- Architecture of my Deep CNN

Conv Layer 1:

Input: 48x48x1 grayscale face
→Convolution(number of kernels: 64, kernel size: 3x3, strde: 1)
→Batch Normalization
→ReLU
→Max Pooling(kernel size: 2x2, stride: 1)
→Dropout(probability of keep = 0.5)
Output: 48x48x64

Conv Layer 2:

Input: 48x48x64
→Convolution(number of kernels: 128, kernel size: 5x5, strde: 1)
→Batch Normalization
→ReLU
→Max Pooling(kernel size: 2x2, stride: 2)
→Dropout(probability of keep = 0.5)
Output: 24x24x128

Conv Layer 3:

Input: 24x24x128
→Convolution(number of kernels: 512, kernel size: 3x3, stride: 1)
→Batch Normalization
→ReLU
→Max Pooling(kernel size: 2x2, stride: 2)
→Dropout(probability of keep = 0.5)
Output: 12x12x512

Conv Layer 4:

Input: 12x12x512
→Convolution(number of kernels: 512, kernel size: 3x3, stride: 1)
→Batch Normalization
→ReLU
→Max Pooling(kernel size: 2x2, stride: 2)
→Dropout(probability of keep = 0.5)
Output: 6x6x512

FC Layer 1:

Input: 6x6x512
→Flatten: compress to 18432
→Linear product(number of the neuron: 256)
→Batch Normalization
→ReLU
→Dropout(probability of keep = 0.5)
Output: 256

FC Layer 2:

Input: 256
→Linear product(number of the neuron: 512)
→Batch Normalization
→ReLU
→Dropout(probability of keep = 0.5)
Output: 512

FC Layer 3:

Input: 512
→Linear product(number of the neuron: 7)
Output: 7 (note that: this is the **final output** of the model)

Take the 7-elements final output as the input of the Softmax function, we'll get a corresponding 7-elements output. Then the index(from 0 to 6) of maximal element of this softmax output is the emotion of the our input face.

For all the layers in this architecture, I'm using the L2 regularization to prevent overfitting. The hyper parameter lambda of the regularization is 1e-7.

- ## Optimization

The optimizer is the AdamOptimizer with:
learning rate = 0.01, beta1=0.9, beta2=0.999, epsilon=1e-08

- ## Training

Run on the Google Cloud Computing platform, with a 8-cores CPU.
Runtime: about 7 hours
Data volume in Train set: 32298
Data volume in Test set: 3589
Batch size: 128
Number of epoch: 30
Number of steps in each epoch: 253 (32298/128 = 252.32, round it to 253)
Total number of the steps : 253*30 = 7590