

[NTU ESOE] 109 Data Mining Final Project Deadline: 2020.12.29.23:00

Student ID: _r10525069_ Name: __林子傑__ Department: __工程海洋學系碩士班__

報告影片：https://drive.google.com/file/d/1D4rU79mdtAKhwV5_GYmqabMa5BLWKJoU/view

1.問題介紹

把圖片依內容分成 4 類，“text”有說明或導覽文字的圖片，“food”有食物的圖片，“people”有人的圖片，“landspace”其他圖片。

2. 請詳述你對調整模型參數的過程邏輯。可參考 **data preview** 以及 **model preview** 的結果來描述。

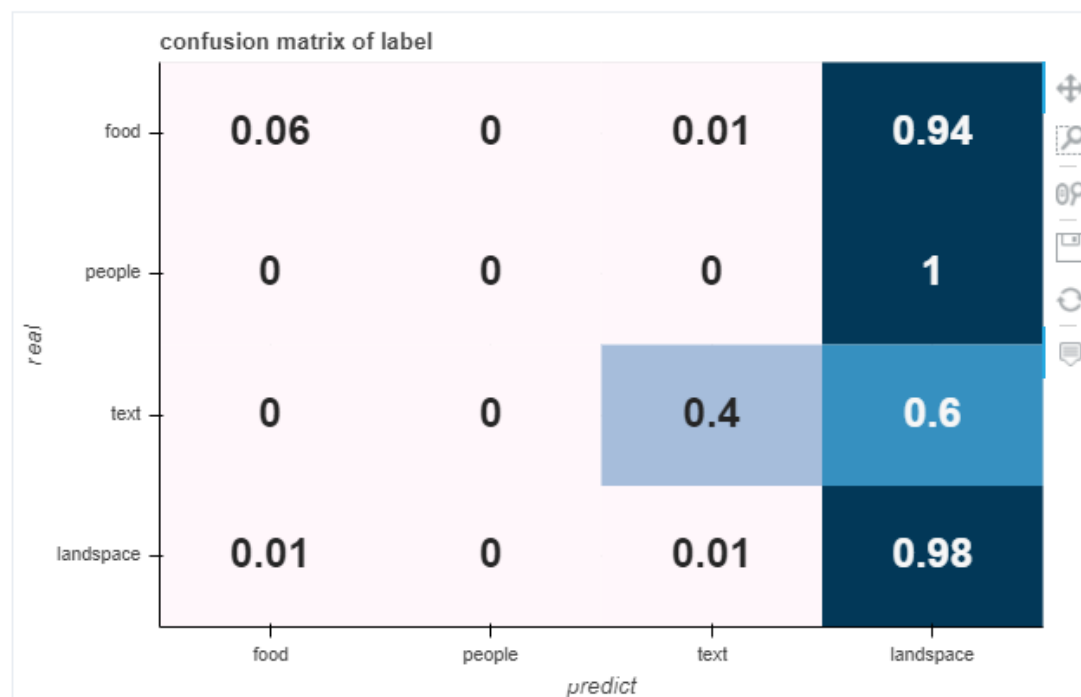
設計的模型如下：

Type	Output Shape	Param #
Conv2D	(None, 30, 30, 32)	896
Conv2D	(None, 28, 28, 32)	9248
MaxPooling2D	(None, 14, 14, 32)	0
Conv2D	(None, 12, 12, 64)	18496
Conv2D	(None, 10, 10, 64)	36928
MaxPooling2D	(None, 5, 5, 64)	0
Flatten	(None, 1600)	0
Dense	(None, 512)	819712
Dropout	(None, 512)	0
Dense	(None, 4)	2052
=====		
Total params: 887,332		
Trainable params: 887,332		
Non-trainable params: 0		

第一次實驗：training 預測結果大都預測為 landspace。

參數名稱	Hidden kernel size	Hidden Activation	Optimizer	Dropout Rate	Epoch	Batch size	Correlation Algorithm
值	3	relu	sgd	0.25	20	32	Pearson

Chart Preview



第二次實驗：發現第一次實驗的 Epoch 參數設太小，有可能是結果不佳的原因，因此把 Epoch 設為 200(最大值)。Training 結果所有的圖片都判斷成正確的分類，但利用 testing 的結果卻不理想，發生 dropout 的情況。

參數名稱	Hidden kernel size	Hidden Activation	Optimizer	Dropout Rate	Epoch	Batch size	Correlation Algorithm
值	3	relu	sgd	0.25	200	32	Pearson

