國立嘉義大學應用數學系

專題製作

Department of Applied Mathematics

National Chiayi University

Project on Mathematics

運用深度學習模型辨識

經傅立葉處理之圖片

學 生：陳範耘、潘陣恩

李涵碩、黃尚崴

指導老師：鄭富國 老師

中 華 民 國 112 年 1 月

**目錄**

[壹、序論3](#_壹、序論)

[一、研究動機及目的3](#_一、研究動機及目的)

[二、致謝5](#_二、致謝)

[貳、研究內容7](#_貳、研究內容)

[一、選擇及使用工具7](#_一、選擇及使用工具)

1. [Python7](#_1、Python)
2. [PyCharm8](#_2、PyCharm)
3. [Spyder9](#_3、Spyder)
4. [TensorFlow11](#_4、TensorFlow)
5. [CIFAR-10012](#_5、CIFAR-100)
6. [VGG 1613](#_6、VGG-16)

[二、開發環境15](#_二、開發環境)

1. [模型訓練之環境15](#_1、模型訓練環境)
2. [主程式執行環境15](#_2、主程式執行環境)

[三、深度學習及卷積神經網路16](#_三、深度學習及卷積神經網路)

1. [深度學習16](#_1、深度學習)
2. [卷積神經網路18](#_2、卷積神經網路)

[四、傅立葉轉換26](#_四、傅立葉轉換)

1. [一維傅立葉轉換26](#_1、一維傅立葉轉換)
2. [二維傅立葉轉換27](#_2、二維傅立葉轉換)

[五、Python程式29](#_五、Python程式)

1. [CNN模型29](#_1、CNN模型)
2. [主程式35](#_2、主程式)

[參、結論41](#_參、結論)

[一、執行結果41](#_一、執行結果)

[二、結果分析43](#_二、結果分析)

[三、未來展望48](#_三、未來展望)

[肆、參考文獻50](#_肆、參考文獻)

[伍、附錄51](#_伍、附錄)

# 壹、序論

## 一、研究動機及目的

自從AI在圍棋上擊敗人類好手後，AI 開始為人所知，不再是從前只能在科幻電影中才能接觸到的高科技產品，而是開始慢慢走進了我們的生活。也因為如此，數學系的應用之一 — 大數據也開始逐漸成為人們討論的話題，有鑑於此，我們想從人工智慧基礎的圖像辨識入門，學習機器學習、深度學習等等的 AI基礎，並嘗試把我們在應用數學系這將近四年的時間裡，學習到的各種數學專業知識，融入我們的專題中，讓我們人工智慧的專題中出現充滿數學與資訊的內容。

在開始研究之前，我們以為這是一個可以以我們現有的能力直接觸及的知識，而在開始研究之後，才發現越深入研究我們越難以負荷，為此我們花費了更多的時間完備了許多資訊相關的專業科目，也因為這樣，我們更確定接下來的道路正是我們的目標。

也是在開始研究之後，我們發現電腦在做如此大量的運算時，會耗費許多時間，所以目前大多數的訓練資料集都是使用低解析度的圖片來進行訓練；但在進行測試時，我們仍會用到高解析度的照片來進行訓練及測試，提高我們的模型準確率以及辨識能力。為了解決高解析度的圖片帶來的負荷，我們對圖片進行傅立葉轉換的加工，會刪除圖片所帶有的不必要的資訊，所以我們想要觀察加工後的圖片會對神經網路造成何種影響。

我們決定使用深度學習中的一個深度神經網路—卷積神經網路來製作我們的視覺辨識模型，使用CIFAR-100來訓練模型，並引入VGG16來完備我們的演算法，最後使用至少100張的不同圖片來測試結果，並分別比較有無經過二維傅立葉轉換處理的圖片，送入模型預測的結果以及各種參數差異。

## 二、致謝

感謝帶領著我們專題的鄭富國老師，有鑑於我們的應用數學系上，目前沒有老師是走人工智慧專業的方向，但富國老師還是同意了我們的專題題目，並從旁提供我們寶貴的建議以及適時的幫助，給予學生們非常大的空間撞擊我們的意見、揮灑我們的創意，我們對您的感謝無以言表，在讓我們深摯地向老師致謝之前，我們對這四年來的經歷有些話想先說說。

在大學一開始就讀應用數學系時，讓我們對於應用數學的幻想有嚴重的破滅，原來應用數學只是單純改了名子的純數學系，在茫茫數學的定理、公設下，看不盡那微積分、高等微積分、微分方程等等專業科目的背後，應用數學系真正的「應用」學習從何時到來；也因此，我們學生中有許多同學對持續鑽研單純的數學失去了信心及喜愛，在漫漫三年好課值得一修再修的洗禮下，淡忘了初衷，直至看到專題，才想起當初對應用數學的幻想，又才堅定「無論如何也要把專題做到更好」，以此不愧對我們在數學系就讀的這四年。

那為什麼突然發表以上的經歷呢？其實和我們對富國老師的感情有很大的關係。我們這屆大一下修老師的離散數學必修課，還記得在那時候，富國老師向我們解釋了密碼學的原理與RSA演算法所運用到的數學知識，當時，這些知識宛如夜空中的流星，劃入了自從大一以來，腦部被純數扭曲搥打而乾涸的小溪，「唰！」那一刻透過「數學」鑿壁的山區流入「應用」的源泉，我們對數學的幻想透過富國老師的啟發而萌芽。自從修過離散後，我們組員也都修了富國老師的 java 兩學期課程，期待在一成不變的數學裡，尋找著千變萬化的應用，而那時的我們，說不定早已定下了用程式輔助數學這個專題的目標，也正是如此，我們非常感謝富國老師的教導，因為前方有老師的引路，才有我們今天得以將傅立葉轉換，製作成深度學習的這項成果。

最後感謝國立嘉義大學對我們的教育之恩，給我們一個很好的學習環境，也謝謝許多在過程中幫助我們的好同學，鼓勵我們和分享了我們許多專題的方向，最後也很開心我們這一組彼此勉勵與鼓勵下，終於完成了這項數學與資訊融合的專題，謝謝大家，謝謝我們敬愛的鄭富國老師。

# 貳、研究內容

## 一、選擇及使用工具

### 1、Python

Python，最受歡迎的程式語言之一[[1]](#_肆、參考文獻)，是一種廣泛使用的直譯式、進階和通用的程式語言。其設計理念是「優雅」、「明確」、「簡單」，強調程式碼的可讀性和簡潔的語法。

