

MLP小專案

曾宏鈞 06160485
徐友笙 05360365
蔡毓丞 06370136
盧君彥 05360153

資料集與問題說明

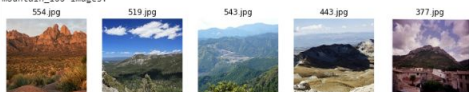

資料集：[Intel Image Classification](#)

內容：

張數：25,000

解析度：150x150

類別：6

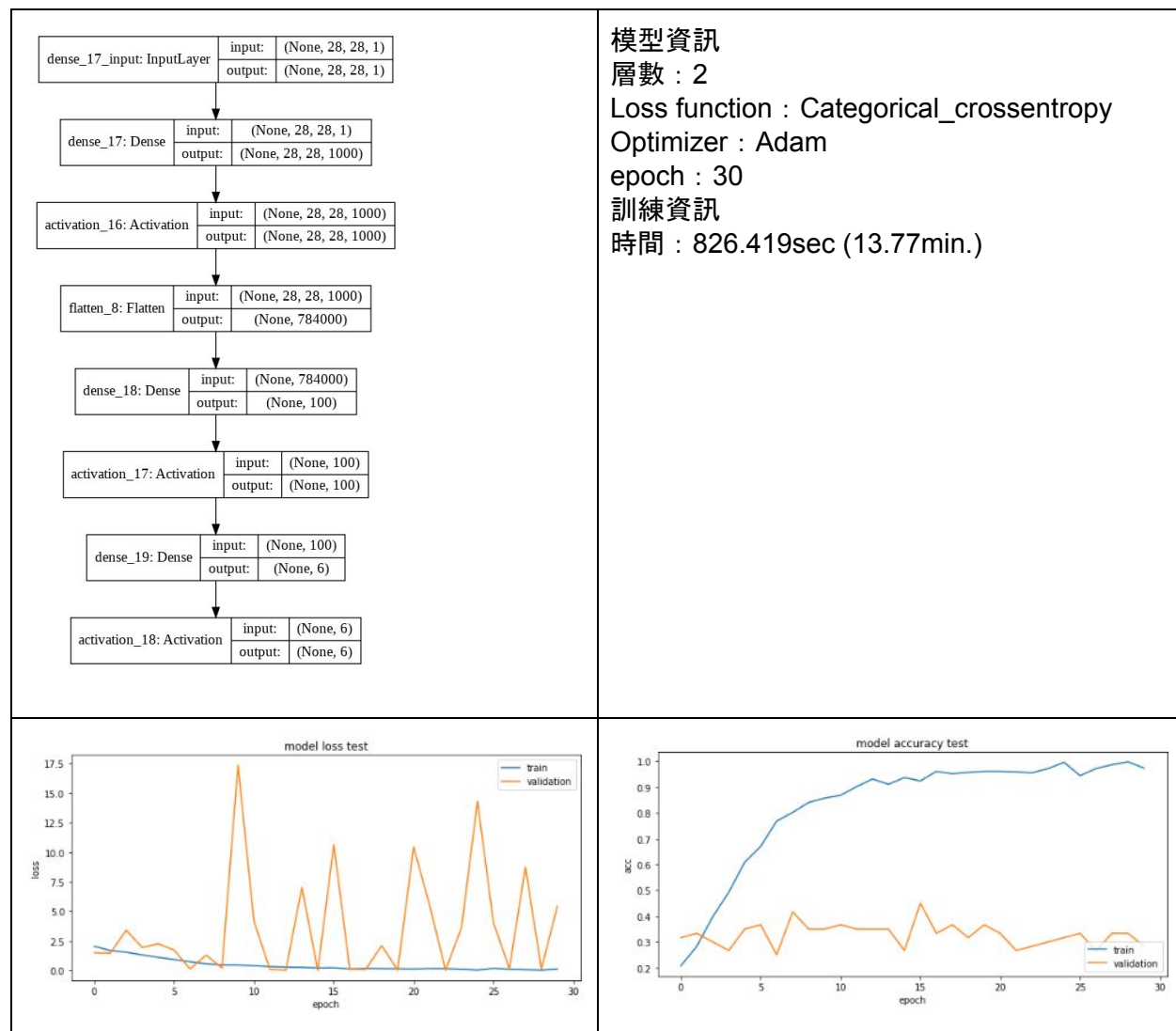
建築物	0
森林	1
冰川	2
山	3
海	4
街道	5
<div>forest_100 images:  mountain_100 images:  glacier_100 images:  street_100 images:  sea_100 images:  buildings_100 images: </div>	Train:540 Validation:60 Test:310