Chart Preview

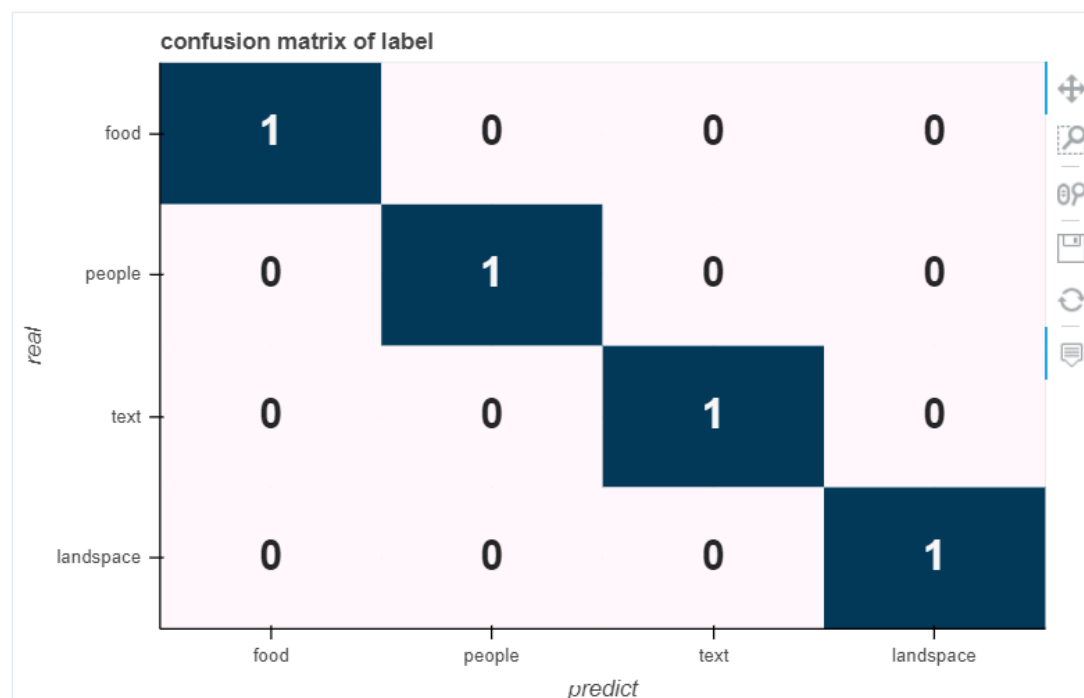
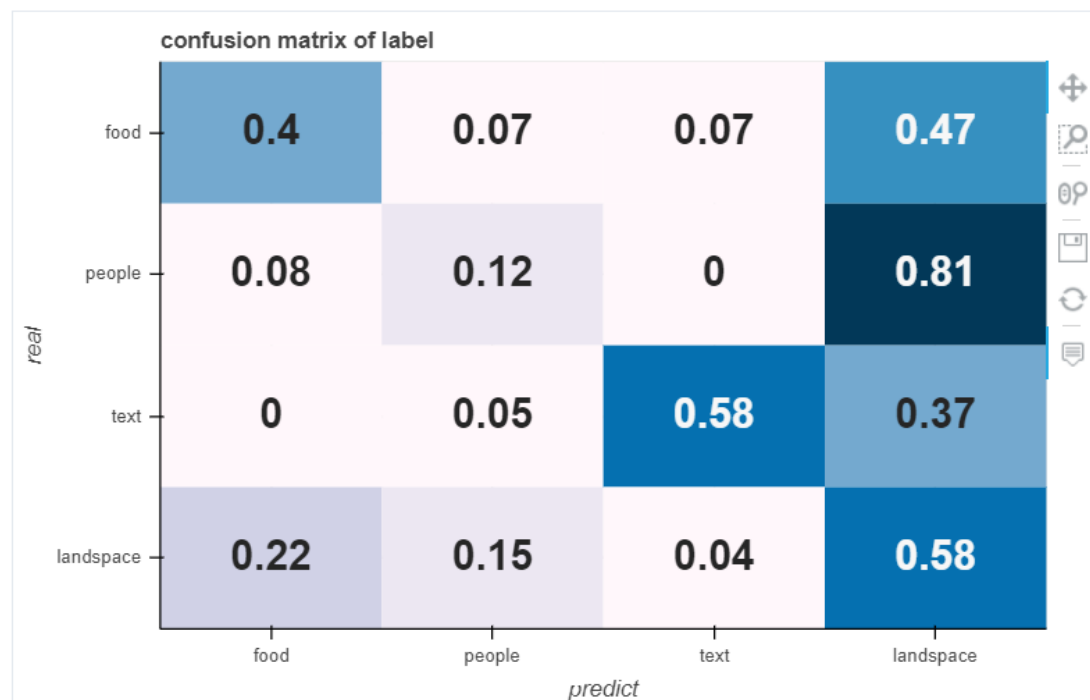


Chart Preview



第三次實驗：調大第二次實驗的 dropout 比例，Training 結果依然為所有的圖片都判斷成正確的分類，Testing 的結果較第二次優，但很有很大的進步空間。