Python還支援多種程式設計範式，包括函數式、指令式、反射式、結構化和物件導向程式設計，而且擁有動態型別系統和垃圾回收功能，能夠自動管理記憶體使用。Python還可以通過外界函式介面如標準庫中的ctypes等，來存取動態連結庫或共享庫中C相容資料類型並呼叫其中函式，以便於向用其他語言編寫的程式進行整合和封裝。又因為它本身設計為可擴充的，所以提供給用戶豐富的API和工具，方便用戶能夠輕鬆地使用Python、C、Cython來編寫擴充模組。

### 2、PyCharm

PyCharm是一個用於電腦編程的整合式開發環境（簡稱IDE）[[2]](#_肆、參考文獻)，也就是一個跨平台的程式開發環境，使用Python開發，其功能之一也是Python編輯器，可以為Python程式用戶提供眾多有用的功能。以下為PyCharm所擁有的其中幾項功能：

1. 代碼分析輔助：擁有補全代碼、突顯語法和錯誤提示。
2. 專案和代碼導航：專門的專案視圖，檔案結構視圖和和檔案、類、方法和用例的快速跳轉。
3. 重構：包括重新命名，提取方法，引入變數，引入常數、pull，push等。
4. 支援Web框架：如 Django, web2py 和 Flask。
5. 整合Python 除錯器。
6. 整合單元測試，按行覆蓋代碼。
7. Google App Engine下的Python開發。
8. 整合版本控制系統：為Mercurial， Git， Subversion, Perforce 和 CVS提供統一的使用者介面，擁有修改以及合併功能。
9. 遠端開發、支援資料庫與SQL。

### 3、Spyder

Spyder是開放原始碼跨平台科學運算集成開發環境(簡稱IDE) [[3]](#_肆、參考文獻)。其使用Python開發。Spyder整合了NumPy，SciPy，Matplotlib與IPython，以及其他開源軟件，比較容易做數據處理相關工作。

與其他科學數值分析專用IDE（如Matlab或RStudio）相比，Spyder有下列特色：第一，開放原始碼，以Python編寫且可以相容於非自由軟體授權。第二，可以使用附加元件擴充，內建互動式工具以處理數據。跨平台的特性使它可以通過Anaconda，Winpython和Python（x,y）。Spyder有以下功能 :

1. 編輯器：支持多語言，具有函數和類查看器，代碼分析特性（pyflakes和pylint獲得了支持），代碼補全，水平與垂直視窗的分離，直接跳入定義等等。
2. 交互端口：Python或IPython端口都在工作區可以調整和使用。支持對編輯器里的代碼直接調試。此外整合了Matplotlib的圖表顯示。
3. 文檔瀏覽器：在編輯器或端口中顯示任意類或函數調用的文檔。
4. 可變的瀏覽窗口：在文件的執行過程中可以創建可變的瀏覽窗口。同時也可以對其進行編輯。
5. 在文件中查找：支持正則表達式與Mercurial倉庫。
6. 其他擴展使用 : Spyder也可以作為PyQt4/PySide的擴展使用（spyderlib模塊）。例如，Spyder當中使用的Python交互端也可以被你用在自己的PyQt4/PySide程序中。

### 4、TensorFlow

TensorFlow是一個開源軟體庫[[4]](#_肆、參考文獻)，用於各種感知和語言理解任務的機器學習因此也有人稱呼它是一個機器學習框架。其用於研究和生產許多Google商業產品，如語音辨識、Gmail、Google 相簿和網頁搜尋。

TensorFlow充當機器學習的核心平台和函式庫。TensorFlow的API使用Keras來允許使用者製作自己的機器學習模型。除了建造和訓練它們的模型之外，TensorFlow還能幫助裝載資料來訓練模型，並使用TensorFlow Serving來部署它。

廣泛的應用程式使用TensorFlow作為基礎，其中它已成功實現自動化圖像字幕軟體，例如DeepDream。另外，2015年10月26日，Google正式啟用了由TensorFlow提供支援的RankBrain，處理了大量的搜尋查詢，替換補充傳統的靜態演算法搜尋結果。

### 5、CIFAR-100

CIFAR-100 是一個圖像數據集[[5]](#_肆、參考文獻)，包含 60000 張 32x32 分辨率的彩色圖像，其中分成10個大類，而其下又個別擁有10個小類，類別之間的交集為空，也就是不會有相同照片同時存在不同類別這種情況發生。總共100個類，每個類包含的600個圖像，各有500個訓練圖像和100個測試圖像。大類別包括：水生哺乳動物、魚類、大型食肉動物、大型人造戶外用品等等。小類別包括：airplane、automobile、bird、cat、deer、dog、frog、horse、ship、truck等等。

### 6、VGG-16

VGG 是英國牛津大學 Visual Geometry Group 的縮寫[[6]](#_肆、參考文獻)，主要貢獻是使用更多的隱藏層，大量的圖片訓練，提高準確率至90%。 VGG中根據卷積核大小和卷積層數目的不同，可分為A，A-LRN，B，C，D，E共6個配置(ConvNet Configuration)，其中以D、E兩種配置較為常用，分別稱為VGG16和VGG19。

###### 

###### 圖(2.1.1) ConvNet Configuration

針對VGG16進行具體分析，VGG16共包含：

1. 13個卷積層（Convolutional Layer），用conv3-XXX表示。
2. 3個全連接層（Fully connected Layer），用FC-XXXX表示。
3. 5個池化層（Pool layer），用maxpool表示。

其中，卷積層和全連接層具有權重係數，因此也被稱為權重層，總數目為13+3=16，這即是VGG16中16的來源（池化層不涉及權重，因此不屬於權重層，不被計數。）。

## 二、開發環境

### 1、模型訓練環境

* 品牌：MSI
* 處理器：Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz
* 已安裝記憶體(RAM) ：16.0 GB
* 系統類型：64位元作業系統，x64型處理器
* 顯示卡：GEFORCE GTX 1050 Ti 7680x4320@60Hz

### 2、主程式執行環境

* 品牌：Dell
* 處理器：Intel(R) Core(TM) i7-10510U CPU @ 2.30 GHz
* 已安裝記憶體(RAM) ：8.0 GB
* 系統類型：64位元作業系統，x64型處理器
* 顯示卡：Intel(R) UHD Graphics 1920x1080@59.99Hz

