574 Project 4 Report

Jingyi Zhao (50245749)
Zheng Kai (50247576)
Can Yao (50243637)

Abstract

The goal of this project is to implement a convolution neural network to determine whether the person in a portrait image is wearing glass or not. The whole procedure includes extracting feature values and labels from the CelebFaces Attributes Dataset, formalizing the pictures, data partition, model training and the final accuracy test. In our final test, the accuracy is 92%.

Introduction

In this project, we implement a CNN to distinguish the person in the picture is eyeglasses or not. In the CNN model, the picture is typically represented by a 3D array of numbers. All the pictures we're using are 64 pixels in height and 64 pixels in width, and each pixel is defined by three integers: red, green, and blue channels.

The key to implement CNN is the "convolution layer". A convolution layer applies a set of "sliding windows" across the image. Which namely are termed filters, and the model can detect the different primitive shapes or patterns through these filters.

In the convolution layer, the filters are passed across the input image, row by row, and they activate when they detect their shape. Rather than isolate the pixel values, the filters treat them in small groups, that's the CNN model can learn meaning features and locate them in any part of the image.

Like most other "deep learning" networks, CNNs tend to have many layers. For example, lower layers may learn simple edges or lines and subsequent higher layers can learn features like shapes(such as ears, noses,etc.)

After the model the training, we still use the following format to evaluation the accuracy of CNNs:

$$E = \frac{N_{\text{wrong}}}{N_{\text{V}}},$$

Where N_{wrong} is the number of misclassification and N_V is the size of the validation dataset. Under the 1-of-K coding scheme, the input will be classified as

$$C = \arg\max_i y_i.$$

Experiment Design

- General Procedure

  The CelebFaces Attributes Dataset is split into a training set, validation set, and the test set. The training set contains 1600 pictures(50% with glass, 50% not) , the validation set contains 400 pictures(50% with glass, 50% not), the test set contains 200 pictures(50% with glass, 50% not). Each picture will be treated into 64*64*3 image array, and apply the filters to train the CNN model. The validation set is used to pick hyper parameter for each model, and the test set is used to evaluate the performance of each model.

- Validation Process & Parameter Tuning

  Validation data is used to minimize overfitting. This set does not adjust the weight of the neural network. If the accuracy based on the training data is increased after the adjustment of the training data, but the accuracy of the verification data is not increased or decreased, it means overfitting occur and early stopping should be introduced. We have successfully picked the best parameters for each model (see section parameter tuning part).

- Convolutional neural network (CNN design)

  The model was consulted from tensorflow and tflearn. The major part is referred from Tensorflow and tflearn official website. We used 9 layers deep neural network to ensure the loss function converges.

Parameter Tuning:

- Increasing learning rate enable a drastically drop in performance. Thus we pick learning rate as 0.001