參數名稱	Hidden kernel size	Hidden Activation	Optimizer	Dropout Rate	Epoch	Batch size	Correlation Algorithm
值	3	relu	sgd	0.5	200	32	Pearson

Chart Preview

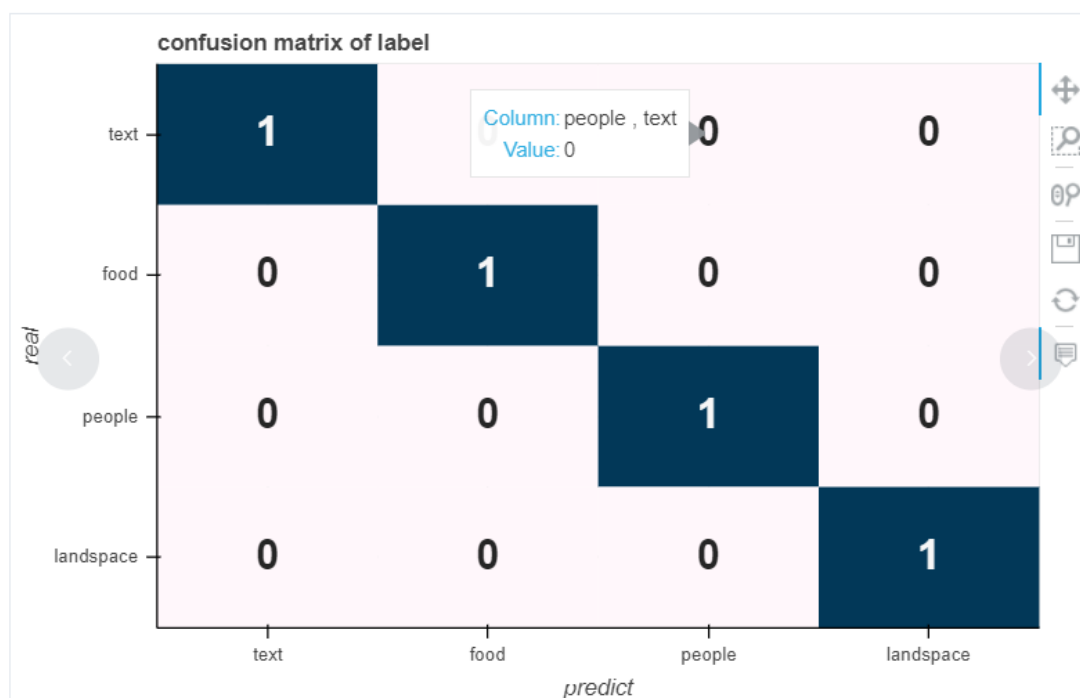
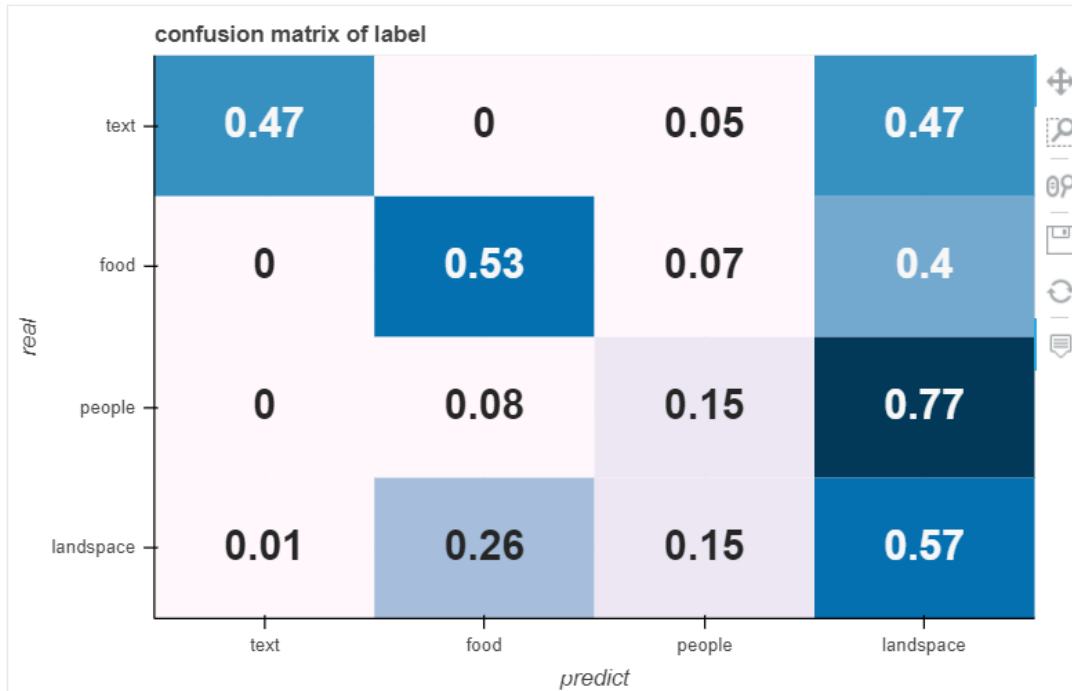