## 三、深度學習及卷積神經網路

### 1、深度學習

深度學習（deep learning）是機器學習的分支[[7]](#_肆、參考文獻)，是一種以人工神經網路為架構，對資料進行表徵學習的演算法。

深度學習是機器學習中一種基於對資料進行表徵學習的演算法。觀測值可以使用多種方式來表示，如每個像素強度值的向量，或者更抽象地表示成一系列邊、特定形狀的區域等。而使用某些特定的表示方法更容易從實例中學習任務（例如，臉部辨識或面部表情辨識）。深度學習的好處是用非監督式或半監督式的特徵學習和分層特徵提取高效演算法來替代手工取得特徵。表徵學習的目標是尋求更好的表示方法並建立更好的模型來從大規模未標記資料中學習這些表示方法。

至今已有數種深度學習框架，如深度神經網路、卷積神經網路和深度置信網路和迴圈神經網路已被應用在電腦視覺、語音辨識、自然語言處理、音訊辨識與生物資訊學等領域並取得了極好的效果。

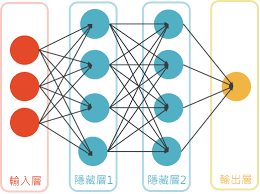
我們可以在無人駕駛汽車中看到深度學習技術，它能夠區分紅燈和綠燈，從路邊區分人，甚至可以測量兩輛車之間的距離。它是一種可以實現手機語音激活、人臉識別和個人設備手勢操作的技術，可以說深度學習已經跟我們生活息息相關。

以下為常見的深度學習庫：

* PyTorch
* ***TensorFlow***
* Theano
* PaddlePaddle
* Deeplearning4j
* Caffe
* roNNie
* ***Keras***
* MXNet
* 備註（***粗斜體***）為本專題主要使用的深度學習庫

### 2、卷積神經網路

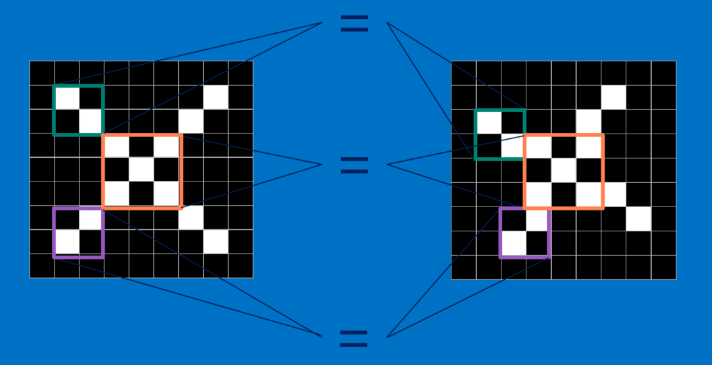
每當深度學習又有什麼重大突破時，這些進展幾乎都和卷積神經網路有關。卷積神經網路（Convolutional Neural Networks,CNN）是目前深度神經網路（deep neural network）領域的發展主力，在圖片辨別上甚至可以做到比人類還精準的程度。

卷積神經網路由一個或多個卷積層和頂端的全連通層組成，同時也包括關聯權重和池化層（pooling layer）。這一結構使得卷積神經網路能夠利用輸入資料的二維結構。與其他的深度學習結構相比，卷積神經網路在圖像和語音辨識方面能夠給出更好的結果。這一模型也可以使用反向傳播演算法進行訓練。相比較其他深度、前饋神經網路，卷積神經網路需要考量的參數更少，使之成為一種頗具吸引力的深度學習結構。

###### 圖(2.3.1)神經網路結構

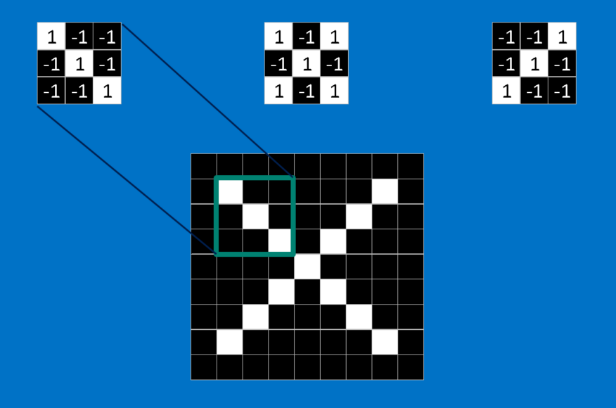
一般CNN會經過以下幾頁的重要的流程。

#### 特徵



###### 圖(2.3.2)比較特徵

CNN 會比較兩張圖片裡的各個局部，這些局部被稱為特徵（feature）。比起比較整張圖片，藉由在相似的位置上比對大略特徵，CNN 能更好地分辨兩張圖片是否相同。



###### 圖(2.3.3)捕捉特徵

一張圖片裡的每個特徵都像一張更小的圖片，也就是更小的二維矩陣。這些特徵會捕捉圖片中的共通要素。以叉叉的圖片為例，它最重要的特徵包括對角線和中間的交叉。也就是說，任何叉叉的線條或中心點應該都會符合這些特徵。

