

Intelligent Integration of Enterprise

Project 3

鳥類百科-鳥類圖像分類器

Group 5

109034551 林奕辰

指導教授：邱銘傳 教授

鳥類百科-鳥類圖像分類器

清華大學工業工程與工程管理學系 林奕辰

摘要

鳥類為日常生活中最常見的脊索動物之一，無論在公園中或校園中都可以輕易的發現鳥類的蹤跡，然而因其種類繁多故正確識別鳥類品種為困難之任務。為了增進民眾對於鳥類的興趣本研究將建立基於人工智慧的鳥類分類器，藉由卷積神經網路識別圖片，並藉由優化參數提高模型準確度。

關鍵字：圖片辨識、鳥類、卷積神經網路

1、背景.....	4
1.1 前言.....	4
1.2 5W1H	4
1.3 研究流程.....	5
2、文獻回顧	6
2.1 卷積神經網路(Convolutional Neural Network, CNN)	6
3、研究架構及方法	6
3.1 資料特徵.....	6
3.2 資料前處理	8
3.3 模型建立.....	9
4、模型訓練及測試.....	11
1. X1.....	11
2. X2.....	11
3. X3.....	12
4. X4.....	12
4. X5a.....	13
4、結果與討論	14
4.1 實際測試.....	14
4.2 未來展望.....	15
參考文獻.....	15

1、背景

1.1 前言

鳥類為日常生活中最常見的脊索動物之一，無論在公園中或校園中都可以輕易的發現鳥類的蹤跡，據調查台灣總共記錄到 87 科 674 種鳥類，且賞鳥活動日益盛行，但是大多數民眾對於野鳥認識甚少，本研究希望藉由建立鳥類百科增進民眾對鳥類的認識，本研究將以 CNN model 建立鳥類識別器，幫助使用者藉由影像辨識鳥類。除了增進一般民眾對於野鳥的知識外，本研究成果也可以使用於學術單位或是與野鳥保育的相關單位，可以串聯野外攝影機的影像資料，及時且全天候的分析獲得的影像資料。本研究使用 Kaggle 上取得之鳥類圖像資料庫進行卷積神經網路 (CNN, Convolutional Neural Networks) 訓練，並藉此建立鳥類圖像分類器。



圖 一、資料集圖片範例

1.2 5W1H

在分類鳥類時最直接的方式就是藉由外觀特徵進行分類，但是由於鳥類種類繁多，導致一般民眾若缺乏相關知識時難以分辨鳥類品種。而學術單位在研究野鳥生態時會使用野外攝影機，若是可以藉由程式自動判斷影像中的鳥類可以大幅提高工作效率。

表 一、5W1H

Who	一般民眾、學術單位。
What	鳥類品種之鑑定
When	需要辨識鳥類品種時
Why	鳥類種類繁多而且特徵難以辨識
How	電腦視覺、機器學習

1.3 研究流程

本研究將自 Kaggle 取得資料集，並對資料進行初步分析。再使用資料進行模型的訓練以及驗證，並以測試及評估模型的績效。藉此流程不斷修正模型以達到參數優化提升準度的目標。

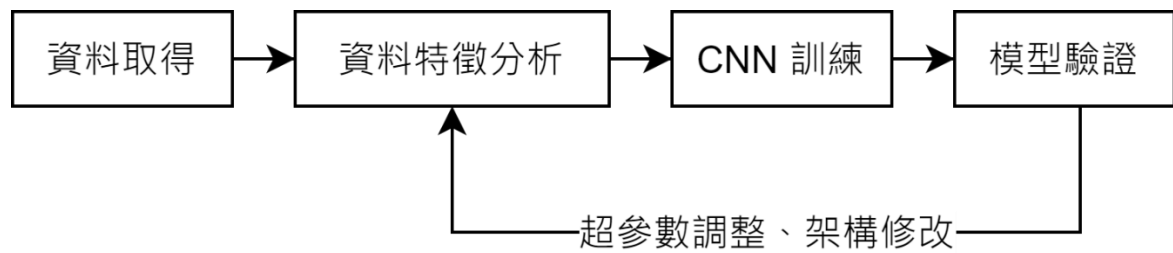


圖 二、研究流程

2、文獻回顧

2.1 卷積神經網路(Convolutional Neural Network, CNN)

