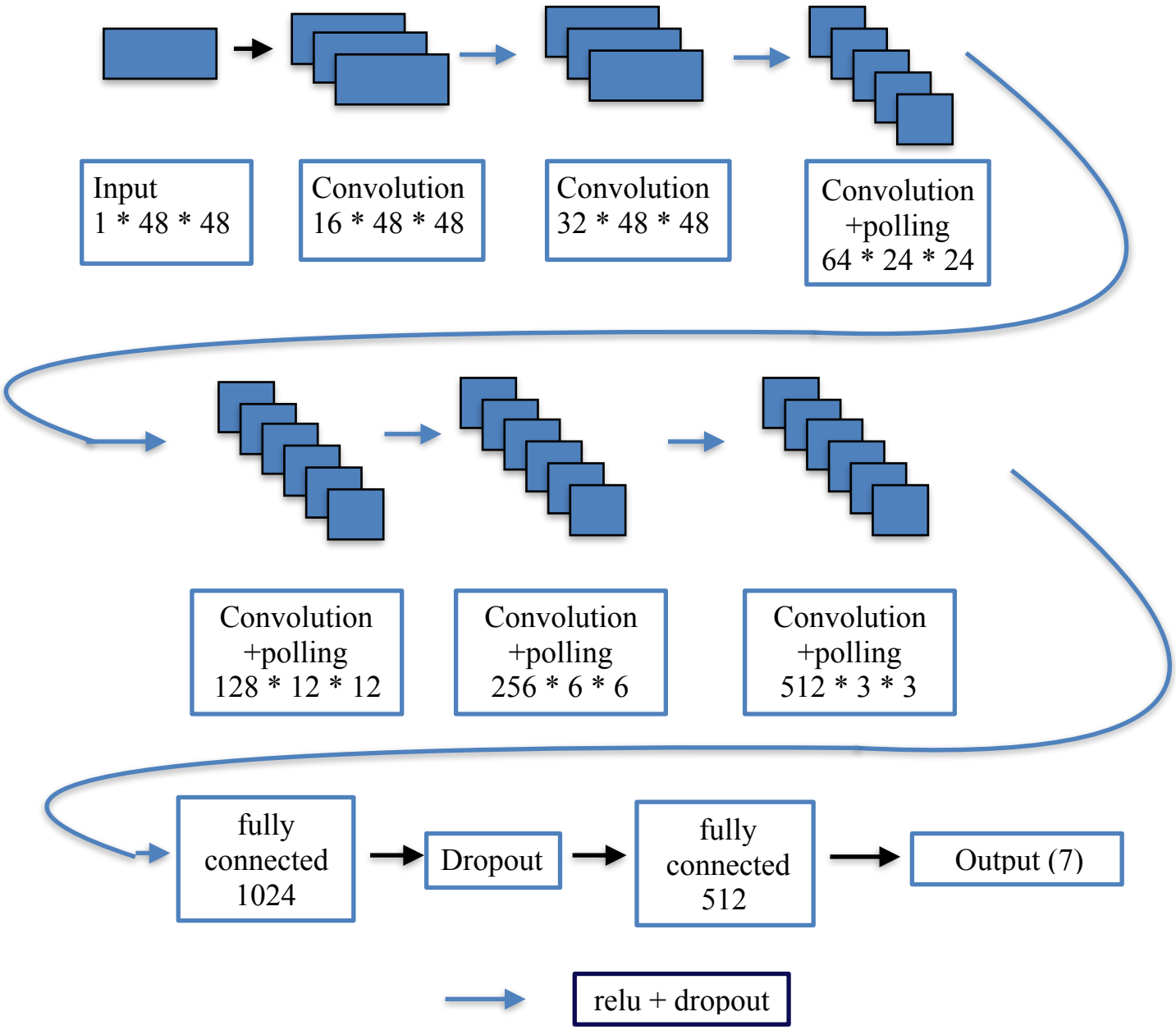
Homework3 Report

Professor Pei-Yuan Wu
EE5184 - Machine Learning

姓名：張雅信
學號：B05902022

1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？



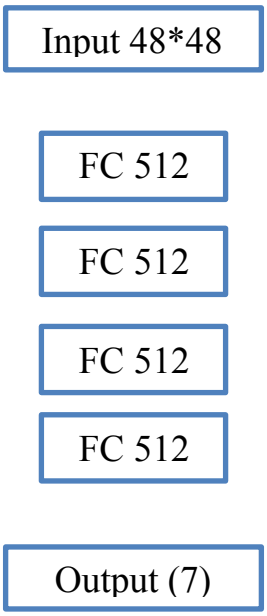
訓練過程：

Make batch —> forward —> loss backward —> ….(loop)

準確率：

Model	Public score	Private score
單一 model	0.69183	0.70075
Ensemble 2 models	0.69796	0.71858