#### 卷積

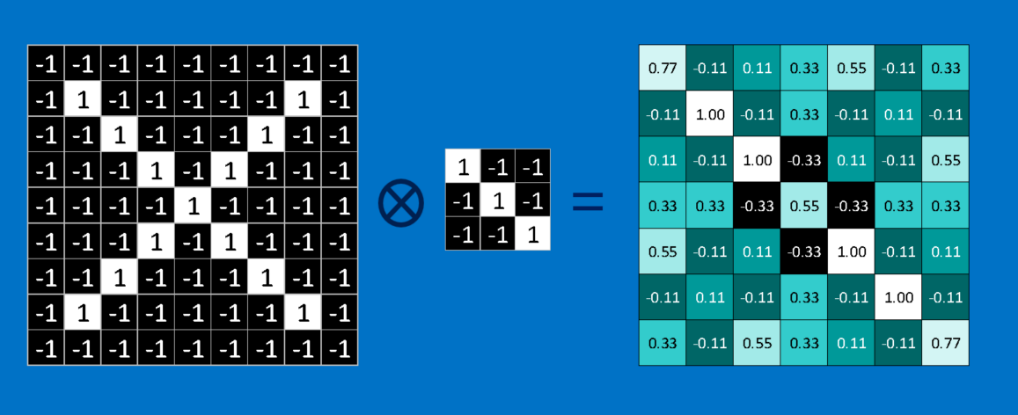
一張含有 文字, 計分板 的圖片

自動產生的描述

###### 圖(2.3.4)卷積運算1

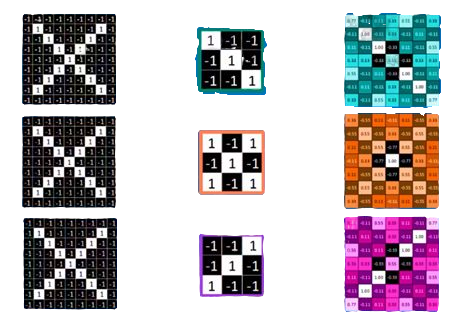
每當 CNN 分辨一張新圖片時，在不知道上述特徵在哪的情況下，CNN 會比對圖片中的任何地方。為了計算整張圖片裡有多少相符的特徵，我們在這裡創造了一套篩選機制。這套機制背後的數學原理被稱為卷積（convolution），也就是 CNN 的名稱由來。

要計算特徵和圖片局部的相符程度，只要將兩者各個像素上的值相乘、再將總和除以像素的數量。如果兩個像素都是白色（值為 1），乘積就是 1 \* 1 = 1；如果都是黑色（值為 -1），乘積就是 (-1) \* (-1) = 1。也就是說像素相符的乘積為 1，像素相異的乘積為 -1。如果兩張圖的每個相素都相符，將這些乘積加總、再除以像素數量就會得到 1；反之，如果兩者的像素完全相異，就會得到 -1。



###### 圖(2.3.5)卷積運算2

我們只要重複上述過程、歸納出圖片中各種可能的特徵，就能完成卷積。然後，我們可以根據每次卷積的值和位置，製作一個新的二維矩陣。這也就是利用特徵篩選過後的原圖，它可以告訴我們在原圖的哪些地方可以找到該特徵。值越接近 1 的局部和該特徵越相符，值越接近 -1 則相差越大，至於接近值接近 0 的局部，則幾乎沒有任何相似度可言。



###### 圖(2.3.6)卷積運算3

下一步是將同樣的方法應用在不同特徵上，在圖片中各個部位的卷積。最後我們會得到一組篩選過的原圖，每一張圖都對應了一個特徵。我們可以將整段卷積運算，簡單想成單一的處理步驟。在 CNN 的運作裡，這個步驟被稱為卷積層，這也代表後面還有更多層。

從 CNN 的運作原理，不難看出它很耗運算資源。雖然我們可以只用一張紙解釋完 CNN 如何運作，但運作途中相加、相乘和相除的數量可以增加得非常快。用數學來表達，可以說這些運算的數量，會隨著：

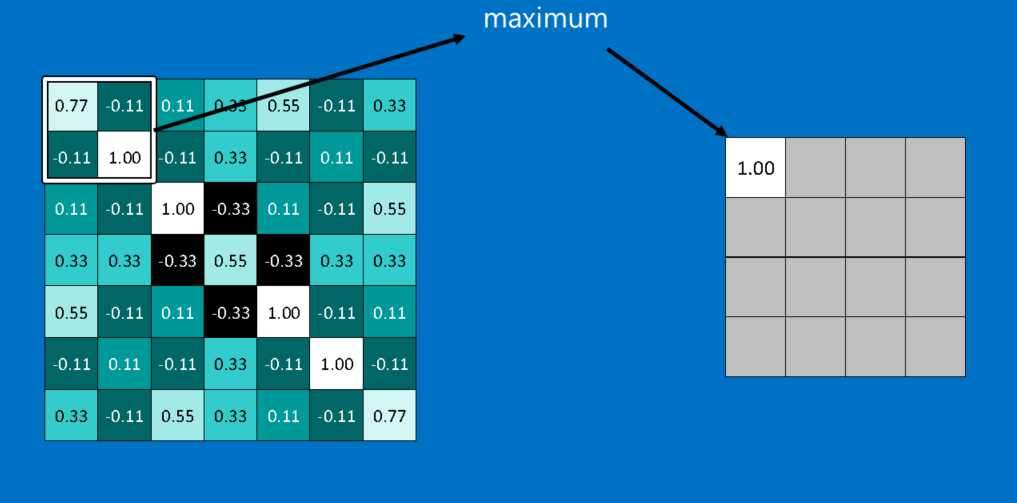
（一）圖片中相素的數量

（二）每個特徵中像素的數量

（三）特徵的數量呈現性增長

有了這麼多影響運算數量的因素，CNN 所處理的問題可以不費吹灰之力地變得非常複雜，也導致一些晶片製造商為了因應 CNN 的運算需求，正在設計和製造專用的晶片。

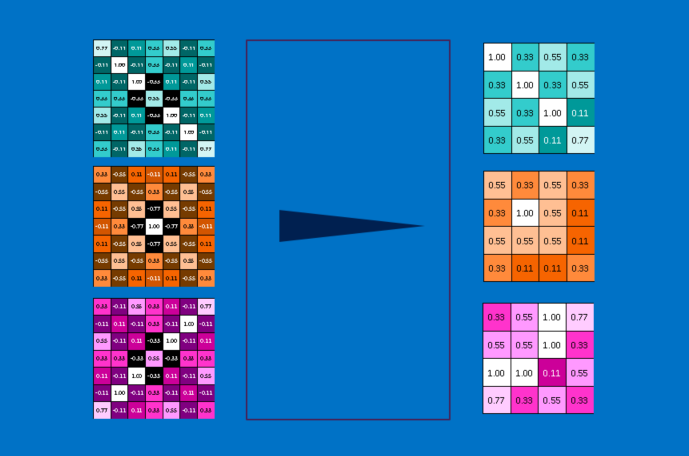
#### 池化



###### 圖(2.3.7)池化1

另一個 CNN 所使用的強大工具是池化（pooling）。池化是一個壓縮圖片並保留重要資訊的方法，它的運作原理只需要小二的數學程度就能弄懂。池化會在圖片上選取不同窗口，並在這個窗口範圍中選擇一個最大值。實務上，邊長為二或三的正方形範圍，搭配兩像素的間隔（stride）是滿理想的設定。