Chart Preview



後來我詢問別人，model 的表現跟參數和訓練圖片量之間的比例有關，一般來說在 3:1 左右比較好，我試著降低模型的參數數量。

Layer (type)	Output Shape	Param #
Conv2D	(None, 30, 30, 8)	224
Conv2D	(None, 28, 28, 8)	584
MaxPooling2D	(None, 14, 14, 8)	0
Flatten	(None, 1568)	0
Dense	(None, 16)	25104
Dropout	(None, 16)	0
Dense	(None, 4)	68
Total params: 25,980		
Trainable params: 25,980		
Non-trainable params: 0		

第四次實驗：參數同第三次的實驗，差異是改變模型， Training 和 Testing 和第三次差異不大。

參數名稱	Hidden kernel size	Hidden Activation	Optimizer	Dropout Rate	Epoch	Batch size	Correlation Algorithm
值	3	relu	sgd	0.5	200	32	Pearson

Chart Preview



Chart Preview



第五次實驗：將第四次實驗的 **Kernel Size** 改成 5，其餘不變。Training 和 Testing 和第四次差異不大。

參數名稱	Hidden kernel size	Hidden Activation	Optimizer	Dropout Rate	Epoch	Batch size	Correlation Algorithm
5							
值	5	relu	sgd	0.5	200	32	Pearson

Chart Preview



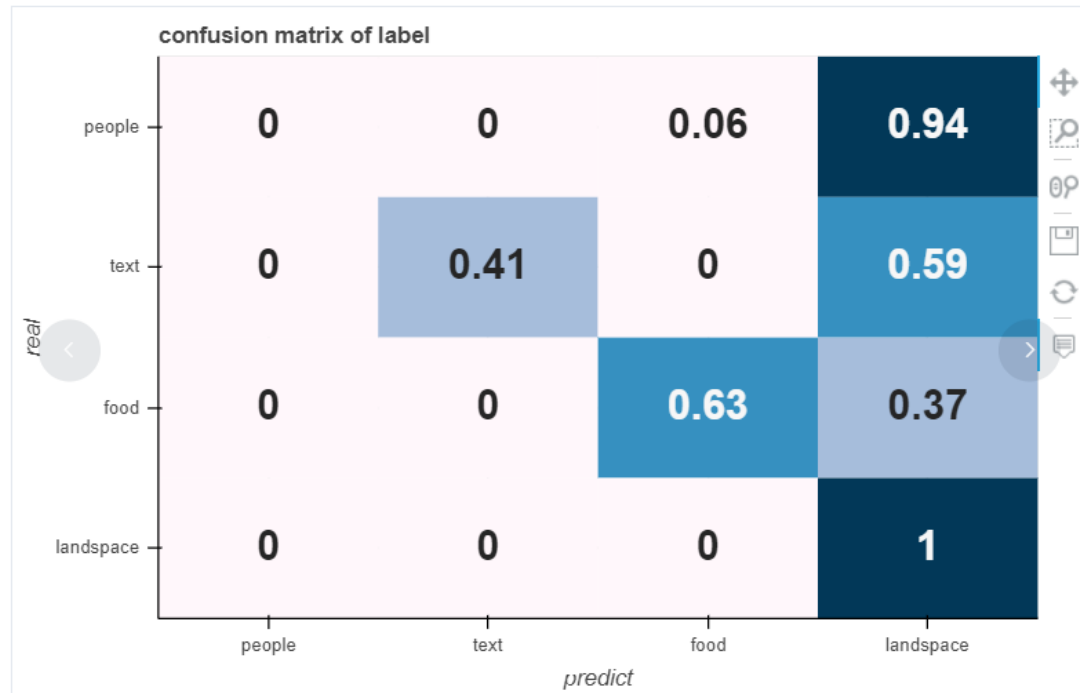
Chart Preview



第六次實驗：將第四次實驗的 Dropout Rate 改成 0.75，其餘不變。Training 的表現下降。

參數名稱	Hidden kernel size	Hidden Activation	Optimizer	Dropout Rate	Epoch	Batch size	Correlation Algorithm
6							
值	3	relu	sgd	0.75	200	32	Pearson

Chart Preview



3. 你最終的訓練以及測試結果為何？請有邏輯的解釋你的結果

最後選定第四次實驗為最後結果，Testing 的結果不盡理想，可能是資料集大小過小所以無法得出很有效的模型，如果要提升 Training 的成效，可能要再加一些資料。