

國立清華大學

智慧化企業整合

Project 3

以深度學習為基礎之  
動物種類辨識及分類

指導教授：邱銘傳 博士

110034554 徐一正

中華民國一一一年一月六日

# 目錄

一、背景介紹.....	1
1.1 問題背景.....	1
1.2 動機與目的.....	1
1.3 問題描述(5W1H).....	2
1.4 資料來源.....	2
二、深度學習模型.....	4
2.1 資料前處理.....	4
2.2 ResNet50.....	5
2.3 模型訓練結果.....	7
三、分析.....	9
3.1 實驗設計.....	9
3.2 L <sub>9</sub> 直交表.....	9
3.3 Minitab 結果.....	10
四、結果與討論.....	12
4.1 結論.....	12
4.2 未來展望.....	12

## 圖目錄

圖 1、動物種類之圖片數量分布.....	3
圖 2、應用 resize 之方式.....	4
圖 3、像素值標準化.....	4
圖 4、資料增強.....	5
圖 5、ResNet50 整體架構圖.....	5
圖 6、ResNet50 引入之新架構.....	6
圖 7、模型架構之程式碼.....	6
圖 8、模型訓練 Acc 圖.....	7
圖 9、模型訓練 Loss 圖.....	8
圖 10、模型測試結果.....	8
圖 11、Main Effects Plot for Means.....	10
圖 12、Main Effects Plot for SN ratios.....	10
圖 13、調整後之模型訓練 Acc 圖.....	11
圖 14、調整後之模型訓練 Loss 圖.....	11

## 表目錄

表 1、5W1H.....	2
表 2、三資料集比例與數量.....	5
表 3、模型超參數設計.....	7
表 4、模型準確率彙整表.....	8
表 5、實驗設計表格.....	9
表 6、L <sub>9</sub> 直交表.....	9
表 7、調整後模型之準確率彙整表.....	11

# 一、背景介紹

## 1.1 問題背景

於 2018 年，包含哈佛大學、牛津大學、明尼蘇達大學在內的多所大學研究人員共同合作，訓練以深度學習為基礎的系統，共辨識、分類 48 種動物，例如大象、長頸鹿和羚羊等。該研究所使用之數據集為 320 萬張野生動物圖片，其來源為位於坦尚尼亞塞倫蓋提國家公園的隱藏式相機所拍攝的照片。團隊目標準確率 99%，目前技術可達 9 成左右，尚有一段差距。此外，相關技術不僅可以應用於研究目的，若有機會增加應用廣度及範圍，可望為攝影師、生態保育相關工作人員、甚至欲整理手機相簿的一般民衆等帶來便利。

## 1.2 動機與目的

- 節省大量時間

蒐集研究用照片的成本已大幅降低，但人類卻通常仰賴人工分類，耗費大量時間和經歷分類圖片後才能做後續研究。

- 方便專家研究

藉由各式相關模型之發展讓生物學、生態學和動物學等許多相關領域轉變為“大數據”科學，輔助專家找出研究領域之更多關鍵，也方便專家執行研究。

- 改善生態保育能力

在動物經過鏡頭時實時辨識、追蹤，有助於提供更完善的生態保育能力，尤其可應用於大範圍的野生動物棲息地。

- 擴大應用範圍及普及程度

即使非前述之專業人士所需，亦可應用此技術於協助動物攝影師分類、生態保育相關工作者、甚至一般民衆分類手機照片等。

### 1.3 問題描述(5W1H)

基於前述之動機及目的，本研究欲進一步釐清問題的本質，以 5W1H 之架構深入分析此問題相關之根本原因、時間、對象、地點、事件、方法等。透過此分析，本研究瞭解此問題之根本原因為現有作業耗費大量時間、人力及生態保育能力尚待進步等，且研究之相關技術有更廣泛的應用價值及意義，可望進一步增加此技術的普及程度；此問題之時間為相關研究需要時、整理手機相簿時等；此問題之對象可能為相關研究人員、攝影師、以手機拍照的一般民衆等；此問題之地點可能為野外動物棲息地、國家公園、動物園等處。為解決上述問題，可以藉由深度學習之方法，以蒐集之照片為基礎，針對各個不同的動物種類進行辨識及分類。

