Qinwen Yang

COMS 402C

Data Science Final Report

Stage 1. Ask A Question

The dataset being used in this project plays an important role in creating a custom model that could best classify the image data. By using these datasets, the questions I expect to answer are:

- What deep learning algorithm should apply to model that could best classifying image?
- Is the model that good at classifying training images also good at classifying testing images or newly input images? If not, how could improve the model and fix it?

  The benefits of this projects as personal is to be more familiar with computer vision which is a popular field of artificial intelligence. As group, this is a cornerstone for applying the image recognition function into our android application.

Stage 2. Get the Data

The dataset I am using for handwriting and image recognition is MNIST and CIFAR10 dataset which are both loaded from "torchvision" package by using pytorch framework. The MNIST dataset contains 60000 training images and 10000 testing images. These images are 28*28 grayscale handwriting images in 10 different classes. The CIFAR10 dataset consists of total 60000 32*32 color images in 10 different classes. There are 50000 training images and 10000 testing images. I haven't cleaned the data since it is not a messy data, these datasets only contain images with following labels in training dataset and only images in testing dataset.

Before using these datasets to create a model or test the model, I have prepared it by preprocessing those images, transfer each grayscale or color image into a tensor array or called matrix with fixed size and shape.

Stage 3. Explore the Data

In order to take a first look at those images, I have created a function which are able to transfer an image from tensor to numpy array and plot 20 images in each dataset use the matplotlib package. The original image in CIFAR10 dataset shows like a normal color camera image and the original image in MNIST dataset shows a black in background and white in foreground. One sample image in CIFAR10 and MNIST dataset are shown respectively.

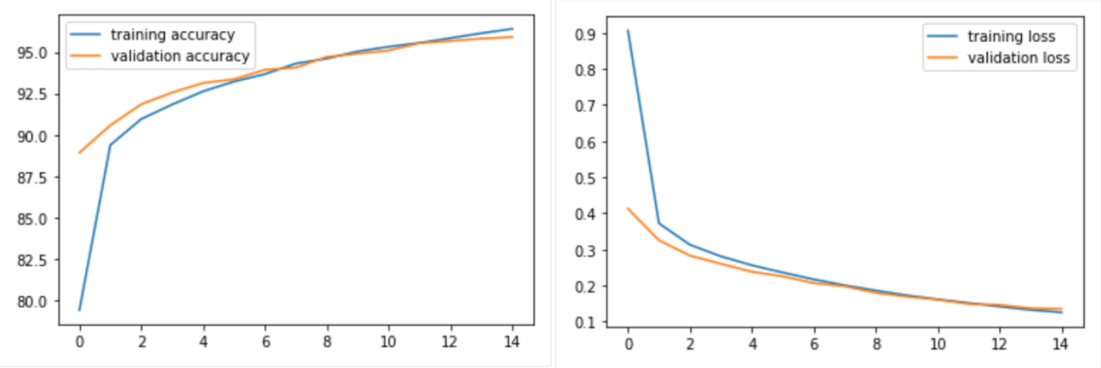
Stage 4. Model the Data

Two different models are generated by working with MNIST dataset, one model is built upon deep neural network, the other model is built upon convolutional neural network. Convolutional neural network is a deep machine learning algorithm which is widely known for image classification since it's efficiency in reducing the number of parameters without losing on the quality of models, especially with a large number of parameters cause by trying to fit more complicated model with the number of layers increase. By fitting MNIST training and testing

dataset separately into two models. The model with deep neural network has training accuracy close to 96%, testing accuracy close to 96% while the model with convolutional neural network have training accuracy close to 99%, testing accuracy close to 99%. By comparing the accuracy results, it clearly shows the accuracy for convolutional neural network model is higher than the deep neural network model which simply means the convolutional neural network model will classify the newly input images better.

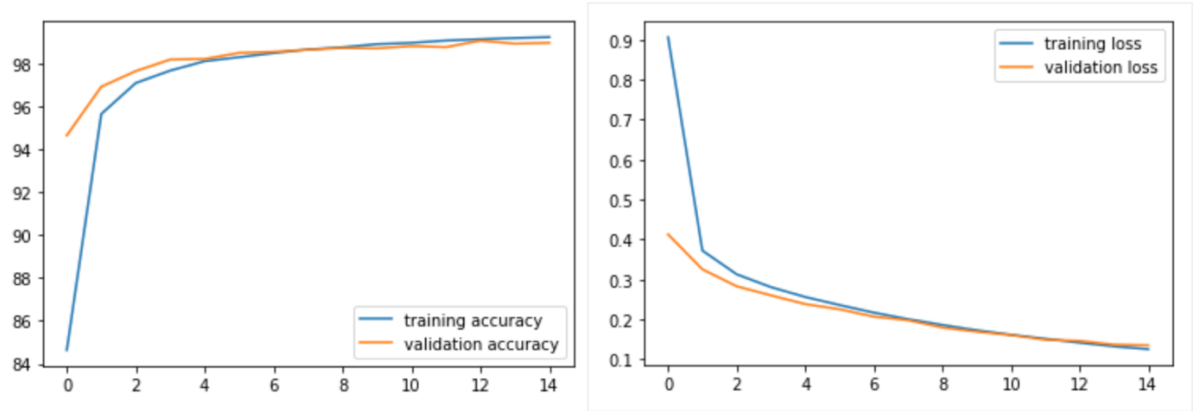
Deep Neural Network (MNIST Dataset)

```
epoch : 15
training loss: 0.1259, acc 96.4000
validation loss: 0.1355, validation acc 95.9100
```