卷積神經網路 Convolutional Neural Networks (CNN) 是一深度學習的演算法，可用於電腦視覺自動辨識、影像分類等應用上，是一以標籤資料進行模型訓練與驗證的監督式學習法。CNN 卷積神經網路模型，訓練完成後可以藉由擷取與影像形狀相關的特徵進行影像辨識(Ketkar & Santana, 2017)。CNN 模型會從數據中自行學習特徵的識別，有別於過往的傳統影像識別方法，需要密集的進行特徵工程。CNN 的基本結構會包含多個交替堆疊的卷積層以及池化層，卷積層會使用多個濾鏡 (kernels) 幫助模型擷取圖像特徵，這種構造使 CNN 架構在參數的使用上更有效率，也讓 CNN 架構的神經網路有高準確度的表現。而池化層可以降低特徵圖的維度讓網路的計算複雜度降低，可以更快的得到運算結果。本研究將使用基於卷積神經網路架構的模型實作圖像分類器，用以達成鳥類分類的用途。

3、研究架構及方法

3.1 資料特徵

本資料集使用 Kaggle 之 250 Bird Species data set，資料集被劃分為訓練集、驗證集以及測試集。資料集最新版本中包含 235 種鳥類的彩色圖片。訓練集中有 32672 張訓練用圖片，驗證集和測試集則各有 1175 張影像。

