Hand Gesture Recognition using Convolutional Neural Networks

By FANG, PengChi

Supervised by

Prof. Dit-Yan Yeung

Introduction

Motivation

The development of Human-Computer Interaction (HCI) is rising in recent years. One of the branches is hand gesture recognition. Like the face recognition, it has the potential to be extensively used in reality. However, most of the current methods build classification algorithms based on complicated handcrafted features, which are not robust and generic enough. Thus, the Convolutional Neural Networks approach, which is powerful for image recognition, seems to be a reasonable solution for the problem.

Objective

The main purpose of our project is to develop a Convolutional Neural Networks model, which has the following properties:
➢ Able to recognize hand gestures with high classification rate
➢ Robust to various hand gestures

Concept

Convolution operation is the reason that Convolution Neural Networks (CNNs) is powerful for image recognition.
The output Y is obtained by convolving the input X with the kernel W. The kernel W acts like an edge detector, which detects a specific image feature throughout the input X. The output Y is a feature map, which records the possibility of occurrence of a certain feature.

Input X Kernel W Output Y

$$y_{00} = x_{00} w_{00} + x_{01} w_{01} + x_{10} w_{10} + x_{11} w_{11}$$

Dot Product
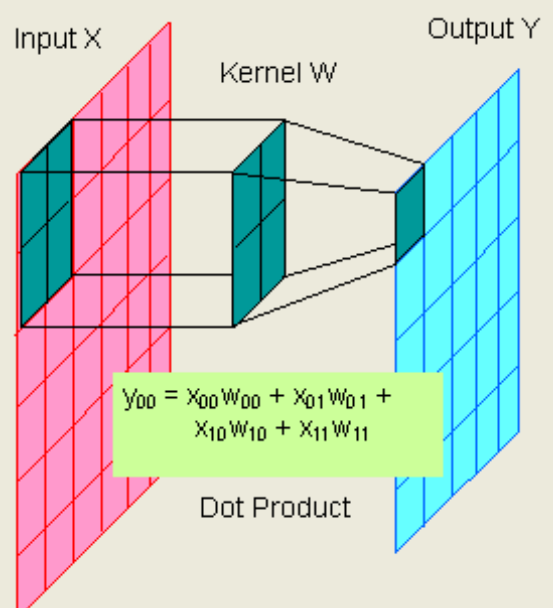
Implementation

CNNs training process requires two fundamental elements: (1) A large scale of input dataset (2) A CNNs model.

In the training process, the CNNs model learns the hand gesture features by scanning all the input images with hundreds of times.
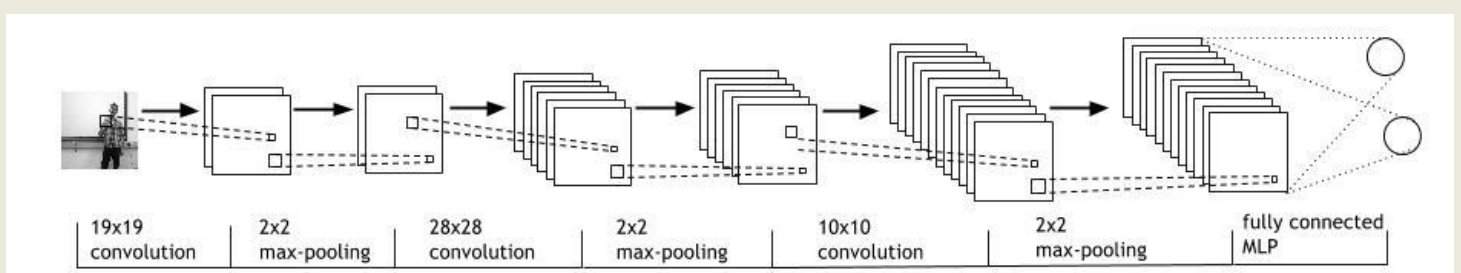
Input dataset (2 gestures)

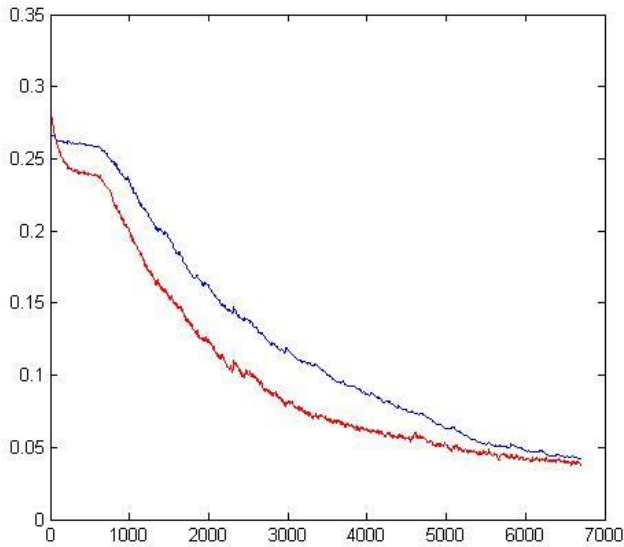

Samples of input datasets: Left - cheduepalle, Right – ok

Type	cheduepalle	ok	Total
Training dataset	1000	1000	2000
Testing dataset	52	72	124
Total	1052	1072	2124

Table of dataset statistics

CNNs model

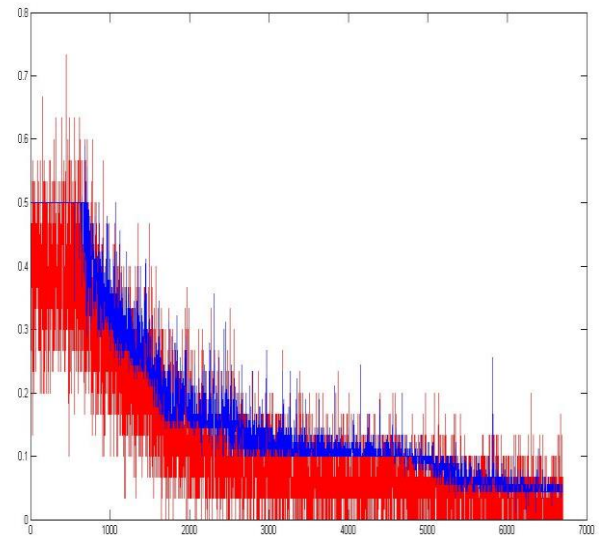| 19x19 convolution | 2x2 max-pooling | 28x28 convolution | 2x2 max-pooling | 10x10 convolution | 2x2 max-pooling | fully connected MLP |

Evaluation

Trend of mean squared error rate

Trend of error rate

Training dataset

The minimal mean squared error = 0.03786

The minimal error rate = 0

Testing dataset

The minimal mean squared error = 0.04214

The minimal error rate = 0.0111

The classification rate is around 0.93 to 0.95.

The best classification rate of existing methods is around 0.96.

Conclusion

The current CNNs model achieve a favorable performance for recognizing 2 hand gestures. Compared with the existing methods, the complexity of classifiers is less complex. For future development, to make the CNNs model more robust and versatile to more gestures, train the model with some larger datasets