**國立臺灣師範大學附屬高級中學第46屆科學展覽會**

**作品說明書封面**

科　　別：

組　　別：

作品名稱：大腸鏡影像檢測有無息肉

關 鍵 詞：人工智慧、VGG-16、息肉（最多3個）

編 號：

摘要

由於尚未得到教授欲提供的腸道腫瘤照片資料，因此先暫時以cifar10進行訓練並建立其VGG-16的模型，又因其中的前幾層都相同，因此後續待拿到資料後即可以進行轉移學習。

1. 研究動機

由於醫生經常要耗費大量時間及精力來人工辨識病人腸道胃鏡之類的照片，辨識是否有息肉的產生，為了盡我們所能減少醫生的辛勞，因此我們希望利用CNN來進行機器學習並直接辨識該大腸鏡照片中是否有息肉，以減少人工辨識所需浪費的人力及時間。谷歌公司曾利用C2D2深度學習演算法為大腸3D建模提高檢查的覆蓋面積。配合我們研究的大腸鏡息肉檢測模型正是如虎添翼。

貳、研究目的

藉由深度學習來協助進行大腸鏡影像有無息肉的辨識。

參、研究設備及器材

1. 、訓練資料

1. cifar10  
 以預先訓練模型中的前幾層，利用轉移學習提高訓練效率。  
 2. 大腸鏡息肉影像  
 以訓練模型主體。

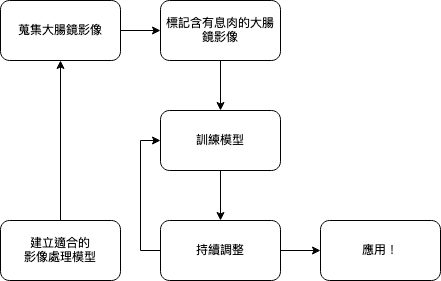
（二）、筆電  
 1. 2015 Mac Book PRO  
 2. 2020 Mac Book PRO

本研究使用的研究器材。

（三）、開發環境  
 1. Google Colab免費方案

本研究使用的開發環境，免費又好用，可以線上共用。其提供12Gb的隨機存取記憶體，與100Gb的硬碟儲存訓練資料。但是每次使用時必須自行引入所有檔案，所有程式碼區塊共用命名空間。

肆、研究過程或方法

以上為我們研究的流程圖。

1. 建立適合的影像處理模型  
    在目前這個階段我們尚未取得大腸鏡息肉照片供我們訓練。但是我們可以利用我們將使用的VGG-16模型的特性「轉移學習」先行訓練模型的前幾層。  
    我們使用keras建立並且訓練模型，我們使用的模型是以VGG-16為基礎，對我們的需求更進一步修改的模型。前面有提到因為要使用cifar10的資料做轉移學習，故取得轉移學習資料時的模型共有10個輸出的神經元。此模型圖可以在附錄（一）找到。
2. 、VGG-16  
    至於使用VGG-16的原因，不外乎是因為他擁有優異的影像分類能力，在原先的設計中能夠分類1000種不同的物品。所以可以在他的模型最後面看到最後一個1000個神經元的全連接層以softmax作為激活函數。VGG-16之所以後面或有一個16是因為VGG-16有13個卷積層和3個全連接層。可以參考附錄（二）的VGG實驗圖，當中的D和E就是VGG-16和VGG-19。之所以選擇VGG-16而不選擇VGG-19是因為我們認為我們欲訓練的模型沒有大到需要19層，畢竟他原先是設計分類1000種物品，而我們只是要辨識有無息肉。此模型甚至利用了一個名為”Multiple Scale training”的技術會在訓練時隨機切取原始圖片[論文一]，預測不同的裁切大小並且在各個位置各取一張，最後取平均成為預測結果。雖然此論文的結論是越深的模型越好，但是論文中也有提到，越深的模型就會需要越多的資料訓練，否則模型多再多層結果也相差不大。  
    論文中提到他在訓練時使用了transfer learning，將第一個訓練的模型作為之後模型前四層的初始權重。因為前幾層的卷基層辨識的都是點線之類的幾何特徵，所以可以轉移。
3. 、轉移學習  
    因為大腸鏡息肉的樣本有限，故我們將會使用cifar10的資料訓練本模型的前幾層以進行轉移學習，提高準確度。
4. 蒐集大腸鏡影像  
    本研究與大學端合作，即將取得一些大腸鏡影像。
5. 標記含有息肉的大腸鏡影像  
    本研究與大學端合作，即將取得一些標記好的大腸鏡影像。
6. 訓練模型  
    （一）、轉移學習