實驗流程

Dataset: train=90, vlaidation=10

一開始我們使用ImageDataGenerator製作分別為90%與10%的training資料與validation資料灰階影像。

模型一

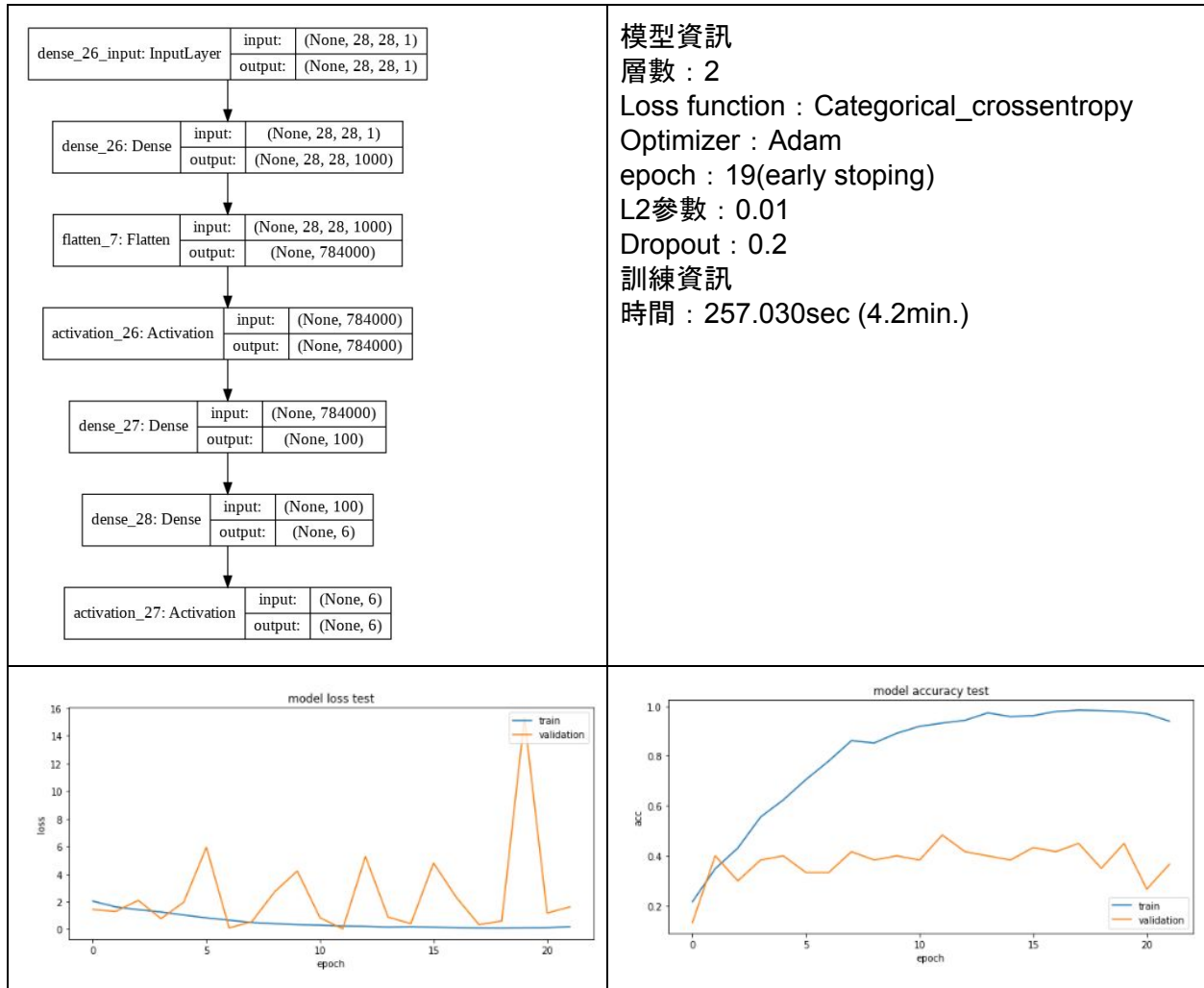


從結果可以看出不管怎麼樣validation的loss在5個epoch後來回震盪。在10個epoch達到高峰，且train的準確度>>validation。且由callback可看出callback在epoch=3時， validation準確度就沒有增加，為0.35。代表我們的模型已經過度訓練，可能發生high-variance問題，因此我們需要regularization。

模型二

因為遇到high-variance問題，因此我們採取了幾個作法：

