Machine Learning Engineering Nanodegree

Capstone project – Image Classification

Domain Background:

Deep learning belongs to neural model paradigm in three paradigms of machine learning. It is the latest version of –[1] connectionism that has been going on since the 1950s. In connectionism, problems that were previously considered obstacles have been solved one by one, and it has recently become the most destructive technology field in artificial intelligence. [2]. In 1979, Kunihiko Fukushima first published Noncognition model applying neurophysiological theory to the artificial neural network. This model, inspired by Torsten Wiesel's award-winning Nobel Prize in neurophysiology, was the beginning of Convolutional Neural Network (CNN). CNN, which imitates human visual cognition processes, is inherently used in a unique way in the field of computer vision. Convolution technology, which shows excellent performance in extracting desired features from various types of data, is used in various fields such as image processing and speech recognition. The reason for using the convolution technique in image processing and signal analysis is to separate and extract features contained in signals such as original image or sound wave. So, I need to access via CNN to identify whether the monitored object is same as the predicted object. A brief description of the data set is provided below:

- The CIFAR-10 dataset (Canadian Institute for Advanced Research) is a collection of images that are commonly used to train machine learning and computer vision algorithms.
- It is one of the most widely used datasets for machine learning research. The CIFAR-10 dataset contains 60,000 32x32 color images in 10 different classes.
- The 10 different classes represent airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. There are 6,000 images of each class.
- Computer algorithms for recognizing objects in photos often learn by example. CIFAR-10 is a set of images that can be used to teach a computer how to recognize objects.
- Since the images in CIFAR-10 are low-resolution (32x32), this dataset can allow researchers to quickly try different algorithms to see what works. Various kinds of convolutional neural networks tend to be the best at recognizing the images in CIFAR-10.
- CIFAR-10 is a labeled subset of the 80 million tiny images dataset. When the dataset was created, students were paid to label all of the images.

Problem Statement:

- Human beings, when fully attentive, do quite well at identifying lane line objects. Computers are not inherently good at doing the same. However, humans have a disadvantage of not always being attentive and do not have the ability to go to inaccessible locations such as space, underwater, wildlife monitoring etc.
- While a computer has the ability to go inaccessible locations.
- As such, if we can train a computer to get as good as a human at detecting and monitoring objects, since it is already significantly better at paying attention full-time, the computer can take over this job from the human to gain details in inaccessible locations.

- From the computer's perspective, the problem is around detecting objects into right categories.
- CIFAR-10 are low-resolution (32x32), this dataset so with low resolution I will implement various algorithms such as the convolutional neural network, SVM Classifier etc.
- Since the data is classified into 10 classes and I'm going to label the data into one of the classes. This is a multi-class classification problem

Datasets and Inputs:

- CIFAR-10 dataset (Canadian Institute for Advanced Research)
- The CIFAR-10 dataset contains 60,000 32x32 color images in 10 different classes.
- The 10 different classes represent airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. There are 6,000 images of each class.

Solution Statement:

To solve the problem presented in the Problem Statement, I will create an image classifier that can identify the object by deep learning using Convolutional Neural Network. To do this, we need to train CNN by inputting data from train.json file. The result will then be represented by a percentage value which is the percentage of the object

- Testing the model with different algorithms and measuring the efficiency in terms of mean square error and r square error of prediction.

Evaluation Metrics

- It will be based on the test data set and on different image classification models. Initially I would test it with the popular model such as CNN then compare the accuracy with other image classification models

Project Design

- Environment-macOS, Python 3.54, Anaconda 4.44, Jupyter Notebook, Keras, TensorFlow
- Workflow –
- Use torch vision to load the data. The dataset is split into three parts, training, validation, and testing. For the training, applied transformations such as random scaling, cropping, and flipping. This will help the network generalize leading to better performance. Also need to load in a mapping from category label to category name. Write inference for classification after training and testing the model. Then processed a PIL image for use in a PyTorch model.
- Try different approaches in algorithms and calculate its accuracy of prediction.

References:

[1] https://www.cs.toronto.edu/~kriz/cifar.html
[2] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
[3] https://cs231n.github.io/convolutional-networks/
[4] Connectionism. In Wikipedia. Retrieved January 15, 2018, from https://en.wikipedia.org/wiki/Connectionism