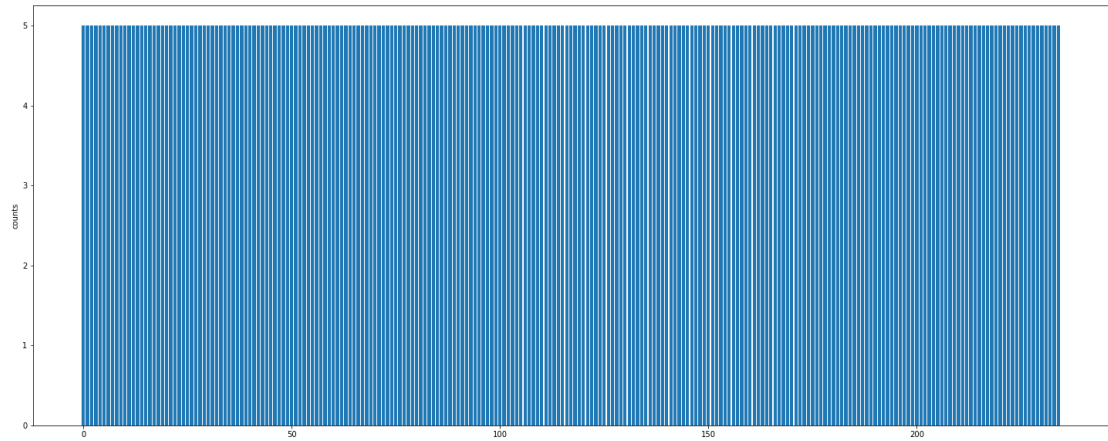


圖 三、驗證集資料分布

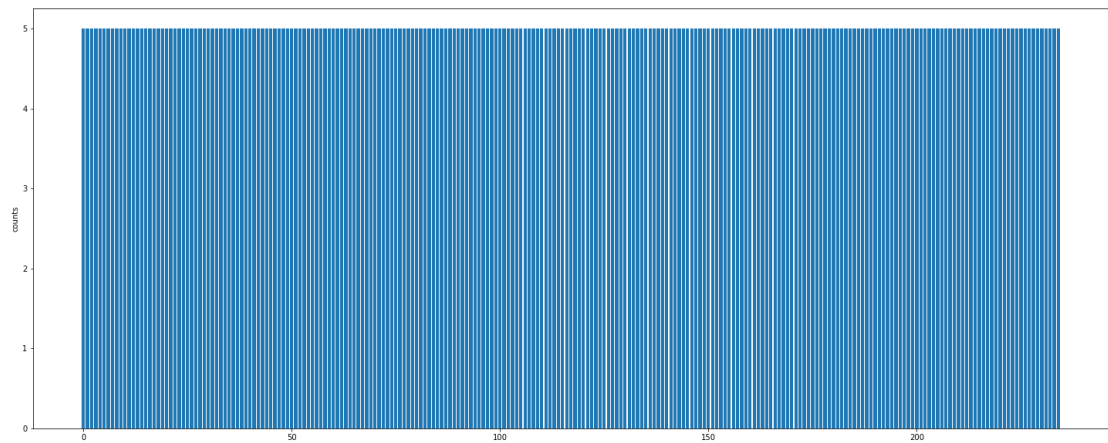


圖 四、測試集資料分布

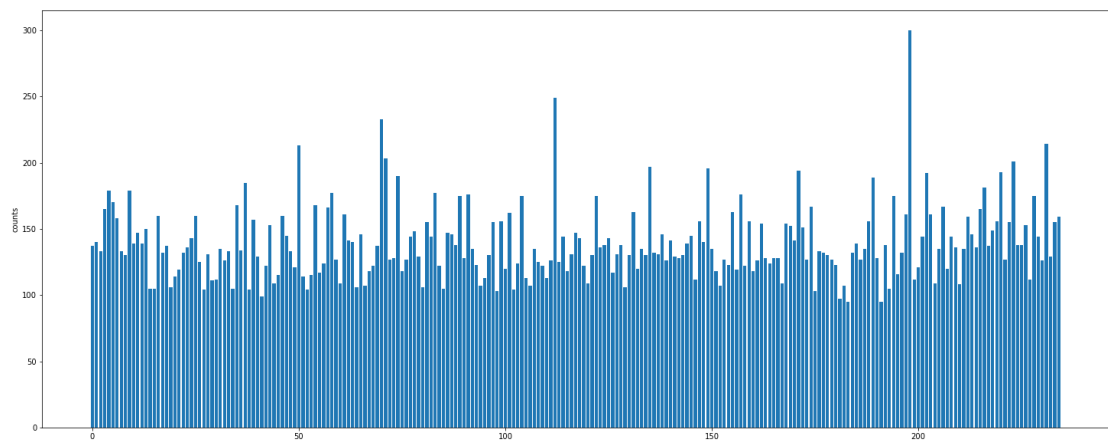


圖 五、訓練集資料分布

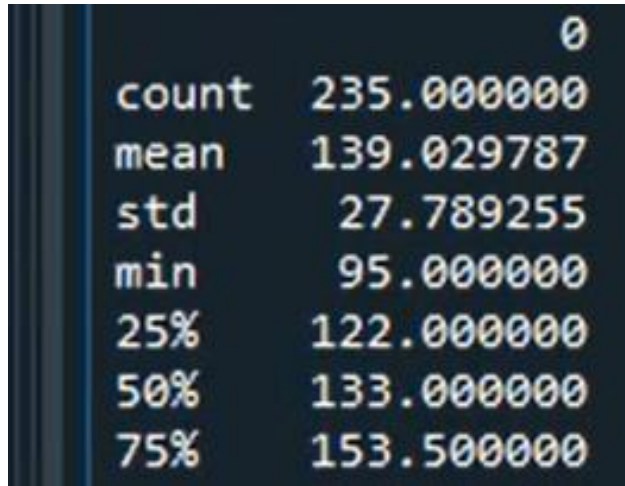


圖 六、訓練集統計資料

從以上圖中可以得知，測試集和驗證集中各鳥類圖片都有 5 張，而訓練集中的各種別圖片數量最少有 95 張，最多的有 300 張，數量之間的差異並不極端，人工檢查後資料集中圖片大致正常，故直接使用本資料訓練模型。