2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model，其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？



訓練過程：

Make batch —> forward —> loss backward —> ….(loop)

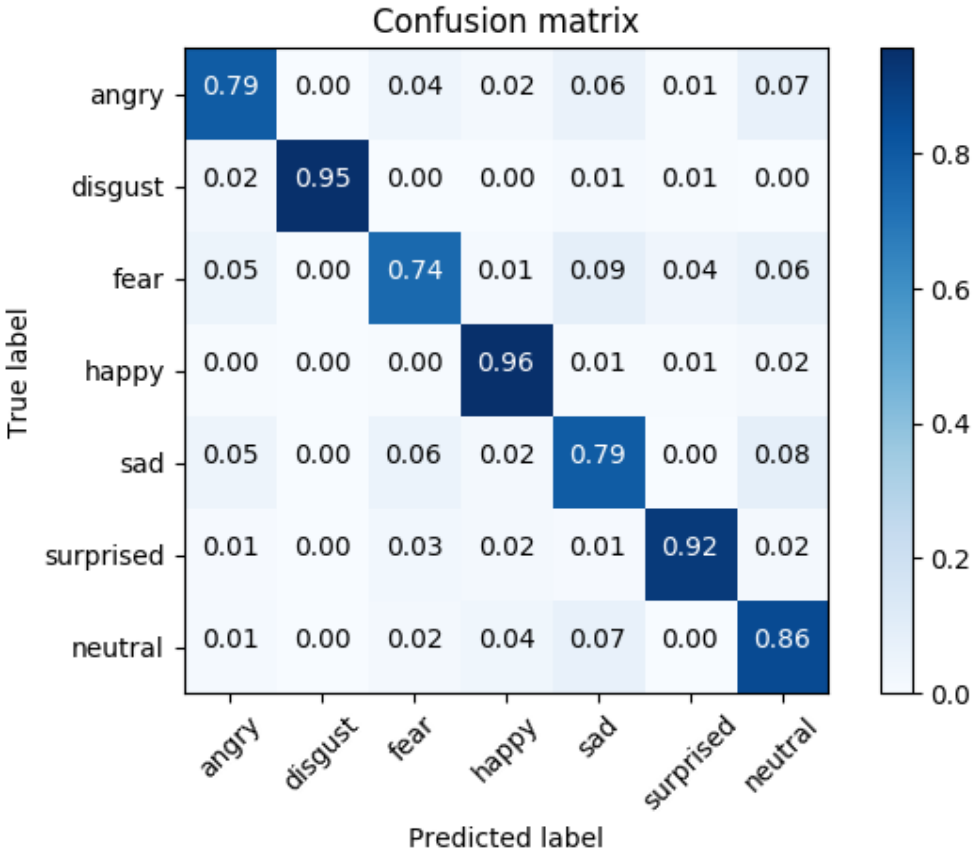
Model	Public score	Private score
CNN	0.36277	0.37977
DNN	0.69138	0.70075

觀察：

DNN 不像 CNN 有Convolution layer，較難察覺各個 pixel 之間的關係，因此在同樣數量的參數下，DNN 的準確率低了許多。

(就算沒有仔細調整架構與Hyperparameter, CNN 也可以輕鬆超過50%)