- 添加L2正規化及Dropout層(丟棄20%)，再加上Early stoping使模型在第19個epoch時，因為連續10個epoch的loss修正validation都高於train，所以就在此時停止，雖然訓練時間較短，但模型validation的準確率並未改善太多。

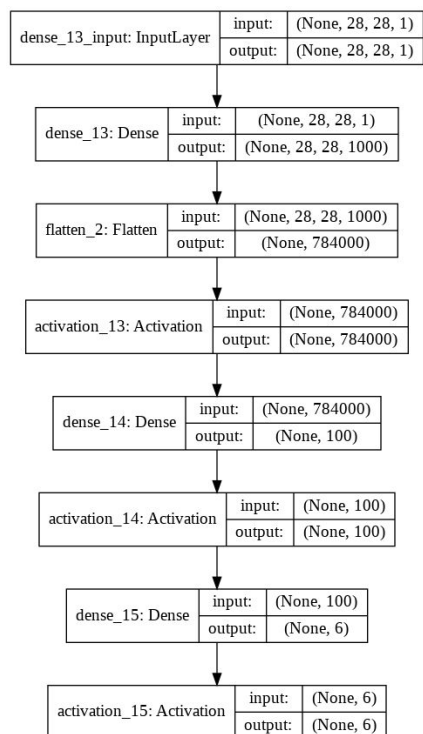


由於已經使用幾種regulation方法，仍獲得不到改善，有可能是資料筆數太少的問題，因此我們作了以下實驗。

Dataset: train=450, validation=50

模型一

與前面一樣，發生high-variance問題，需要regulation



模型資訊

層數：2

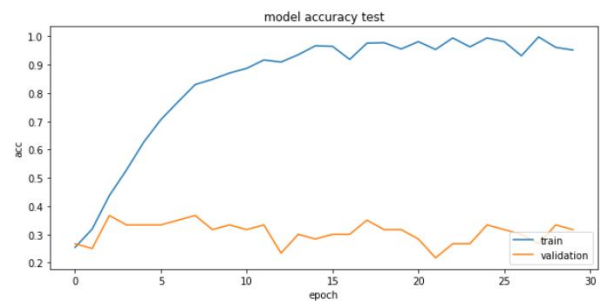
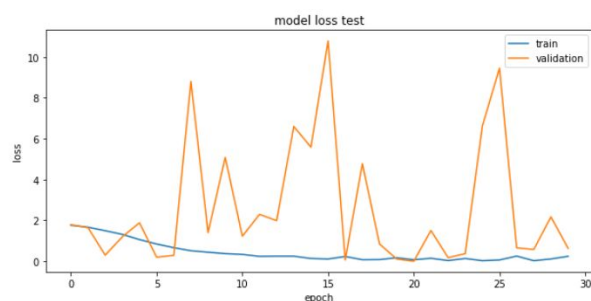
Loss function：Categorical_crossentropy