3.2 資料前處理

先將所需資料放置到所屬資料夾，使用 Keras 的功能將資料從資料夾中讀出，將圖片轉為(150,150,3)尺寸的 tensor，並將色階值均一化，類別 encoding 使用 sparse。此方式會將類別對應為一整數，如{'AFRICAN CROWNED CRANE': 0,'AFRICAN FIREFINCH': 1,'ALBATROSS': 2,}，程式碼以及轉換後的圖如下圖，右邊的亞歷山大鸚鵡畫質明顯較低畫素為 150*150。影像色階均一化可以使模型因為輸入值浮動較小，可以更準確的預測影像類別。

```

train_datagen=ImageDataGenerator(rescale=1/255)
val_datagen=ImageDataGenerator(rescale=1/255)
test_datagen=ImageDataGenerator(rescale=1/255)

train_generator=train_datagen.flow_from_directory(train_dir,
                                                  target_size=(150,150),
                                                  color_mode='rgb',
                                                  class_mode='sparse',batch_size=32,shuffle=True)

val_generator=val_datagen.flow_from_directory(validation_dir,
                                              target_size=(150,150),
                                              color_mode='rgb',
                                              class_mode='sparse',batch_size=32,shuffle=True)

test_generator=test_datagen.flow_from_directory(test_dir,
                                                target_size=(150,150),
                                                color_mode='rgb',
                                                class_mode='sparse',batch_size=32)

```

圖 七、圖片前處理



圖 八、處理後影像範例

3.3 模型建立

本次選用 Xception net 的架構，並在輸出的部分加入自定義的 layers 並調整超參數以達到提高模型準度的目標。下表為該模型的基本資料，準確度為使用 ImageNet validation dataset 測試之結果。

表 二、模型基本資料

Model	size	Top-1 acc	Top-5 acc	Parameters	depth
Xception	88MB	0.79	0.945	22910480	126

為了優化模型本研究測試了不同超參數以及激活函數的組合，其組合以及績效如下表所示。由於本次使用的 encoding 為 sparse encoding 故損失函數使用 sparse categorical cross entropy 計算。

model	x1	x2	x3	x4
base	xception			
epoch	20	20	20	5
0	global_max_pooling2d			
1	flatten			
2	Dense 235 relu	Dense 235 sigmoid	Dense 1024 relu	Dense 1024 elu

3

Dense 235 softmax

loss	sparse_categorical_crossentropy			
optimizer	adam	adam	adam	adagrad
test acc	0.9523	0.9549	0.9609	0.9787

x5a

base	xception
epoch	20
0	global_max_pooling3d
1	flatten
2	Dense 1024 elu
3	Dense 235 softmax
loss	sparse_categorical_crossentropy
optimizer	adagrad
test acc	0.9898

4、模型訓練及測試

1. X1

X1 模型在 flatten 層後連接 235 個節點的全連結層激活函數為 relu (Dense 235 relu)，之後連接 235 個節點的全連結層激活函數為 softmax (Dense 235 softmax)，optimizer 為 adam。訓練結果如下圖所示。本模型的測試集準確度為 **0.9523%**。

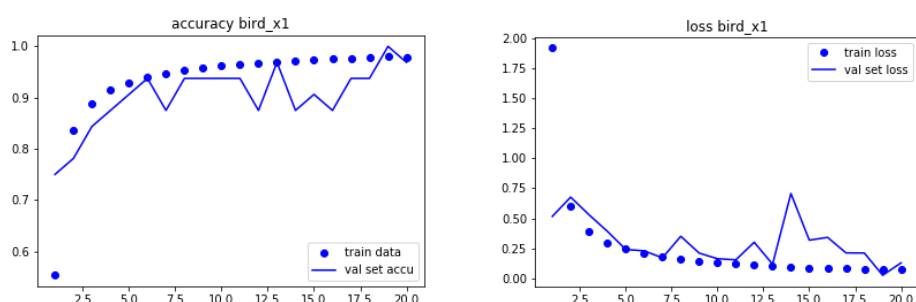


圖 九、X1 訓練績效

2. X2

X1 模型在 flatten 層後連接 235 個節點的全連結層激活函數為 sigmoid (Dense 235 sigmoid)，之後連接 235 個節點的全連結層激活函數為 softmax (Dense 235 softmax)，optimizer 為 adam。訓練結果如下圖所示。本模型的測試集準確度為 **0.9549%**。在更換激活函數後準度並無明顯提高，推測是 flatten 層後的全連結層寬度不足造成資訊瓶頸，使得重要參數遺失，若想提升準度可以加寬此層。