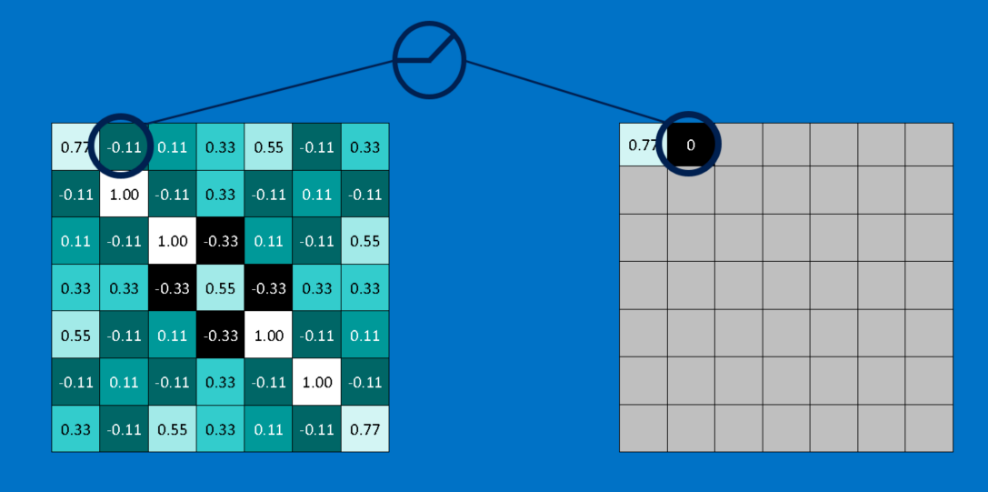
原圖經過池化以後，其所包含的像素數量會降為原本的四分之一，但因為池化後的圖片包含了原圖中各個範圍的最大值，它還是保留了每個範圍和各個特徵的相符程度。也就是說，池化後的資訊更專注於圖片中是否存在相符的特徵，而非圖片中哪裡存在這些特徵。這能幫助 CNN 判斷圖片中是否包含某項特徵，而不必分心於特徵的位置。這解決了前面提到電腦非常死板的問題。



###### 圖(2.3.8)池化2

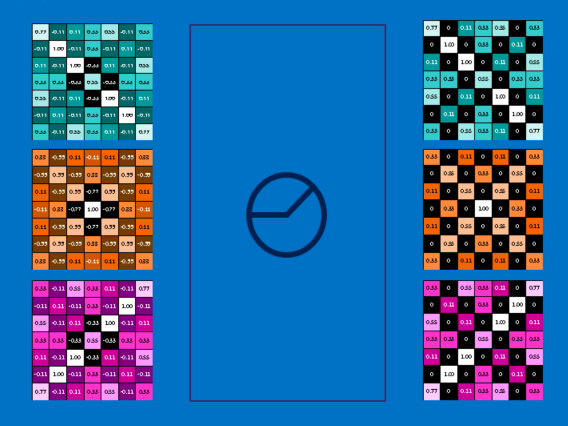
所以，池化層的功用是將一張或一些圖片池化成更小的圖片。最終我們會得到一樣數量、但包含更少像素的圖片。這也有助於改善剛才提到的耗費運算問題。事先將一張八百萬像素的圖片簡化成兩百萬像素，可以讓後續工作變得更輕鬆。

#### 線性整流單元



###### 圖(2.3.9) 線性整流單元1

另一個細微但重要的步驟是線性整流單元（Rectified Linear Unit，ReLU），它的數學原理也很簡單，將圖片上的所有負數轉為 0。這個技巧可以避免 CNN的運算結果趨近0或無限大。



###### 圖(2.3.10) 線性整流單元2

線性整流後的結果和原圖會有相同數量的像素，只不過所有的負值都會被換成零[[8][9]](#_肆、參考文獻)。

## 四、傅立葉轉換

### 1、一維傅立葉轉換

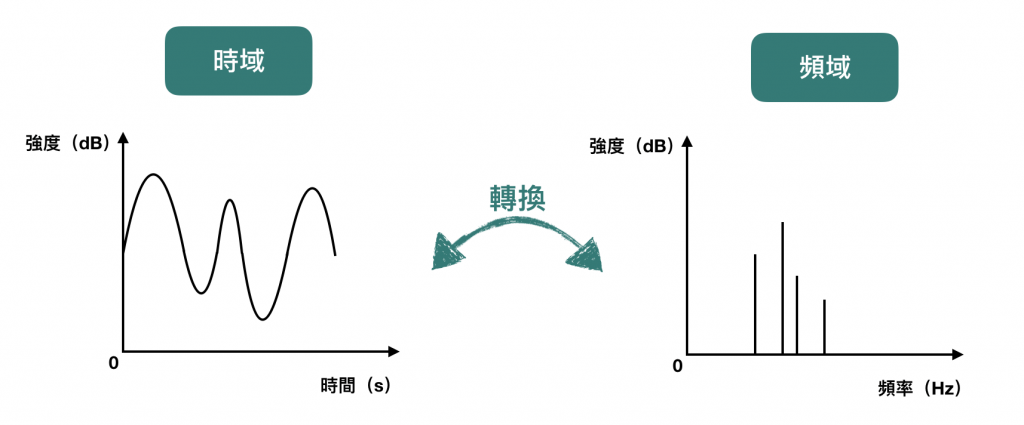
傅立葉轉換可將一個時域訊號，轉換到頻域上去。不管這個訊號是以時域的方式表現（波形, waveform），還是以頻域的方式表現出來（頻譜, spectrum），指的都是同一個訊號 ，只是藉由傅立葉轉換來改變我們對他的觀點[[10]](#_肆、參考文獻)。

* 時域

描述數學函數或物理信號對時間的關係。一個信號的時域波形可以表達信號隨著時間的變化。

* 頻域

是描述信號在頻率方面特性時用到的一種坐標系。頻域圖顯示了在一個頻率範圍內每個給定頻帶內的信號量。



###### 圖(2.4.1)時域頻域轉換圖[[11]](#_肆、參考文獻)

### 2、二維傅立葉轉換

相較於一維傅立葉轉換，對一張的靜態的平面影像而言，雖然沒有時間的概念在裡面，取而代之的是二維空間域的觀念。二維傅立葉變換的目的是將信號由空間域轉到頻域。

對於平面上的連續非週期信號 二維連續傅立葉變換定義為：

一張含有 文字 的圖片

自動產生的描述

影像處理的資料大多不是連續信號，而對於平面上的不連續信號，我們則需使用二維離散傅立葉變換。假設輸入的影像 水平方向長度是N，垂直方向長度是M，二維離散傅立葉變換定義為：