本研究在訓練模型時遭遇了相當多的困難，特別是在第一次訓練時。我們為了要取得可以拿來轉移學習的模型，必須要先用cifar10的資料進行訓練。但是在訓練資料提取的程式碼和模型輸入的程式碼過程中遇到了不少的問題，最後終於成功了。

當中我們使用categorical\_crossentropy作為我們模型的損失函數，該損失函數適合分類的問題，而其特姓為驗證答案的輸入端須為所有可能數的維度的向量。

首先讀取資料後，必須將其轉為三層的向量，並且將所有像素的RGB轉為0到1的浮點數，然後和所有其他的圖片經過處理後的向量串在一起成為訓練集的列表。同時處理圖片的標記，因為cifar10共有十類圖片，將所有圖片的標記轉為一個十維的向量，如下圖，表示第一類圖片的標記。

  
 詳細程式碼請看附錄（三）。  
 而我們替cifar10特製的模型可以在附錄（六）找到。當中將原本VGG-16的最後一層1000個神經元的全連接層換成10神經元的以切合需求。

1. 持續調整  
    在訓練的過程中我們遭遇很多困難與挑戰，以下將描述我們遇到的問題與可能的解決方法。  
    （一）、模型參數的調整  
    一開始我們使用harse\_categorical\_crossentropy作為學習使用的損失函數，但是後來發現此損失函數並非最理想的。因為偏差值一直很大。所以將其調整為categorical\_crossentropy。  
    （二）、訓練資料的品質  
    我原先以為就算不修改訓練時的順序，應該對模型訓練沒有關係，因為此模型並沒有長期記憶或是短期記憶的問題。但是在開始訓練後我觀察到正確率極大幅度的上升，原來是模型以為答案只有這一種，所以那個答案的機率特別高。但是就在我把順序打亂後，模型的正確率急遽下降，逐漸逼近0.1，相當於亂猜，我們初步認為是訓練用的照片太不清楚，其實人也看不太出來那些是什麼。  
     
     
    例如說這張，看起來有點像吹風機，也有點像藍色小手槍，但是這張圖片其實是  
   飛機。像是這樣的圖片品質我們很難訓練出高品質的模型。同樣的例子還有以下這個。  
     
   乍看之下像綿羊，細看像老鼠，誰也沒想到這是一隻鳥，我們需要增進我們訓練圖片的品質。  
    （三）、訓練參數的調整  
    1. 輸入圖像大小  
    我們最初將模型輸入的圖片大小設定為32px×32px，因為我們訓練轉移學習的部分使用的訓練資料來源是cifar10，其圖片原始大小為32px×32px，我們原先以為模型輸入只要設為原始圖片大小即可，但是經過實驗發現，圖像大小設置到接近224px×224px，可讓模型訓練的效率提升。我們參考VGG原作者的實驗將模型輸入端口的圖像大小長寬都分別設置為224px，我們認為這應該是因為VGG系列的模型非常深，且擁有非常多的kernal，雖然每個kernal大小都很小，但是訓練起來非常費力，若是使用更大的圖片，就能夠在維持kernal大小的前提下提升對細節的精確度，此外，也能更有效率的進行訓練。  
    2. 訓練批次的調整  
    在一開始時，我們直接將五萬張的資料餵給我們可愛的模型。但是她顯然無法一次接受這麼大的資訊量，因為當我們把一次的訓練量降到500張時，模型開始有了進步，但是她的準確度依舊時好時壞，最高不超過25%。

最後我們決定一次訓練100張，同一批訓練50次（恰可將模型對這100張辨識的結果達到100%）取250批，終於有了顯著的成果。

1. 應用

本研究最主要的用途即為協助醫生進行辨識，由於訓練後得到的準確度並非100％，因此在辨識上仍然有極小的機率進行誤判，為了確保不會因誤判而造成延誤治療，因此還是同時還需要醫生的判斷。不過我們認為應該在醫生檢查過後，再進行辨識，以確保判斷得到的結果與醫生相同，而不要讓醫生因為已經有了電腦判斷的結果而產生主觀意見，使他無意中忽略了電腦沒判斷到的部分。我們也希望這個技術可以因此造福大眾，減少更多因誤判導致的延誤治療。不過若真的要從根本解決息肉的問題，注意飲食及避免煙、酒才是至關重要的解決方法。

伍、研究結果  
 一、發現圖片大小與模型訓練效率的關聯性

在我們的實驗中可以發現圖像較大時，可以捕捉到更多的圖片細節。

二、

陸、討論

ㄧ、文獻探討

在我們的研究中，我們使用的VGG是Oxford的Visual Geometry Group組提出的。此模型證明了在某種程度上越深的模型的效果越好。在人工智慧進行大腸保健這方面上，已經有谷歌公司為病人的大腸3D建模提高醫生人工檢查的覆蓋率，目前無人提出開源的人工智慧模型檢測有無息肉。