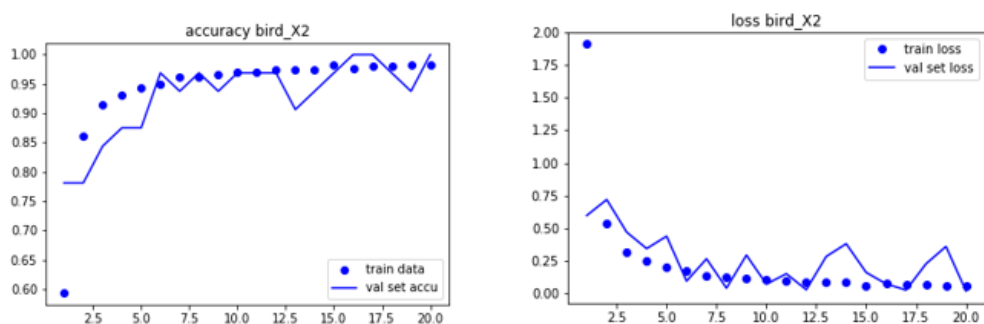


圖 十、X2 訓練績效

3. X3

X3 模型在 flatten 層後連接 1024 個節點的全連結層激活函數為 relu (Dense 1024 relu)，之後連接 235 個節點的全連結層激活函數為 softmax (Dense 235 softmax)，optimizer 為 adam。訓練結果如下圖所示。本模型的測試集準確度為 0.9609%。在加寬全連結層之後準確度有略微提升。

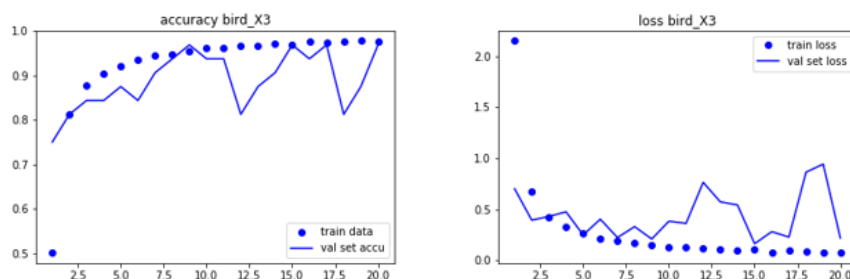


圖 十一、X3 訓練績效

4. X4

X4 模型在 flatten 層後連接 1024 個節點的全連結層激活函數為 elu (Dense 1024 elu)，elu 可以避免梯度為 0 導致神經元死亡情形，並可以更快的訓練，之後連接 235 個節點的全連結層激活函數為 softmax (Dense 235 softmax)，optimizer 為 adagrad。本次訓練時由於模型收斂速度快，為了避免過度擬合只訓練 5 個世代就結束訓練。本模型的測試集準確度為 0.9787%，較前述參數組合高出不少。

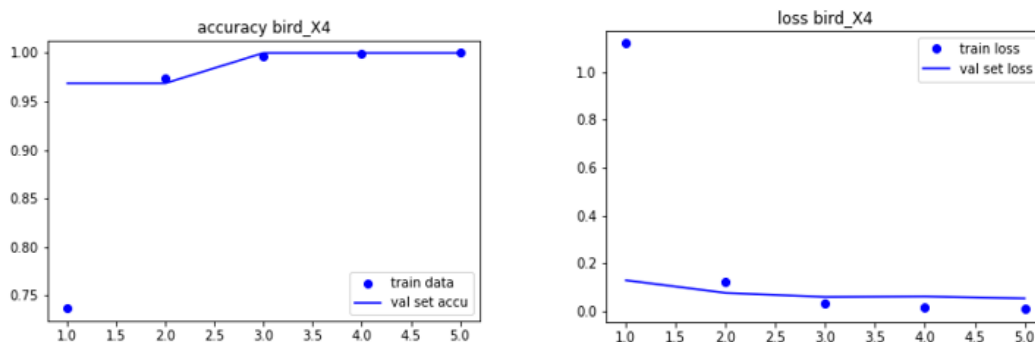


圖 十二、X4 訓練績效

4. X5a

本模型沿用 X4 模型的架構但是為了避免過擬，在資料輸入時會對影像進行擴增，隨機對輸入影像進行旋轉、位移、縮放或是翻轉，擴增後結果如圖十三。使用擴增後的資料訓練可以避免過度擬合，並可以增加模型的泛化能力，在測試集中可以有較佳表現。由於可以避免過度擬合本次實驗訓練 20 個世代。本模型的測試集準確度為 0.9898%，為所有模型中準確度最佳的。

