Machine Learning Engineer Nanodegree
Capstone Proposal

Egor Ermilov
2017-06-21

Proposal

Domain Background

Image recognition is a classical problem in computer vision. It solves many different problems such as human-computer interaction, detection, control, navigation, etc.

Deep Neural Networks are very efficient at finding difficult patterns in the data (including image recognition). Adding a convolutional layer makes a huge impact on the neural networks and their abilities to identify objects on an image.

As we already saw from the previous Udacity Machine Learning Nanodegree project, a neural network with two convolutional layers could identify objects with hight accuracy.

Emotion recognition is now being actively researched. A good example of the research: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2731770/

Problem Statement

However, the goal of this project is to train a deep neural network to recognize human emotions. The model should be able to recognize emotions by a given human face photo with high accuracy. It makes a classification problem with face photos as inputs and emotion types as outputs.

It should be able to use not only the test photos from the original dataset, but also any custom photo uploaded.

Datasets and Inputs

This project will use the dataset from the Emotion and identity detection from face images Kaggle competition (https://inclass.kaggle.com/c/facial-keypoints-detector/data).

The dataset consists of 3,761 gray-scaled images of 48x48 pixels in size and a 3,761 label set of seven elements each.

Each element encodes an emotional stretch:
0 = anger,
1 = disgust,
2 = fear,
3 = happy,
4 = sad,

5 = surprise,
6 = neutral.

Value counts for the classes:
0 437
1 457
2 424
3 758
4 441
5 459
6 1202

Class 6 appears much more often than other classes, therefore the classes are unbalanced.

Solution Statement

For this project a convolution neural network will be built and trained on the Kaggle dataset. The dataset will be firstly preprocessed, normalized and split into training, testing and validation datasets. The application will be written in Python and use Tensorflow for building the model.

The approximate architecture of the network would be the following:

– Input layer
– Convolutional layer 1
– Pooling layer 1
– Convolutional layer 2
– Pooling layer 2
– Fully connected layer 1
– Fully connected layer 2
– Output layer

Benchmark Model

In the field of emotion recognition it is hard to be 100% accurate. People themselves don't always understand their emotions, let alone computers. The most of the reported models have an accuracy about 80%.

The benchmark model for this project would be this model:
http://www.paulvangent.com/2016/04/01/emotion-recognition-with-python-opencv-and-a-face-dataset/
which uses a very similar dataset and has an accuracy of 82.5%

Evaluation Metrics

Since the classes are unbalanced, the evaluation metrics would be:
– accuracy (the ratio of correctly predicted observation to the total observations)
– precision (the ratio of correctly predicted positive observations to the total predicted positive observations)
– recall (the ratio of correctly predicted positive observations to the all observations in actual class)

Project Design

Firstly, the dataset will be downloaded and preprocessed. The images will be transformed from strings into numpy arrays, normalized and reshaped. The dataset would be split into training, testing and validation datasets in the 80/10/10 proportion.

Then a neural network model will be built. The architecture is described above. The model will be trained on the training dataset. The performance will be evaluated every epoch on the validation dataset, and finally on the testing dataset.

The model would be refined by adding/removing layers, adjusting hyperparameters, choosing different activation function, etc.

Once the model is refined, trained and evaluated, it would be saved.

The saved model would be used to upload a custom photo and evaluate the performance.