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？ 並說明你觀察到了什麼？ [繪出 confusion matrix 分析]



從上圖可看出，disgust、happy、surprised 三者比較不易搞混，angry、fear、sad 三者容易互相搞混。 sad 也容易與 neutral 搞混。

----------Handwritten question----------

Collaborator: b0590274 魏佑珊

4. (1.5%,each 0.5%)CNN time/space complexity:

For a. b. Given a CNN model as

```
model = Sequential()
model.add(Conv2D(filters=6,
                 strides=(3, 3),
"""Layer A"""    padding ="valid",
                 kernel_size=(2,2),
                 input_shape=(8,8,5),
                 activation='relu'))
model.add(Conv2D(filters=4,
                 strides=(2, 2),
"""Layer B"""    padding ="valid",
                 kernel_size=(2,2),
                 activation='relu'))
```

And for the c. given the parameter as:

kernel size = (k,k);
channel size = c;
filter size = f;
input shape = (n,n);
padding = 1;
strides = (s,s);

a. How many parameters are there in each layer(Hint: you may consider whether the number of parameter is related with)

Layer A: (5*2*2+1)*6 = 126

Layer B: (6*2*2+1)*4 = 100

b. How many multiplications/additions are needed for a forward pass(each layer).

Layer A:

6 * 5 * 2^2 * 3^2 = 1080 multiplications

6 * (5 * 2^2-1) * 3^2 = 1026 additions

Layer B:

4 * 6* 2^2 * 1 = 96 multiplications

4 * (6 * 2^2-1) * 1 = 92 addition

c. What is the time complexity of convolutional neural networks?(note: you must use big-O upper bound, and there are l layer, you can use C_l, C_{l-1}as lth and l-1th layer)

For each layer, it takes

$f * c * k^2 * \lceil \frac{n}{s} \rceil^2$ multiplications and

$f * (c * k^2 - 1) * \lceil \frac{n}{s} \rceil^2$ additions

And the time complexity of a CNN is $O(\sum_{1}^{l} c_l * f_l * k_l^3 * \lceil \frac{n}{s} \rceil^2)$

5. (1.5%,each 0.5%)PCA practice:Problem statement: Given 10 samples in 3D space.(1,2,3),(4,8,5),(3,12,9),(1,8,5),(5,14,2),(7,4,1),(9,8,9),(3,8,1),(11,5,6),(10,11,7)

a. (1) What are the principal axes?

μ = (5.4, 8, 4.8)

covariance matrix = $\begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix}$

principal axes:

(0.3999, 0.3376, -0.8521)

(-0.6782, 0.7344, -0.0273)

(-0.6166, -0.5888, -0.5226)

b. (2) Compute the principal components for each sample.

(1, 2, 3) -> [-1.4814, 0.7087, -3.362]

(4, 8, 5) -> [0.0394, 3.026 , -9.7899]

(3, 12, 9) -> [-2.4187, 6.5326, -13.6189]

(1, 8, 5) -> [-1.1601, 5.0605, -7.9401]

(5, 14, 2) -> [5.0212, 6.836 , -12.3716]

(7, 4, 1) -> [3.2972, -1.837 , -7.194]

(9, 8, 9) -> [-1.3699, -0.4741, -14.9632]

(3, 8, 1) -> [3.0481, 3.8133, -7.0829]

(11, 5, 6) -> [0.9735, -3.9517, -12.8622]

(10, 11, 7) -> [1.747 , 1.1055, -16.3011]

c. (3) Reconstruction error if reduced to 2D.(Calculate the L2-norm)

$L = \sum \|(x - \bar{x}) - (\sum_{k=1}^{K} (x - \bar{x}) \cdot w^k \cdot w^k)\|_2$

list of L2 norm (for each data)=

(2.2510, 0.7302, 3.188, 1.930, 4.2516,

2.5276, 2.1395, 2.2785, 0.2038, 0.9774)

sum = 20.4778