本研究將 5W1H 分析之架構整理如表 1 所示：

表 1、5W1H

Why	耗費大量時間、人力及生態保育能力尚待進步
When	相關研究需要時、整理手機相簿時
Who	相關研究人員、攝影師、以手機拍照的一般民衆
Where	野外動物棲息地、國家公園、動物園
What	針對多種動物種類進行辨識及分類
How	以深度學習之方法辨識並分類照片

### 1.4 資料來源

於 Kaggle 網站下載 Animal Images Dataset 資料集，其中狗、貓、兔子、鳥類、魚類、老鼠等六個常見動物種類之各種圖片。本資料集共有 11959 張動物圖片，而資料集中六個動物種類之圖片數量分布如圖 1 所示：

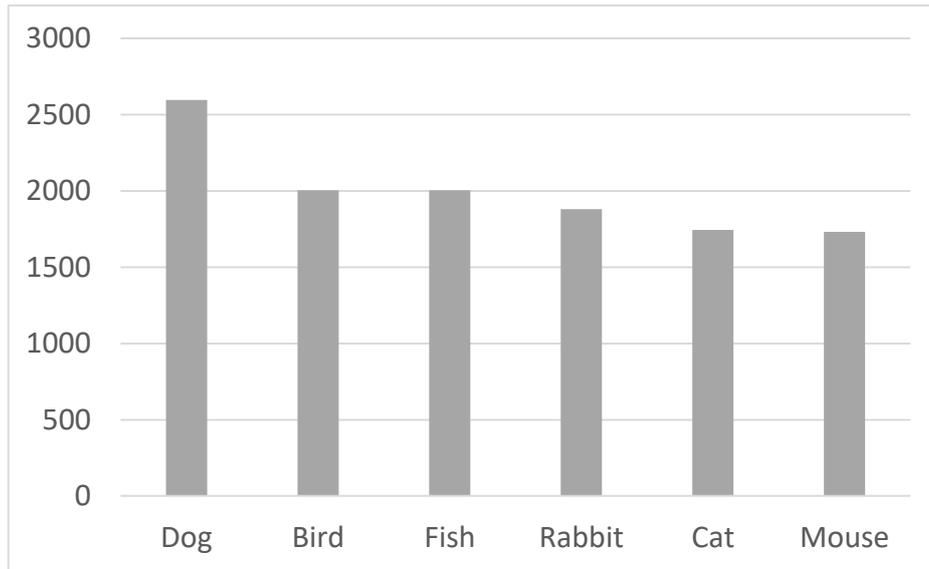


圖 1、動物種類之圖片數量分布

## 二、深度學習模型

### 2.1 資料前處理

本研究得到資料集後首先檢查資料集的內容，由於少數圖片過於模糊，導致幾乎無法辨識，因此刪除此部分內容。而於原標籤出現錯誤的情況，本研究予以修正，確認最終之資料集包含的圖片皆正確且清晰。

由於後續模型之 input 欲使用 224x224 之尺寸的圖片，因此本研究使用 resize 的方式使資料集中每張圖片的大小統一。使用之方式如圖 2 所示：

```
def preprocess_image(image_path, train_data = False):  
  
    # Resize the image to our desired value  
    image = tf.image.resize(image, size = IMAGE_SIZE)  
  
    return image
```

圖 2、應用 resize 之方式

本研究將圖片之像素值進行標準化，意即原圖片之像素值除以 255，以使得其像素值從原本的 0~255 等比例縮放至 0~1，以利後續模型使用，增快運算速度。使用之方式如圖 3 所示：

```
def preprocess_image(image_path, train_data = False):  
    # Read the image  
    image = tf.io.read_file(image_path)  
  
    # 將原圖像像素值除以255，以等比例縮放至0-1之間  
    image = tf.image.decode_jpeg(image)
```

圖 3、像素值標準化

本研究將原資料集進行資料切割，分為訓練集、驗證集、測試集三個部分，資料集圖片之數量比例為 80%、10%、10%。三資料集之與相對比例與圖片數量以表 2 所示：

表 2、三資料集比例與數量

資料集 \ 項目	比例	數量
Train set	80%	9567
Val set	10%	1196
Test set	10%	1196

為了避免圖片數量不足，本研究利用 Keras ImageDataGenerator 將資料集之圖片進行旋轉、縮放與移動，增加圖片張數，其中包含利用套件 One-hot encoding 自動將資料之標籤進行有效編碼，以利後續模型訓練。使用之方式如圖 4 所示：

```
# 創建ImageDataGenerator, 包含Data Augmentation
train_datagen = ImageDataGenerator(rotation_range=10, zoom_range=0.1, width_shift_range=0.1, height_shift_range=0.1)
train_batches = train_datagen.flow_from_directory(TRAIN_PATH,
                                                target_size=IMAGE_SIZE,
                                                interpolation='bicubic',
                                                class_mode='categorical',
                                                shuffle=True,
                                                batch_size=BATCH_SIZE)
```