柒、結論

捌、參考資料及其他

一、參考資料

Karen Simonyan & Andrew Zisserman. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. International Conference on Learning Representations

Y. Lecun, L. Bottou, Y. Bengio and P. Haffner. (1998). *Gradient-based learning applied to document recognition*, in Proceedings of the IEEE

, vol. 86, no. 11, pp. 2278-2324, doi: 10.1109/5.726791

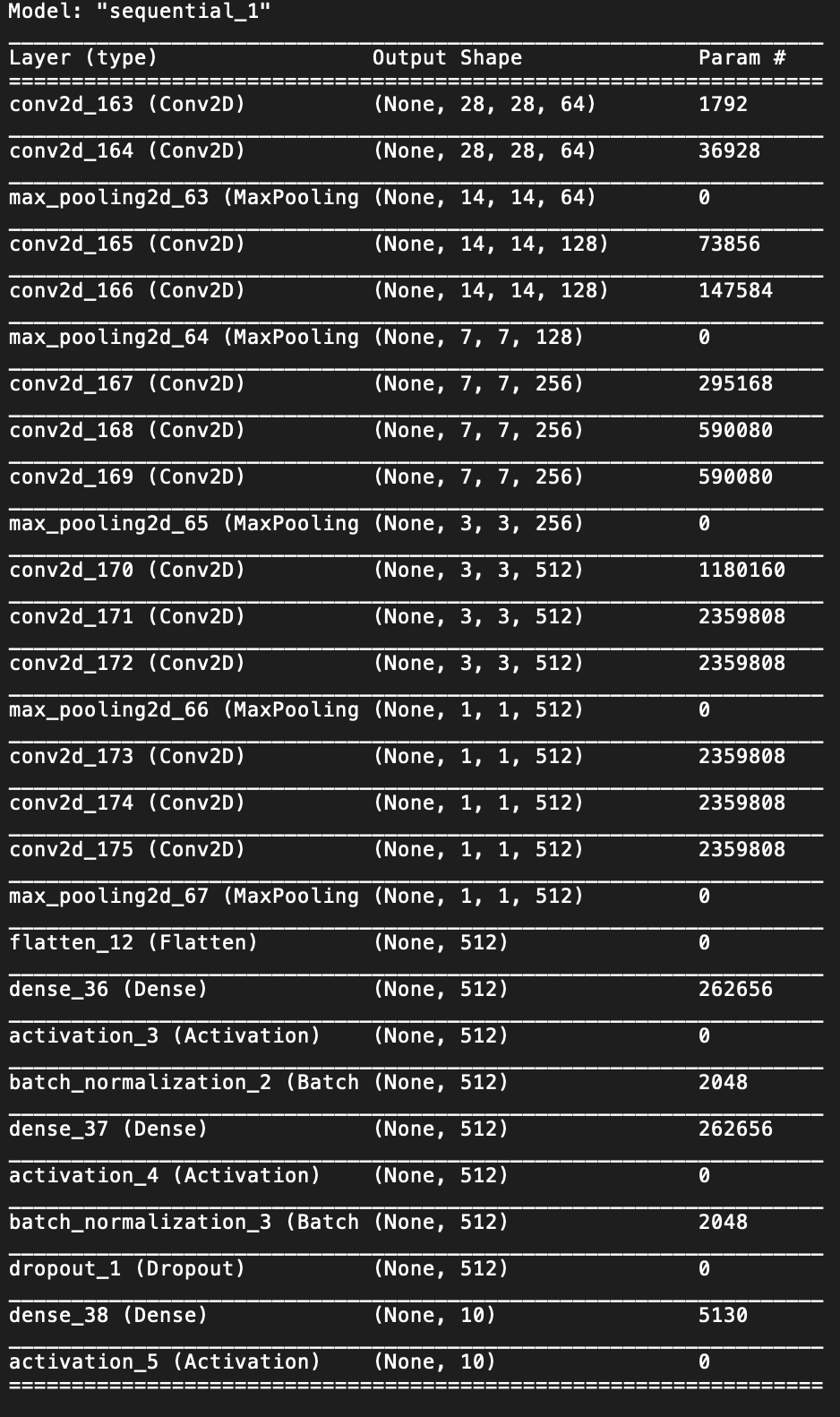
荷葉田田 ( 2019 年 03 月 12 日 )。VGG16和VGG19介紹【部落格影音資料】。取自