將以上的概念應用在灰階圖片上，空間域 表示在位置 的灰階值，二維傅立葉轉換後的結果，我們可以看到它是由哪些灰階度週期變化組成，此時，可以想像許多不同的灰階圖，各個灰階圖視覺上像個小漣漪，這些小漣漪會疊加成最後的灰階圖片，將此灰階圖進行二維傅立葉轉換，可以得到這些小漣漪的頻率，也就是所謂的圖像頻率。

我們得到圖像頻率後，可以將特定頻率的值設為0，逆二維傅立葉轉換後可實現濾波功能，例如，圖像的高頻訊號就視覺上來看，會讓界線較為明顯，如果可以去除低頻訊號，保留高頻訊號，就有機會保留較多的圖像邊緣。

* 逆二維連續傅立葉變換定義為：



* 逆二維離散傅立葉變換定義為：



[[12]](#_肆、參考文獻)

這就構成了圖像上的高通濾波器（High pass filter），只不過高頻部份不一定是邊緣，也有可能是高頻雜訊等其他來源，在沒有透過模糊等處理來抑制雜訊的情況下，檢視圖像濾波的結果時，若提高亮度，仍然可以看到一些非邊緣的紋路。

在圖像處理與分析上，傅立葉轉換是個非常有用的工具，除了能認識圖像頻率的意義之外，對於模糊、邊緣等各種處理方式，也會因此而融會貫通[[13]](#_肆、參考文獻)。

## 五、Python程式

### 1、CNN模型

我們運用Deep learning的其中一種深度學習框架，也就是上面說的CNN來當我們的深度學習模型。

首先我們會需要用來訓練模型的圖片，當然可以使用自己的圖片，甚至建立自己的Database，但是自己建立圖片庫會產生許多問題，如手機拍的圖片資料量過大、上網找的圖片有版權等；所以我們決定使用CIFAR-100，關於CIFAR-100先前已介紹過，我們直接進入程式步驟：

1. 引入函式庫

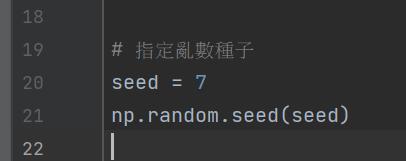
一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.1)

引入numpy、tensorflow、keras等函式庫。

1. 指定亂數種子



###### 圖(2.5.2)

深度學習網絡模型中初始的權值參數通常都是初始化成隨機數，而使用梯度下降法最終得到的局部最優解對於初始位置點的選擇很敏感，爲了能夠完全復現作者的開源深度學習代碼，隨機種子的選擇能夠減少一定程度上，算法結果的隨機性，也就是更接近於原始作者的結果，即產生隨機種子意味着每次運行實驗，產生的隨機數都是相同的。

1. 輸入資料集

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.3)

1. 正規化

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.4)

資料正規化是指將原始資料的數據按比例縮放於 [0, 1] 區間中，且不會改變原本的分佈，我們在進行影像辨識或圖像預處理時，也會習慣將像素值縮放到 [0,1] 之間（即除以255）。

1. 資料編碼

一張含有 文字 的圖片

自動產生的描述

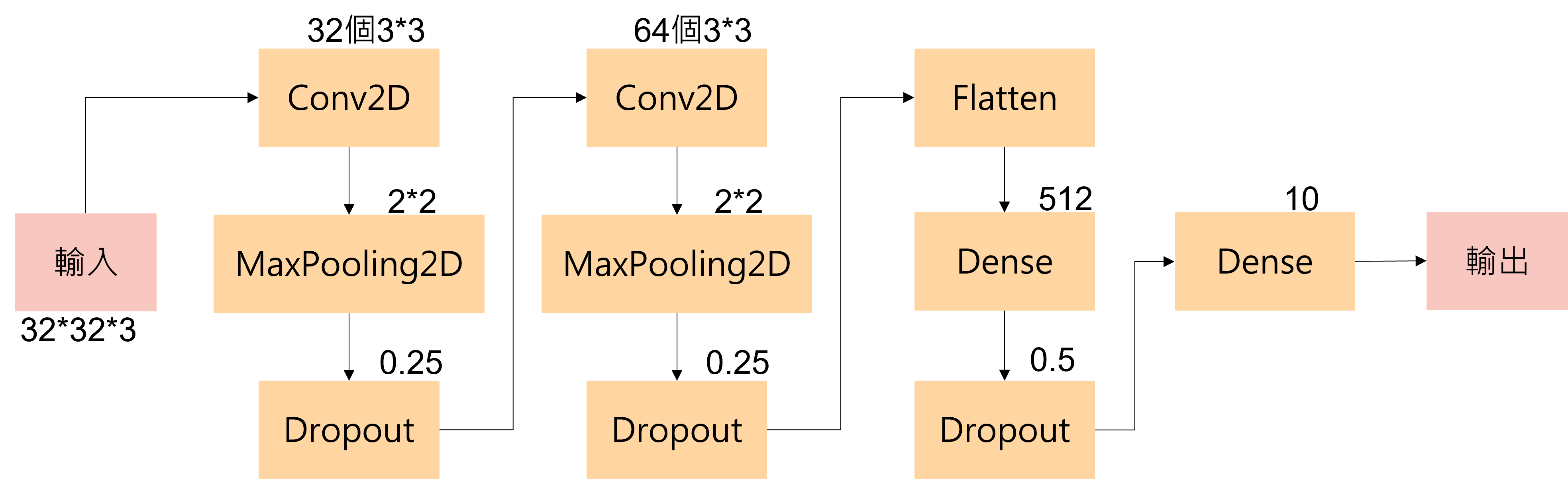
###### 圖(2.5.5)

要訓練模型時，data必須要是數值形式，所以要將類別型的特徵進行轉換，常用有兩種編碼方式：

* 標籤編碼（Label Encoding）：類似於流水號，依序將新出現的特徵類別編上新代碼，已出現的類別編上已使用的代碼；適合非深度學習時使用。
* 獨熱編碼（One Hot Encoding）：改良標籤編碼數字大小沒有意義的問題，將不同特徵類別分別獨立為一欄；適合非深度學習時使用。