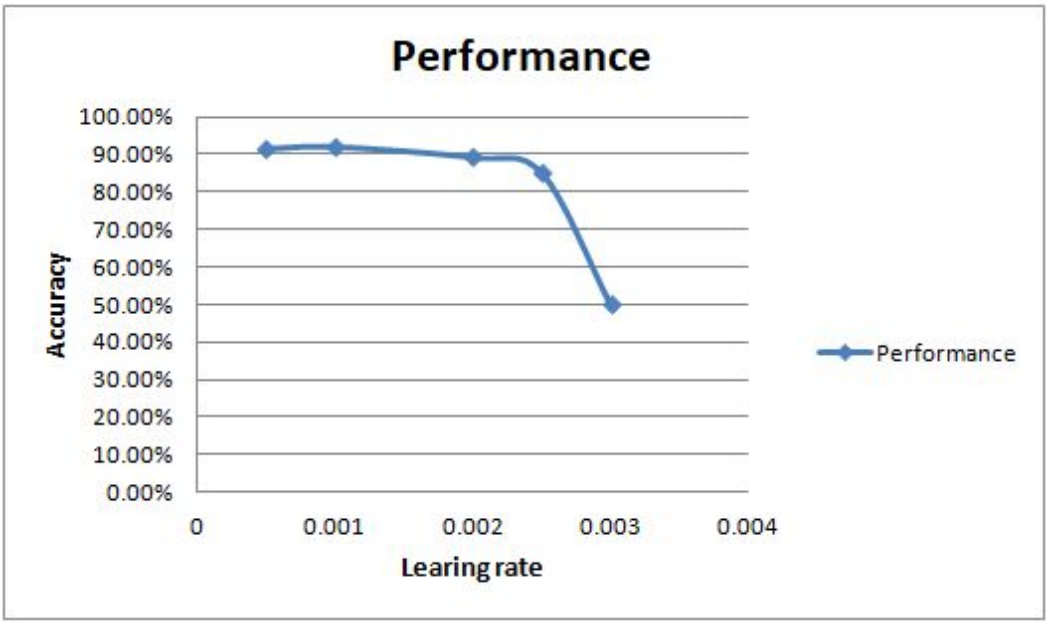

Figure 1-1:CNN accuracy at different learning rate

- The size of mini-batch have impact on the CNN model, that the larger the stride the slower and lower the performance is. we pick the mini-batch size as 10.

Batch size	Accuracy
10	91.8%
50	85.1%
100	77.1%

Data analysis

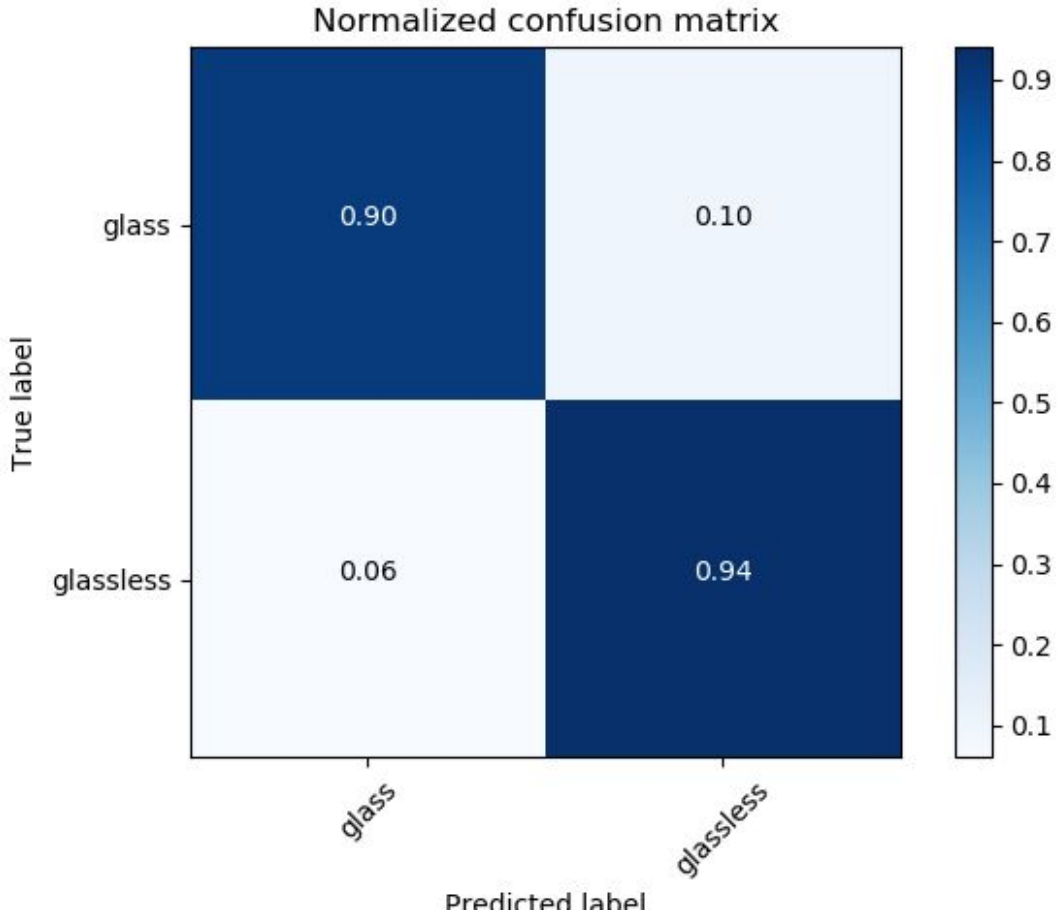

Figure 1-2: Confusion matrix of CNN model on Test data

	Training (1600)	Validation (400)	Test (200)
CNN	93.4%	91.8%	92%