Convolutional Neural Network (MNIST Dataset)

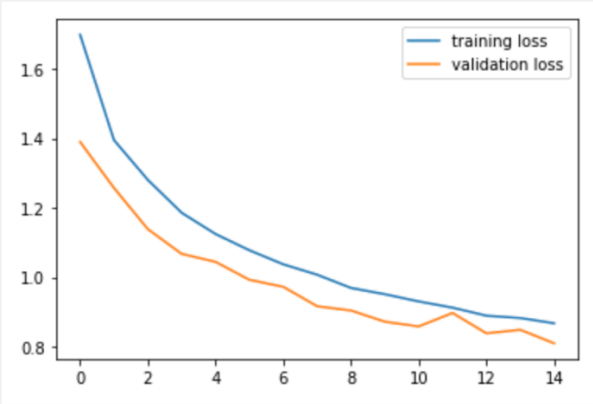
```
epoch : 15
training loss: 0.0248, acc 99.2367
validation loss: 0.0311, validation acc 98.9700
```

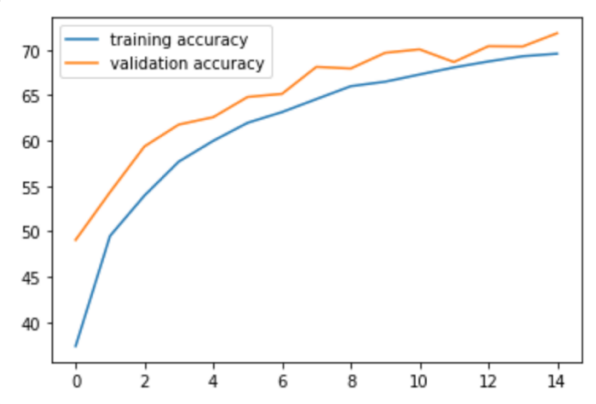


CIFAR10 is a dataset with bunch of 3D (color images depth = 3) RGB color images which will be more complex than MNIST dataset with only 2D (grayscale images depth = 1) grayscale images, in order to best classify the color images from CIFAR10 dataset, I have built the model with convolutional neural network. The model contains different layers, such as convolution layer, relu layer, pooling layer, fully connected layer. Each layer will bring various effect to the model. As an original input image transforms to an array of pixels, when the image pass through the convolution layer, the convolution layer will extract and learn specific image features by involving kernel (filter) over the image, the kernel also forms by an array and it contains a specific feature like curve, edge and it is like a feature detector. Normally, it will have lots of different filters for helping extract diverse image features for an image. After the convolution layer, it will output a feature map which can then follow with relu layer and then with pooling layer. For pooling layer, I have used maxpooling which will scale down the feature map size, but it will still keep with feature of interest at the same time. The training accuracy I eventually got for that model is close to 70% and the validation (test) accuracy is close to 72%. Overfitting is one of the challenge people will facing when creating a model, it always happened when a model able to generalizes a training images well but can't generalize unseen images well. The reason for the model doesn't generalize the unseen images well is it learns too much detail

and noise in the training data instead of training to learn a general contour for an image. In order to avoid overfitting issues, I have used drop out layers and apply data augmentation techniques such as make training image rotate, horizontal flip to increase the diversity of the image available for the model.

```
epoch : 15
training loss: 0.8673, acc 69.5900
validation loss: 0.8097, validation acc 71.8500
```





Use the model created to test the testing dataset:

Data Augmentation:

Stage 5. Communicate the Data

After exploring and modeling the data, the final work is to prepare a presentation that could show audience with the work I have done. In the presentation, there will be an oral explanation of the dataset I used, and the process for using convolutional neural network algorithm to create the model. In the first presentation, I have done recognize handwriting images by using other algorithm like support vector machine and decision tree. I understand some basic knowledge about computer vision at that time. Due to require more deep understanding about this topic and try to classify more complex dataset, I choose to learn neural network which include deep neural network and convolutional neural network. These algorithms are general used for classifying complicated images.

This project brings us a good experience on image recognition, how computer vision actually works. It might able to implement at a later time into our overall group application to allow capture image with smartphone camera and recognize the image. This will be a good function to apply to bring a better experience to user.