1. 定義模型

接著定義神經網路模型，我們規劃的CNN如下圖所示（以兩層為例）：



**0**

###### 圖(2.5.6) [[14]](#_肆、參考文獻)

上圖有3組卷積和池化層、3個Dropout層、1個Flatten層、2個Dense層。因為CIFAR-100是預測100種分類，所以輸出層是100，程式如下：

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.7)

此2組卷積和池化層分別是32個和64個（3,3）過濾器，最後的Dense輸出層物件是100個神經元，啟動函數是softmax函數。然後呼叫summary()函式顯示模型的資訊。

1. 編譯模型

定義好模型後，我們需要編譯模型來轉換成TensorFlow計算圖，程式如下：

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.8)

此compile9()函式的損失函數是categorical\_crossentropy，優化器是adam，評估標準是accuracy準確度。

1. 訓練模型

在成功編譯模型後，就可以開始訓練模型，程式如下：

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.9)

此fit()函式的第1個參數是訓練資料集X\_train，第2個參數是訓練資料集Y\_train，並且分割出驗證資料集20%，期訓練週期是50次，批次尺寸是128。

1. 評估與儲存模型

使用訓練資料及訓練模型後，我們可以使用測試資料集來評估模型效能，程式如下：

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.10)

上述程式碼呼叫evaluate()函式來評估模型，然後使用save()函式儲存模型結構和權重。

### 2、主程式

訓練好模型後，我們需要寫一個主程式，用於將不同於data圖片的測試圖片，送入我們訓練好的模型進行視覺辨識，在終端輸出執行結果，並直接生成一個Excel表格，方便我們進行資料分析。

我們分為兩個程式：一是未將測試圖片經二維傅立葉轉換，直接送入模型辨識；二是將測試圖片經二維傅立葉轉換後，送入模型辨識。

#### 未將測試圖片經二維傅立葉轉換之程式

1. 引入函式庫

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.11)

引入pandas、tensorflow、numpy、os、time等函式。

1. 引入VGG16

因為前面用於訓練CIFAR-100的卷積層數不夠，所以我們引入VGG16，利用VGG16中間層萃取的特徵重新訓練資料。

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.12)

一張含有 文字 的圖片

自動產生的描述 VGG16的卷積和池化層程式如下：

###### 圖(2.5.13)一張含有 文字 的圖片 自動產生的描述

1. 定義陣列

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.14)

array\_of\_image、array\_of\_image\_name為儲存測試圖片集的一維陣列；show\_svr為儲存結果的二維陣列，大小是3\*100。

1. 一張含有 文字 的圖片

   自動產生的描述處理測試圖片

###### 圖(2.5.15)

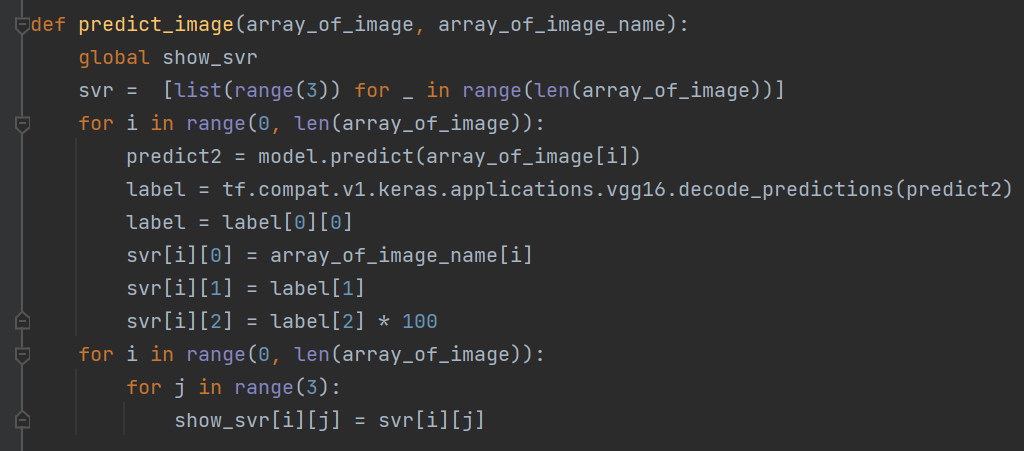
load\_img函式為讀取測試圖片，其中target\_size函數是轉換圖片大小為224\*224。