*https://blog.csdn.net/qian2213762498/article/details/88422941*

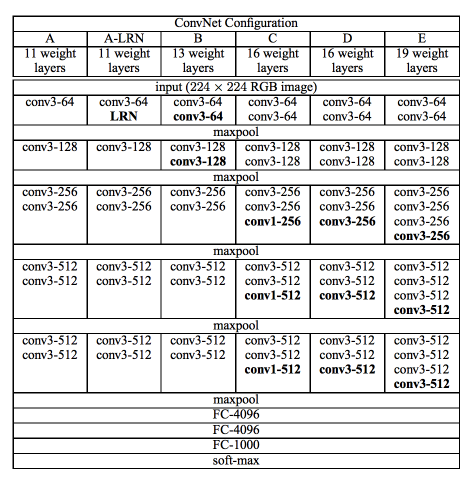
I code so I am ( 2017 年 12 月 19 日 )。CNN 經典模型應用【部落格影音資料】。取自

*https://ithelp.ithome.com.tw/articles/10192162*

1. 附錄  
    特別需要提及的是由於我們開發環境的關係（Google Colab），所有程式共用命名空間，其特性更方便開發者進行人工智慧相關的開發。故下列程式碼中各函數與變數名稱是共用的。

（一）、模型結構1  
 

（二）、VGG實驗圖

  
（三）、產生訓練資料之程式碼  
# 產生訓練資料用

import os

data\_path = '/content/sample\_data/cifar10/train/'

Xtrain = []

Ytrain = []

classes = sorted(os.listdir(data\_path))

print(f'classes: {classes}')

class\_num = 0

for img\_class in classes:

for img in os.listdir(f'{data\_path}{img\_class}/'):

image = load\_img(f'{data\_path}{img\_class}/{img}',target\_size=(32,32))

image = np.array(image)

image = image.astype('float32')

image /= 255

Xtrain.append(image)

Ytrain.append([0]\*class\_num + [1] + [0]\*(9-class\_num))

class\_num += 1

Xtrain = np.array(Xtrain)

Ytrain = np.array(Ytrain)

（四）、預測圖片之程式碼

# 此格為測試模型用

test\_img\_path = '/content/sample\_data/10020.jpg'

image = load\_img(test\_img\_path,target\_size=(32,32))

image = np.expand\_dims(image, axis=0)

# image = np.array(image)

image = image.astype('float32')

image /= 255

features = model.predict(image)

print(features)

#print('Predicted:', decode\_predictions(features, top=3)[0])

1. 、訓練模型之程式碼  
   EPOCH = 3

history = model.fit(

Xtrain,

Ytrain,

epochs=EPOCH

)

（六）、建立模型之程式碼（包含編譯，設定損失函數等等）  
# 此格為建立模型用，再度執行會覆寫model

from keras.models import Sequential

from keras.layers import Dense, Activation, Dropout, Flatten

from keras.layers import Conv2D

from keras.layers import MaxPooling2D

import numpy as np

from keras.preprocessing.image import ImageDataGenerator, load\_img, img\_to\_array

image = load\_img('/content/sample\_data/10020.jpg',target\_size=(32,32))

# image = np.expand\_dims(image, axis=0)

image = np.array(image)

image = image.astype('float32')

image /= 255

# print(image)

input\_shape = (32, 32, 3)

model = Sequential([

Conv2D(64, (3, 3), input\_shape=input\_shape, padding='same',

activation='relu'),

Conv2D(64, (3, 3), activation='relu', padding='same'),

MaxPooling2D(pool\_size=(2, 2), strides=(2, 2)),

Conv2D(128, (3, 3), activation='relu', padding='same'),

Conv2D(128, (3, 3), activation='relu', padding='same',),

MaxPooling2D(pool\_size=(2, 2), strides=(2, 2)),

Conv2D(256, (3, 3), activation='relu', padding='same',),

Conv2D(256, (3, 3), activation='relu', padding='same',),

Conv2D(256, (3, 3), activation='relu', padding='same',),

MaxPooling2D(pool\_size=(2, 2), strides=(2, 2)),

Conv2D(512, (3, 3), activation='relu', padding='same',),

Conv2D(512, (3, 3), activation='relu', padding='same',),

Conv2D(512, (3, 3), activation='relu', padding='same',),

MaxPooling2D(pool\_size=(2, 2), strides=(2, 2)),

Conv2D(512, (3, 3), activation='relu', padding='same',),

Conv2D(512, (3, 3), activation='relu', padding='same',),

Conv2D(512, (3, 3), activation='relu', padding='same',),

MaxPooling2D(pool\_size=(2, 2), strides=(2, 2)),

Flatten(),

Dense(4096, activation='relu'),

Dense(4096, activation='relu'),

Dense(10, activation='softmax')

])

model.summary()

model.compile(

optimizer='adam',

loss='categorical\_crossentropy',

metrics=['accuracy']

)