圖 4、資料增強

## 2.2 ResNet50

本研究使用之深度學習模型為 ResNet50，其模型架構如圖 5 所示：

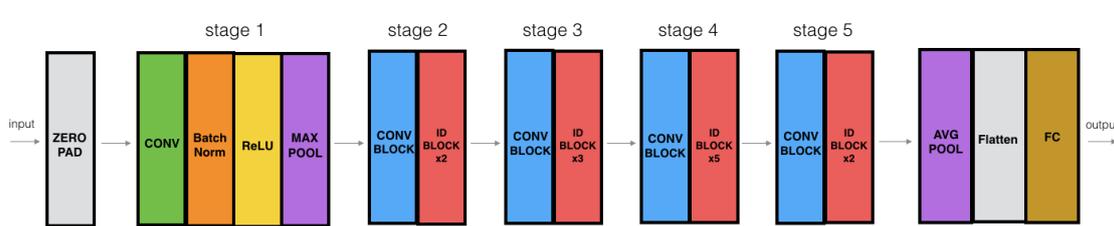


圖 5、ResNet50 整體架構圖

ResNet50 於 2015 年被提出，在 ImageNet 比賽 classification 任務上獲得第一名，其後許多方法都建立在 ResNet50 的基礎上完成。

依照 ImageNet 比賽的歷屆冠軍資料，可以發現網路層數不斷增加，從 2012 年至 2017 年持續上升。由此可知在沒有 Overfitting 的情況下，加深網路可以帶來好的效果。而欲持續提升模型的效果，隨網路層數不斷加深，隨之可能帶來其他問題，例如隨著層數增加，模型準確率接近飽和且開始不穩定。這往往不是因為 Overfitting 所造成，而是因為網路疊了太多層，導致過高的訓練誤差。為了減輕訓練較深神經網路的困難度，作者提出了一種全新的網路，將殘差網路引進模型中，稱之為深度殘差網路，其引入的架構如圖 6 所示。即使網路沒有學習到新的特徵，也保證不會退化，允許模型得以加深，保持複雜度的同時亦保有實用性及穩定性，使得此模型逐漸在檢測、分類、識別等領域得到廣泛應用。

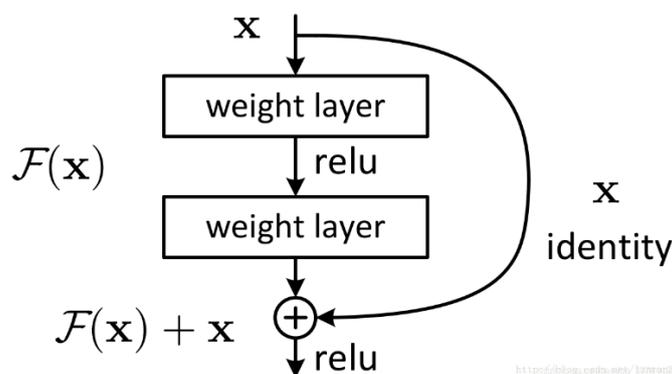


圖 6、ResNet50 引入之新架構

模型架構之程式碼如圖 7 所示：

```
net = ResNet50(include_top=False, weights='imagenet', input_tensor=None,
              input_shape = (IMAGE_SIZE[0], IMAGE_SIZE[1], 3))
x = net.output
x = Flatten()(x)
x = Dropout(0.5)(x)
output_layer = Dense(NUM_CLASSES, activation = 'softmax', name='softmax')(x)
net_final = Model(inputs = net.input, outputs = output_layer)
for layer in net_final.layers[:FREEZE_LAYERS]:
    layer.trainable = False
for layer in net_final.layers[FREEZE_LAYERS:]:
    layer.trainable = True
net_final.compile(optimizer = Adam(learning_rate = 1e-5), loss = 'categorical_crossentropy', metrics = ['accuracy'])
print(net_final.summary())
```

