Deep Sea Image Classification

Cloud Computing

Jennifer Crosby
Benjamin Lee
Polly McKim

PMH

Problem Statement

Researchers currently spend thousands of hours sifting through countless photos to determine which ones have items of interest for their research. Can deep learning be used to solve this problem?

Can we find cute photos like this?

While filtering out photos like this?

Project Overview

- Acquired deep sea research expedition images from NOAA
- Manually classified images as interesting or not based on creature presence
- Built and deployed a machine learning model to learn to classify images based on their content labels
- Used AWS architecture to greatly increase speed and scalability of the model
- Delivered a model of over 75% accuracy
- Used a CNN (convolutional neural network) algorithm as our model

Sources Used

- NOAA repository of deep sea images
- CVision AI (http://cvisionai.com/) houses NOAA images on an online server
- 1500 images + additional created through image augmentation

PM

Libraries & Features Implemented

Python libraries to install:

Pandas, numpy , matplotlib, sklearn, pytorch, pytorchvision, warnings,

PIL, image_folder, random

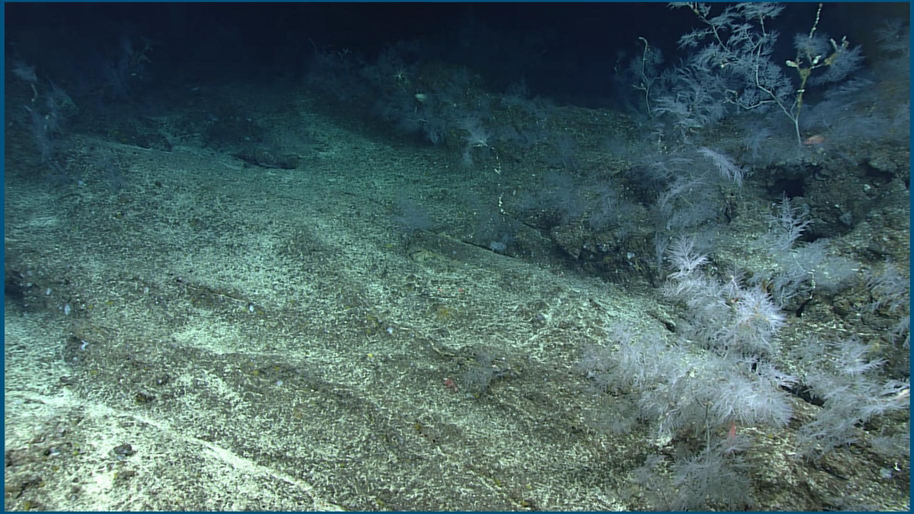
Features used in modeling:

CNN, Adam optimizer, back propagation

PM

Examples of Inputs

Not Interesting



Interesting

PM

Target file

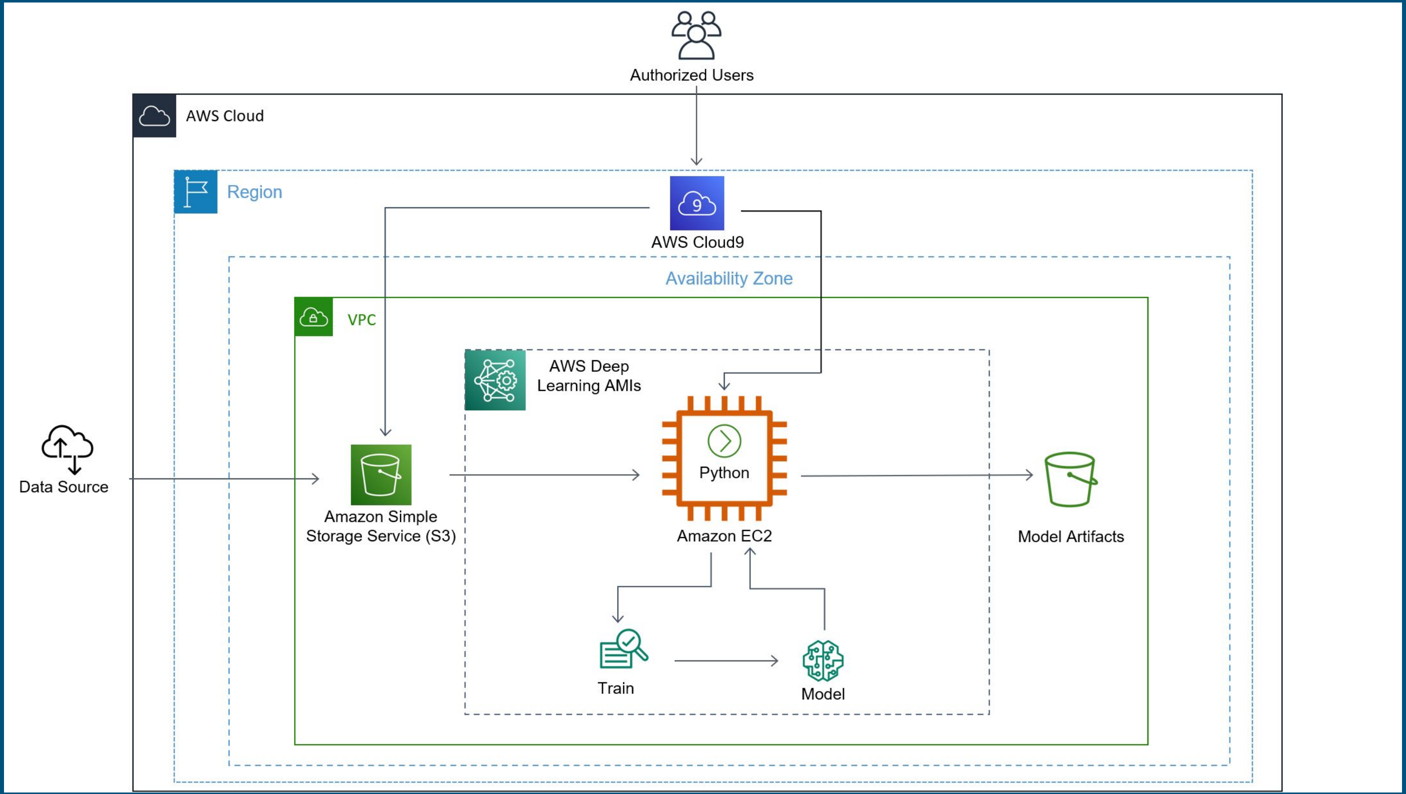
18	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_1378.png	interesting
19	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_2247.png	interesting
20	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_2277.png	interesting
21	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_3866.png	interesting
22	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_4555.png	interesting
23	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_4675.png	interesting
24	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_4885.png	interesting
25	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_5274.png	interesting
26	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_5604.png	interesting
27	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_6113.png	interesting
28	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_6323.png	interesting
29	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_7222.png	not_interesting
30	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_7462.png	not_interesting
31	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_7852.png	not_interesting
32	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_7882.png	not_interesting
33	56321614_EX1708_VID_20170926T212000Z_ROVHD.mp4_7972.png	not_interesting
34	56321615_EX1708_VID_20170926T200000Z_ROVHD.mp4_539.png	interesting
35	56321615_EX1708_VID_20170926T200000Z_ROVHD.mp4_629.png	interesting

The 1500 images were manually classified

AWS Architecture Used for Project

- EC2 (AMI Name Deep Learning Base AMI (Ubuntu 18.04) Version 31.0); Type: g3.4xlarge
- VPC
- AWS Cloud9- Cloud IDE

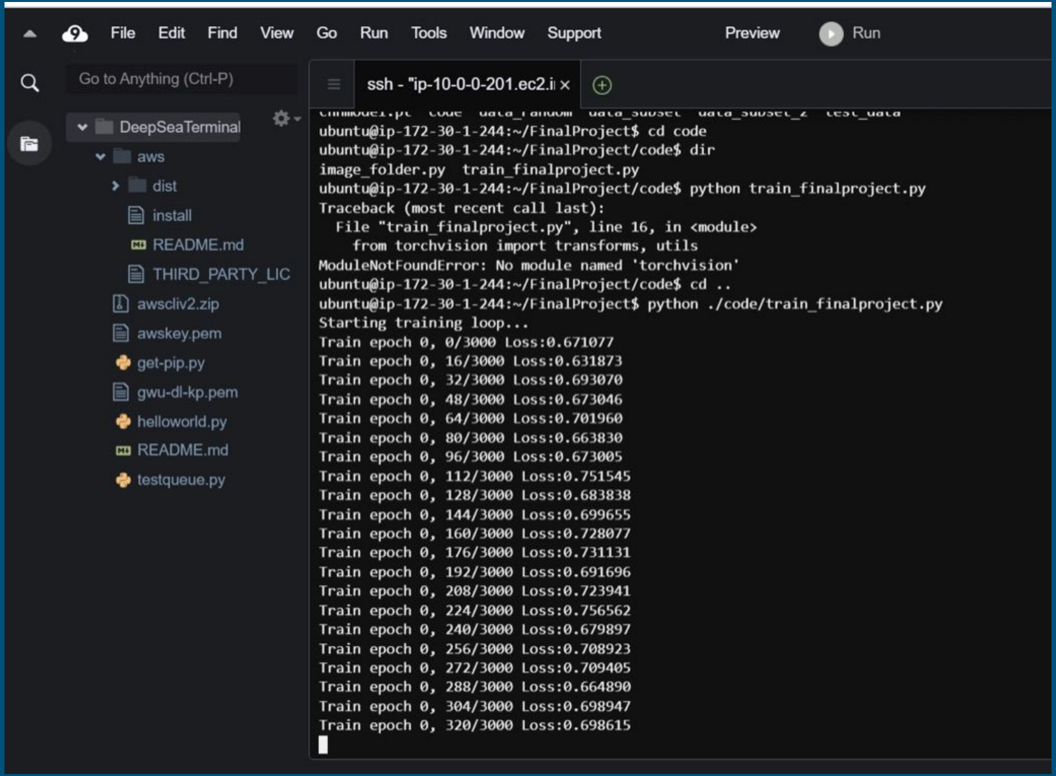
Architecture for Deep Learning EC2 Deployment



