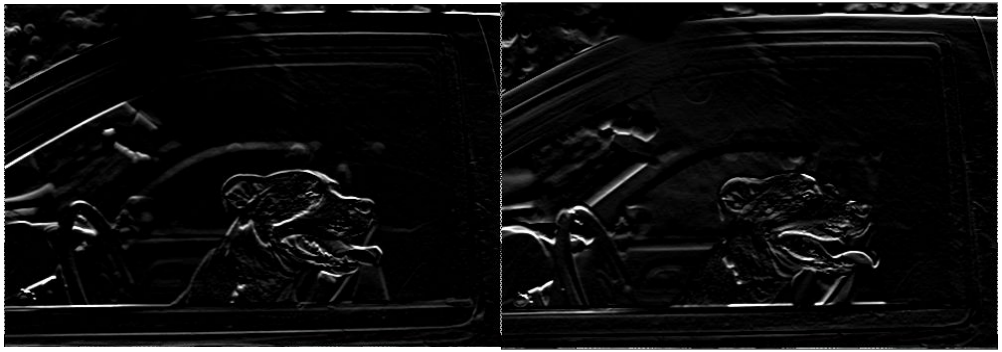
- Vision:
  - Complete understanding of visual scenes
- 108K + images
  - Not professional photography
- Average per image:
  - 35 objects
    - Breeds of dogs not labeled
  - 26 attributes
  - 21 pairwise relationships

Filters: 64 x (3 x 3 x 3)

1st Conv2D Layer







Model Performance:

- Loss: 0.0995
- Accuracy: 97.52%

- 10 epochs
- Best epoch: 2
- No augmentation

Dog Images	Training Set	Validation Set
3,235	1,995 Images	1,240 Images
3%	62%	38%

Next Steps

1. Batch Generator
   a. Augmentation as regularization
   b. Batch size flexibility

2. Consider similarity of images
   a. ImageNet vs Visual Genome

3. Label breeds in VG dataset
   a. Crowd source
   b. Build breed app

4. Language bias deteriorate CV task performance ex. Hot Dogs

References:

Photo by Clay Banks on Unsplash

Images:

https://barkpost.com/humor/doodle-or-fried-chicken-twitter/

Papers:

https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/

Very Deep Convolutional Networks For Large-Scale Image Recognition` (arXiv): https://arxiv.org/pdf/1409.1556.pdf