圖 7、模型架構之程式碼

模型超參數初步設計如表 3 所示：

表 3、模型超參數設計

參數	設定
Image size	(224,224)
Epoch	7
Batch size	32
Dropout	0.5
Activation function	Relu, Softmax
Optimizer	Adam
Learning rate	1e-5

### 2.3 模型訓練結果

模型訓練之結果如圖 8 至圖 9 所示，而訓練、驗證、測試之準確率如表 5 所示。

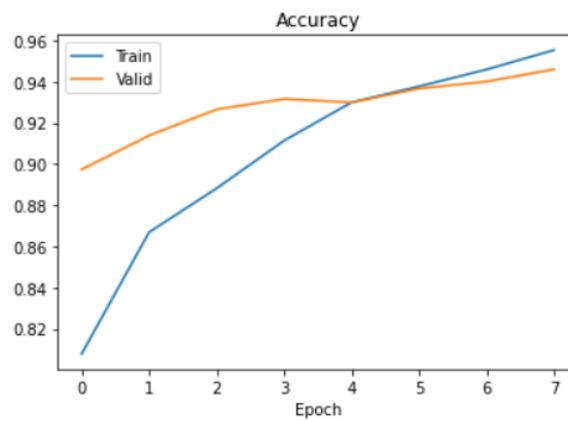


圖 8、模型訓練 Acc 圖

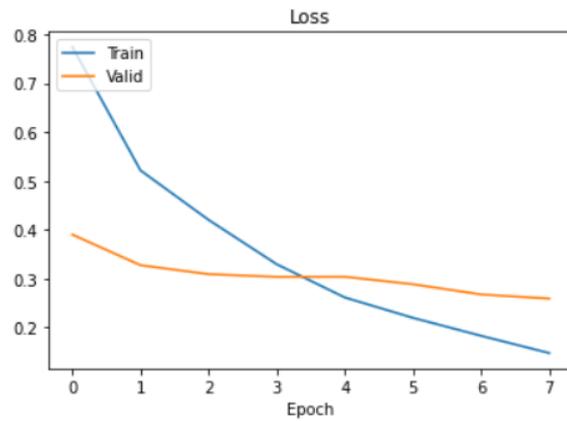


圖 9、模型訓練 Loss 圖

表 4、模型準確率彙整表

資料集 \ 準確率	準確率
<b>Train</b>	0.95
<b>Val</b>	0.93
<b>Test</b>	0.87

模型之測試結果如圖 10 所示，成功分辨測試圖片中的動物種類。



圖 10、模型測試結果

## 三、分析

### 3.1 實驗設計

本研究欲進一步調整超參數以優化模型，使用 3 因子 3 水準之實驗設計，其中選定之三個因子分別為 Epoch、Dropout、Batch Size。實驗設計之表格如表 5 所示：

表 5、實驗設計表格

Factor Level	Epoch	Dropout	Batch Size
Level 1	6	0.4	32
Level 2	8	0.5	64
Level 3	10	0.6	128

### 3.2 L<sub>9</sub> 直交表

由於使用 3 因子 3 水準之實驗設計，因此本研究採用 L<sub>9</sub> 直交表，各實驗之設計如表 6 所示。此直交表設計完成後，依照各實驗之超參數設計分別訓練模型。

表 6、L<sub>9</sub> 直交表

Factor Experimant	Epoch	Dropout	Batch Size	Accuracy
1	6	0.4	32	0.92
2	6	0.5	64	0.93
3	6	0.6	128	0.92
4	8	0.4	64	0.88
5	8	0.5	128	0.9
6	8	0.6	32	0.87

7	10	0.4	128	0.89
8	10	0.5	32	0.9
9	10	0.6	64	0.89

### 3.3 Minitab 結果

將各個超參數組合的模型訓練結果輸入 minitab 軟體中，分析之結果如圖 11、圖 12 所示，由此可以看出選定之各超參數（因子）對模型準確率的影響力及最佳超參數組合。

