Machine Learning Engineer Nanodegree

Capstone Proposal

Gaofeng Bai
August 9st, 2019

Proposal

Domain Background

The dogs vs cats problem is proposed for my capstone project in the Nanodegree program. This is in computer vision domain. Computer vision was first proposed in the late 1960s, and it was used to describe an artificial intelligence computer that can mimic human visual system. Now, computer vision is an interdisciplinary problem that solves how computer can understand digital images or videos in higher level(Wikipedia). The technique is used broadly in many fields, like robotics, self-driving cars and image recognition, etc. The dogs vs cats problem is an image recognition problem. Image recognition has a very wide application, like visual search and face recognition on social media software, etc. The dogs vs cats is a Kaggle competitions that is proposed in 2013 (https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/overview). This project may help visual search, and automated image organization. It also can help participator have better understand of the knowledge of deep learning and image recognition(https://imagga.com/blog/the-top-5-uses-of-image-recognition/).

Problem Statement

The dogs vs cats is defined as an image classification problem. As mentioned in the requirement of the Udacity dogs vs cats project github readme, the deep learning algorithm is needed to be used to recognize if a dog or cat is in a target image. The input of the project is one colored image. The out put of the project is a result that tells if the image is a cat image or dog image. The test data is provided by Kaggle website. For each test image, the probability of the image is a dog will be submitted. The submissions will be scored based on the log loss function:

$$LogLoss = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log\left(\hat{y}_i\right) + \left(1-y_i\right)\log\left(1-\hat{y}_i\right)\right]$$

Equation 1

- *n is the number of images in the test set*
- $\hat{y}_i$ is the predicted probability of the image being a dog
- y_i is 1 if the image is a dog, 0 if cat
- *log*() is the natural (base e) logarithm

The score can be used to measure the performance of the program. To pass the capstone project the score should be higher than the first 10% of the score that is list in Kaggle Public Leaderboard.

Datasets and Inputs

The dataset is provided by Kaggle(https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data). It contains 25000 images of dogs and cats for training. The labels of these images are part of their filename. There are also 12500 images that can be used to test, and their filenames are numeric id. This dataset is officially provided in this competition, so it should be very suitable for this project. The Oxford-IIIT Pet

Dataset (http://www.robots.ox.ac.uk/~vgg/data/pets/) also can be used as training data, which include 37 category pet datasets. It includes 2371 cat images and 4978 dog images. These images have large variations in scale, pose and lighting, so using this dataset to train the network may avoid over-fitting.

Solution Statement

The solution of the problem is to use deep learning algorithm to train a neural network. A transfer learning will be used to increase the speed and accurate of the model. Several models, such as DenseNet 16.08, Xception 16.10, NASNet 17.07, etc, can be used in the transfer learning that are listed in Udacity dogs vs cats github website. The best model will be used in the capstone project. The training and testing data is mentioned above. The LogLoss function that is mentioned before can be used to calculate the score of the model. The solution can be evaluated by comparing the score with Kaggle Public Leaderboard.

Benchmark Model

As mentioned before, Kaggle provides a LogLoss function to calculate the score that can be used to evaluate the performance of the model. Kaggle Public Leaderboard provides a list of other competitors' score, so the score can be used to benchmark the model.

Evaluation Metrics

The test images will be used to test the model. The model need to evaluate each image in the test dataset and calculate the probability of a dog in the image. The result will be submitted to Kaggle, and the LogLoss equation is used to evaluate the model. The figure 1 is an example of the result that need to submit to the Kaggle and calculate the score. The LogLoss equation is also used to evaluate other competitors' models that are listed in the Kaggle Public Leaderboard and the project.

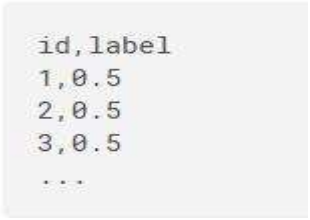

Figure 1: Example of result of test dataset

Project Design

1. Input dataset from Kaggle(https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data) and the Oxford-IIIT Pet Dataset
2. Transfer each image to four-dimensional tensor, which is the required format for the TensorFlow package. The size of each image need to be scaled into same size, for example (1, 224, 224, 3).
3. The Kaggle training dataset will be used to train the model. The Oxford-IIIT Pet Dataset probably will be used if Kaggle training dataset is not sufficient, which means the final score is not in the first 10% of the Kaggle Public Leaderboard of this competation. The Kaggle test dataset will be used to test the model, and the LogLoss equation function will be used to evaluate the model to calculate the scrore that is mentioned before.
4. Data augmentation will be used to increase the training dataset, which can help to avoid overfitting and it can make the model has better generalization.

5. Some models, such as DenseNet 16.08, Xception 16.10, NASNet 17.07, etc, will be test and selected to do the transfer learning. The best model will be used in the capstone project. The process of this step basically is:
    1. Load the pre-trained model
    2. Extract Features
    3. Create my own model
    4. Train the model
    5. Test the model
6. Probably will use the model to make an application, such as WeChat application, Web application or Android application.