Optimizer：Adam

epoch：30

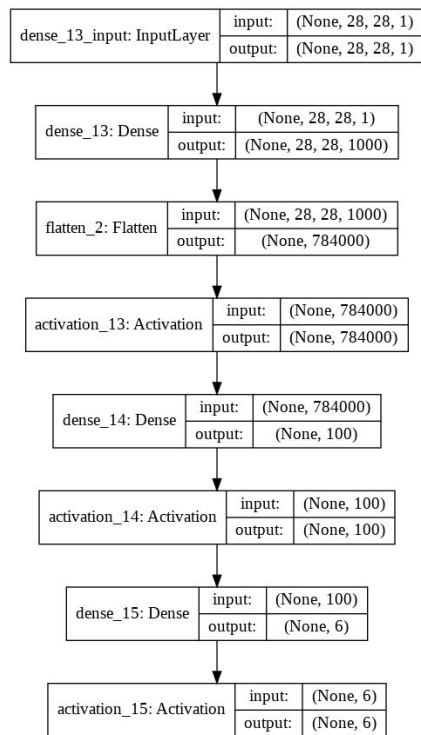
訓練資訊

時間：2095sec (35min.)



模型二

觀察發現train與validation差不多疊合，但準確度低，再觀察loss的曲線學習平緩，發生High-bias，需要optimization，應該繼續實驗添加加BatchNormalization層，或使用其他的最佳化演算法。



模型資訊

層數：3

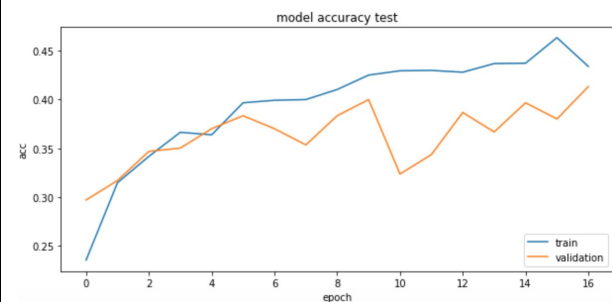
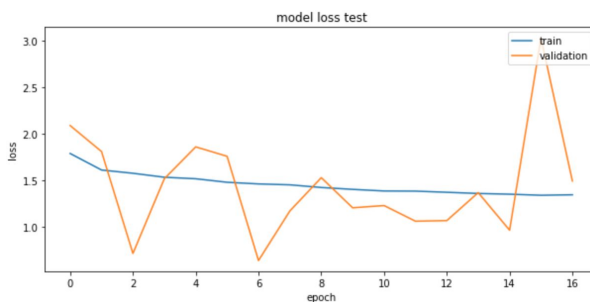
Loss function：Categorical_crossentropy

Optimizer：Adam

Epoch：16 (early stopping)

訓練資訊

時間：6099sec (35min.)



心得

這次的實驗我們使用MLP來解MLP的圖像分類問題，雖然是圖像分類，但這次的圖片本身資料蠻複雜的，是生活中不同場景的辨識，再加上時間有限，大大加深了訓練的困難度。在本次的實驗中我們可以觀察到我們的模型大多發生high-variance的問題，準確度 validation的資料遠低於train，因此我們的模型產生過度擬和，雖然我們有使用各種regulation的方法，包含dropout、early stoping、l1及l2但都不見有明顯的改善問題，雖然l2會使模型訓練速度加快，且l2的參數不能給太大，大約0.001會比較好。最後推論是因為原先為了colab的訓練只能調batch設1，且模型太深也會超過colab記憶體限制，但如果太多資料又要訓練太久(10hr)，不過經過推測應該是我的原本資料複雜度高，加上訓練的圖片不夠多(100張做資料擴增)所導致的。目前正在嘗試使用但結果也不見得比較好，最後目前測試到500張，訓練時間就達2個小時，目前雖然train的曲線與validation的曲線較近，但準確度 <0.5 ，仍然發生high-bias的問題，因為在很早就發生early stopong，且loss已經無法學習，可能走到平坦區。可能尚待更多訓練時間來調整模型，雖然有點可惜，但希望此次經驗可以記取失敗的地方使期末專案更容易一些。至於High-bias的問題，也正在嘗試添加BatchNormalization層，試著讓模型走到平坦區可以繼續學習。