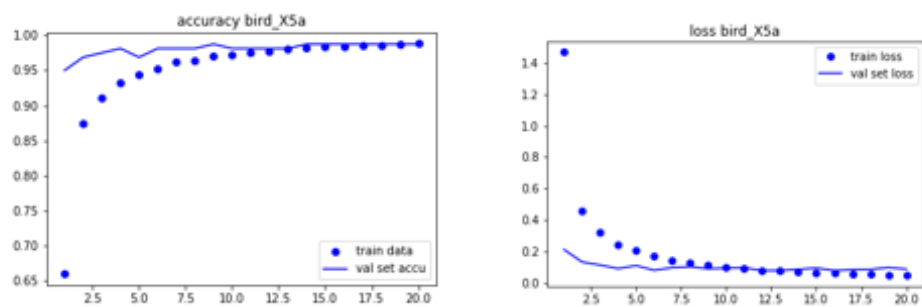


圖 十三、X5a 訓練績效

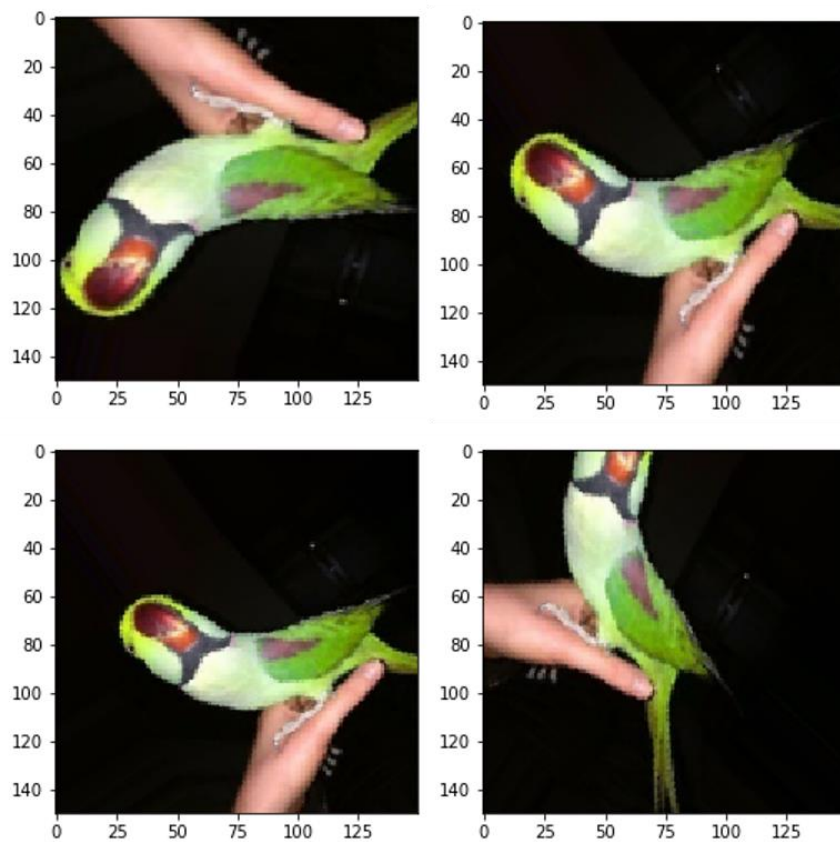
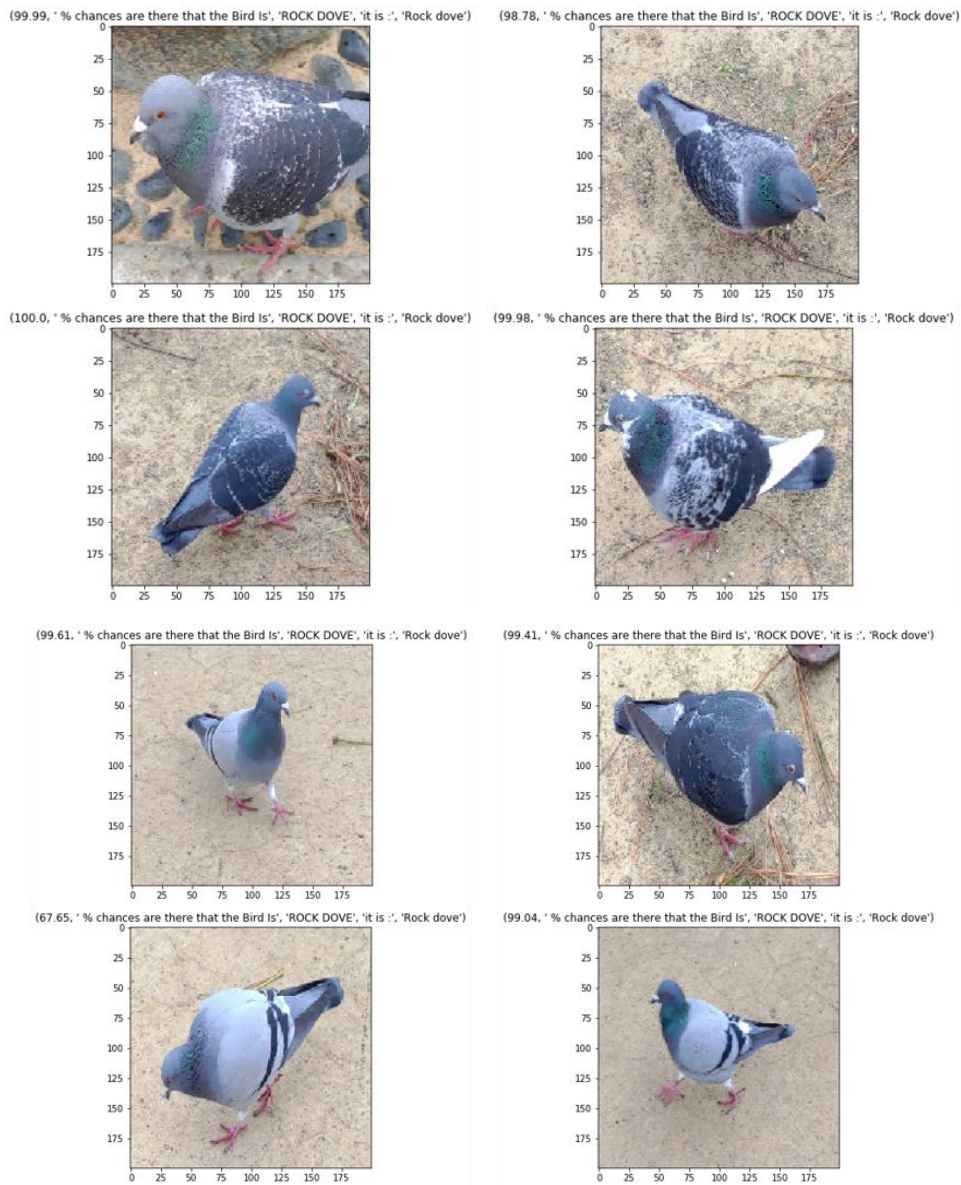


圖 十四、影像擴增後的圖像

4、結果與討論

4.1 實際測試

以下圖片為在清大校園內拍攝的鴿子，使用績效最好的模型進行判斷，模型將以下照片都判斷為鴿子。因為本次使用資料集包含多國鳥類，所以只有少部分類別是在清大校園內找到，故本次測試時只有使用鴿子的圖像。由測試結果中可以看出本模型有實用化的潛力。



4.2 未來展望

經過本次實驗建構出可辨識鳥類種類的神經網路，而在針對此一任務有不錯的準度，而未來希望可以使用本土鳥類的資料集進行訓練，讓本模型可以支援本土鳥類的辨識。也希望可以建立藉由聲音辨識鳥類種類的模型，由於多數時候在野外環境，因鳥類躲藏習性難以直接獲得鳥類照片，如可以藉由聲音辨識鳥類，則可以協助研究單位利用戶外收音設備長時間的觀察棲地中的鳥類。

參考文獻

Ketkar, N., & Santana, E. (2017). *Deep Learning with Python* (Vol. 1): Springer.