img\_to\_array函式為將圖片轉成numpy形式。

reshape函式為轉換圖片維度。

preprocess\_input函式為將圖片送入模型。

append函式為將圖片資料存入陣列。

1. 預測圖片

###### 圖(2.5.16)

第3列為宣告一個local二維陣列svr儲存結果。

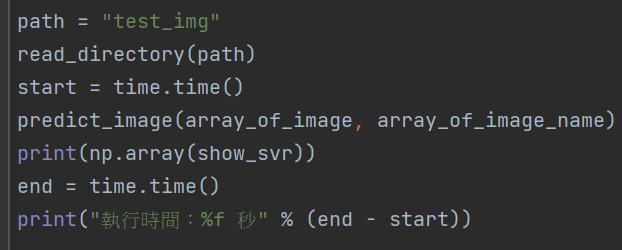
model.predict函式為預測圖片。

decode\_predictions函式為將預測資料轉文字。

第7～10列為將結果儲存至svr中。

第11～13列為將svr中的內容複製到show\_svr中。

1. 輸出影像



###### 圖(2.5.17)

第一列為將path儲存為測試圖片集之路徑。

呼叫read\_directory函式並送入path。

start = time.time()為儲存預測開始之時間。

呼叫predict\_image函式。

print(np.array(show\_svr))為輸出執行結果。

end = time.time()為儲存預測結束之時間。

print("執行時間：%f 秒" % (end - start))為輸出預測時間。

1. 生成Excel

一張含有 文字 的圖片

自動產生的描述

###### 圖(2.5.18)

前兩列為讀取執行結果之資料。

第3列為將資料存入Excel中。

第4列為把資料放在叫project\_nonftt2的頁面中，預設格子寬度為5。

第5、6列為儲存Excel並關閉。

#### 將測試圖片經二維傅立葉轉換過之程式

經二維傅立葉轉換過之程式差別地方在於：

1. 一張含有 文字 的圖片

   自動產生的描述處理測試圖片

###### 圖(2.5.19)一張含有 文字 的圖片 自動產生的描述

第5、6列為二維傅立葉轉換。

第8～10列為設定高通濾波，保留高頻。

第12～14列為逆二維傅立葉轉換。

1. 一張含有 文字 的圖片

   自動產生的描述生成Excel

###### 圖(2.5.20)

第4列為把資料放在project\_ftt2的頁面中，預設寬度為5。

# 參、結論

## 一、執行結果

### 1、圖片經二維傅立葉轉換之輸出結果



###### 表(3.1.1)

### 2、圖片不經二維傅立葉轉換之輸出結果



###### 表(3.1.2)

## 二、結果分析

### 1、程式執行辨識的執行時間

1. 經二維傅立葉轉換：輸出時間為：33.600873秒

一張含有 文字, 螢幕擷取畫面, 監視器, 黑色 的圖片

自動產生的描述

###### 圖(3.2.1)

1. 不經二維傅立葉轉換：輸出時間為：31.445531秒

一張含有 文字, 螢幕擷取畫面, 監視器, 黑色 的圖片

自動產生的描述

###### 圖(3.2.2)

綜合兩者的輸出時間，我們可以得出經二維傅立葉轉換之程式的執行時間高於不經二維傅立葉轉換之程式的執行時間，但是兩者差距僅2秒，約1.068542倍，所以已結果來說兩者差距可忽略。

### 2、準確率對比圖

我們將兩個程式輸出的Excel做比較，並製作成表格，增加可視度外也方便直接比較：

###### 圖(3.2.3)

###### 圖(3.2.4)

###### 圖(3.2.5)

###### 圖(3.2.6)

###### 圖(3.2.7)

###### 圖(3.2.8)

###### 圖(3.2.9)

一張含有 文字 的圖片

自動產生的描述

###### 圖(3.2.10)

統整之後我們比較準確率發現經二維傅立葉轉換之程式準確率比不經二維傅立葉轉換之程式準確率還高的情況有63次，比例差為1.7027027倍，經二維傅立葉轉換之程式準確率明顯高於不經二維傅立葉轉換之程式準確率。

綜合上述之比較，時間相差約1.068542倍，準確率比較高的次數相差約1.7027027倍；以準確率提升的比率來看，我們認為：將測試圖片經二維傅立葉轉換後再送入訓練好的模型進行視覺辨識預測，是比沒有經二維傅立葉轉換的方式好的，透過二維傅立葉轉換的特性來找到圖片頻率，並通過高通濾波保留高頻進行預測來降低資訊量，進而將背景複雜的圖片複雜度降低，大幅提升準確率。

這個情況跟我們執行程式前猜想的結果大致相同，只有執行時間是我們意想之外的結果，執行前的猜想是經二維傅立葉轉換之程式的執行時間會低於沒有經二維傅立葉轉換之程式的執行時間，我們認為造成此結果的原因可能有：

1. 程式不夠精簡
2. 硬體設備不夠完備，包括：沒有獨顯、RAM不足等問題
3. 演算法不完備

我們會以此為目標繼續完備我們的模型，持續精進程式技巧，完成我們的目標。

## 三、未來展望

深度學習是未來的趨勢，也是世界趨之若鶩研究的目標，從我們身邊的手機、各種軟體到人類基本一定會碰到的醫學、甚至是還在開發期的自動駕駛都是深度學習的範疇；在這世界以及市場的需求下，人工智慧的知識及研究、與人工智慧的人才，都是不可多得的存在。

我們的專題在於研究圖片辨識更好的演算法，在提升準確率的同時融入在數學系學習的專長，完成數學與資訊融合的成果，以此為基礎往更深的地方研究，只能說延伸是非常非常多的；我們預計尋找更多關於人臉的圖片資料，並運用攝影機時時寫入圖像資料，實現更實用的人臉辨識，加入紅外線儀器可以監測來上課的同學的體溫、加入 Live2D 虛擬人偶可以實現透過肢體及臉部表情控制人偶動作、甚至是導入個人資料及臉部詳細資訊可以實現2D 人臉辨識等延伸，我們希望可以在畢業前完成：能夠時時追蹤攝影機捕捉到的人的臉的模型以及演算法，我們將以這個目標持續努力。

而對於未來的期望，我們研究的專題正是步入人工智慧領域的一個很好的基礎，雖然我們的訓練資料只有上萬筆，測試資料只有幾百筆的情況下，還無法稱是大數據分析，但是當我們的知識及能力上升後，處理的資料就從一萬筆、十萬筆、甚至上千萬筆起跳，這時更能凸顯我們數學系專長應用之一的大數據分析的能力，運用我們比其他人還要多的數學知識來強化我們的能力，期望我們在未來的工作中，都能透過記起這個專題的融合性以及實用性，創造出更能造福社會的作品。

# 肆、參考文獻

[1] <https://zh.wikipedia.org/zh-tw/Python>

[2] <https://zh.wikipedia.org/zh-tw/PyCharm>

[3] <https://zh.wikipedia.org/zh-tw/Spyder>

[4] <https://zh.wikipedia.org/zh-tw/TensorFlow>

[5] <https://www.cs.toronto.edu/~kriz/cifar.html>

[6] <http://deanhan.com/2018/07/26/vgg16/>

[7] <https://zh.wikipedia.org/zh-tw/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0>

[8] <https://brohrer.mcknote.com/zh-Hant/how_machine_learning_works/how_convolutional_neural_networks_work.html>

[9] [How Convolutional Neural Networks work - YouTube](https://www.youtube.com/watch?v=FmpDIaiMIeA&t=757s)

[10] <https://hackmd.io/@jkrvivian/B1wHF21ib?type=view>

[11] <https://ithelp.ithome.com.tw/articles/10228205>

[12] <https://zh.m.wikipedia.org/zh-tw/%E4%BA%8C%E7%B6%AD%E5%82%85%E7%AB%8B%E8%91%89%E8%AE%8A%E6%8F%9B>

[13] <https://www.ithome.com.tw/voice/145429>

[14] [TensorFlow 與 Keras：Python 深度學習應用實務](https://www.books.com.tw/products/0010831367)

# 伍、附錄

## 專題網站

<https://lhopita1.github.io/fanyun/graduation_project/>

[](https://lhopita1.github.io/fanyun/graduation_project/)