Cloud Architecture Demo

JKC

Python Running the Model on the Deep Learning EC2



JKC

76%

Overall Accuracy Achieved

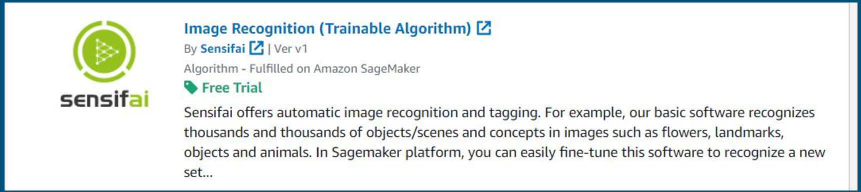
```
Validation Loss 0.67914
Validation Accuracy: 76.0
ubuntu@ip-172-30-1-244:~/FinalProject$
```

EC2 Deep Learning Findings & Cautionary Tales

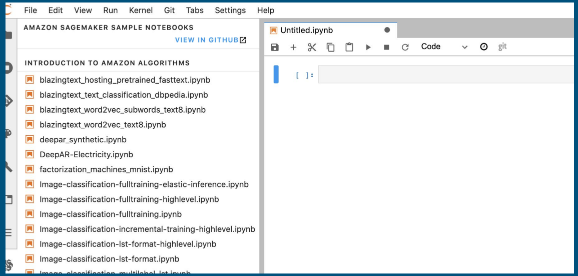
- G8, Ran much faster than on a typical personal computer. Ideal P3.
- Auto Scalability or using more GPUs would likely have improved the run time even more but would have added to the cost of running the model. Need to make sure have limits available to implement.
- Understanding of EC2 Ubuntu choices for deep learning is important (ie make sure the family you select has the capacity and features you need)
- GPUs do not exist in all availability zones or on every option of deep learning Ubuntu (Don't select a T when you really need a G or a P!).
- In the future, the process would be simplified by figuring out how to feed the data directly from the S3 to the running python program on the EC2

Sagemaker

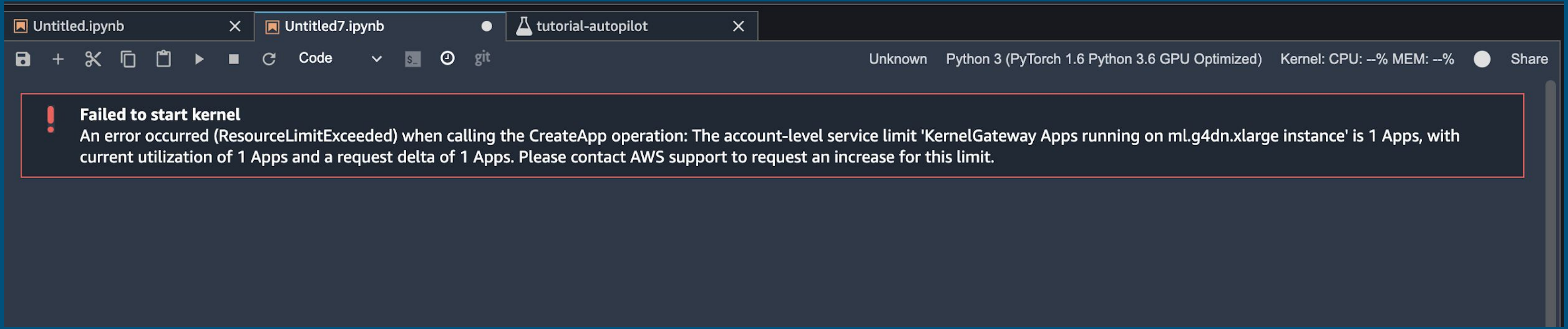
Sagemaker Marketplace



Prefab Notebooks



Instance Limit Issues



PM & JKC

Sagemaker Findings & Cautionary Tales

- Sagemaker notebooks/instances, like much of AWS, have predetermined blocks on certain kinds of notebooks so people don't inadvertently run up a large bill
- AWS is very slow to respond to requests to raise these limits
- While Sagemaker worked well for smaller tutorial type models that we did to learn about the platform, we were unable to get adequate access to run a deep learning model within the sagemaker ecosystem

Amazon SageMaker

JKC

Questions?