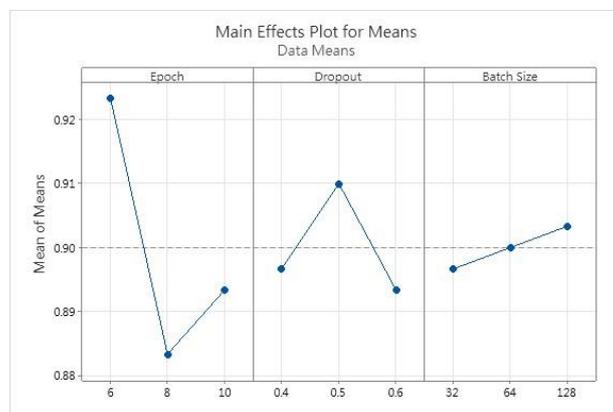


圖 11、Main Effects Plot for Means

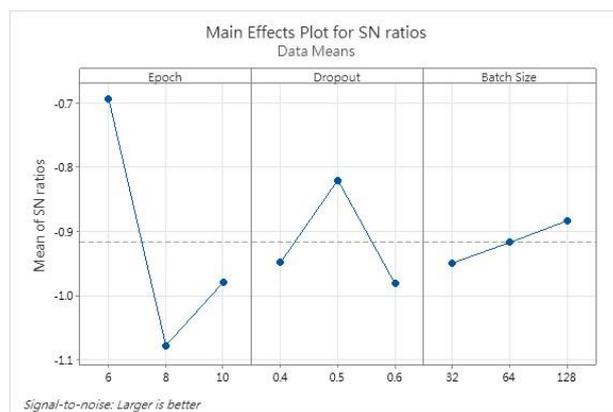


圖 12、Main Effects Plot for SN ratios

本研究選定最終的模型超參數組合，Epoch 為 6，Dropout 為 0.5，Batch Size 為 128，進一步調整並優化模型。調整後模型之訓練結果如圖 13、圖 14 所示：

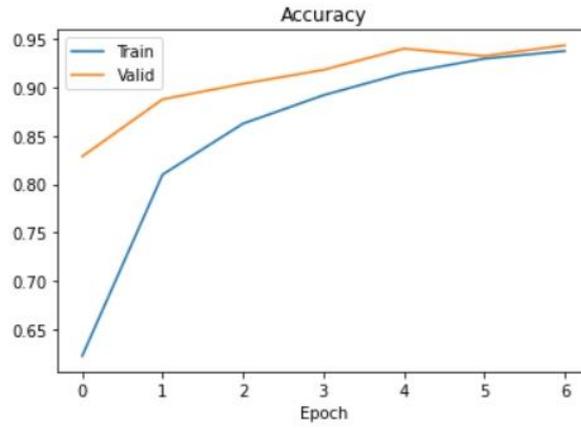


圖 13、調整後之模型訓練 Acc 圖

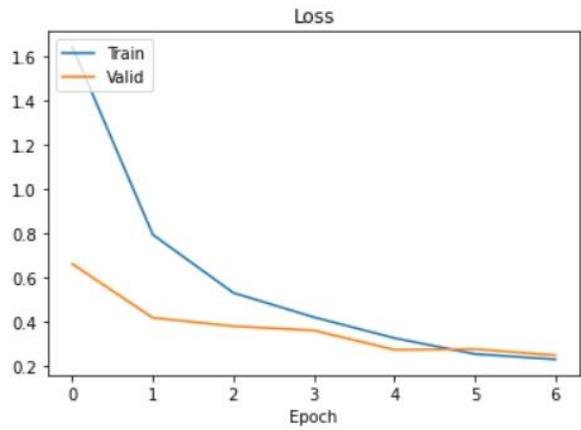


圖 14、調整後之模型訓練 Loss 圖

調整後模型之訓練、驗證、測試的準確率如表 7 所示：

表 7、調整後模型之準確率彙整表

資料集 \ 準確率	準確率
Train	0.94
Val	0.94
Test	0.93

## 四、結果與討論

### 4.1 結論

- 以現有資料集之多個種類的動物進行辨識並分類，準確率達一定水準，經超參數優化後使得準確率可達 93% 以上。
- 應用實驗設計的方法求得模型所需參數之最佳組合，亦可使用 minitab 等軟體找出重要因子。
- 模型達初步應用目的，完全於現實情境中執行仍需強化及調整。

### 4.2 未來展望

- **增加資料類別**

由於動物種類繁多，未來可蒐集其他類別資料，加入更多種類的動物圖片，使模型更完善。

- **擴大豐富程度**

現階段資料僅以大項區分，然各種動物又可細分成多樣品種，可針對不同品種擷取特徵，擴增其豐富程度。