(演算法與資料分析)期中報告

利用 ELA 搭配神經網路進行影像偽造識別

國立金門大學

資訊工程系

組長　魏仲彥

組員　徐伯元

指導教授: 張伯銀教授

目錄

(一)前言

圖片裡面包含者許多資訊，許多假圖片可以利用光線等因素，使用肉眼判斷出來，但近年來電腦科學領域快速的成長，包括人工智慧、大數據、機器學習、深度學習等，逐步與一般生活結合 。然而，在過程中不免會出現有心人士利用這些技術去做一些侵犯他人權益的事情，網路上的圖片或是影片不一定眼見為憑，許多圖片也會因此而無法識別。假的圖片、影片席捲全球，所以識別圖片的技術也就從而誕生出來，除了使用肉眼判斷圖片，檢查圖片真偽有很多辦法，像從照片格式、元資料、RGB…，本專題將會使用本專題會使用 Convolutional neural network(CNN; 運用卷積神經網路)的 Conv2D 建立模型，並對影像進行 Error Level Analysis(ELA;錯誤級別分析)，從而判斷影像是否進行過人工修改。 並使用 CNN 運算來判斷圖片的可信程度。

(二)文獻探討

deepfake 對生活的影響

深偽技術，又稱為深度偽造，是深度學習和偽造的混合詞，指 AI 的人體圖像合成技術應用。近年來因為技術進步，普通人也可以使用 Deepfake 進行圖片偽造，遭成了許多隱患。deepfake 常被用在偽造媒體、影片，甚至聲音也可以展現出來。在教育上，可以呈現出古人的樣貌，像是數位愛因斯坦，在警界中 deepfake 可以協助調查犯罪，還原模糊或是損毀的影像。但 deepfake 所帶來的負面問題遠比它的正面多，隱私問題、造謠抹黑、放假消息，許多科技隱患也就此慢慢誕生。

傳統上是採用基於圖學的 3D 模型重建追蹤技術。較新的研究則是採用深度學習來達到換臉效果，為了解決深度學習的訓練難度和生成品質，又進一步融合了生成對抗網路技術。表情偽造是將其他人臉圖像的表情替換到目標人臉上，從而達到目標人物做指定表情的目的，換臉偽造和表情偽造還常常結合語音偽造技術。通過文字到語音合成和語音轉換技術來製作虛假語音。對於日益增長的 deepfake 技術，可能會發生下面的隱患: 假造企業主發言，詐騙頻傳或干擾營運; 色情片移花接木，貶損人格聲譽; 網路銀行應用影響伸至個人資料。

影像辨識技術

RGB 識別技術

紅綠藍是光的三原色,每個圖片上都是由紅綠藍三個像素點所製成的,但是,不是每一個格子都存在著 RGB 這三種信息,一般來說,需要經由個個像素點之間的演算法來控制,所以每個格子其實有一定的數字關係,當格子與附近的數字沒有關係或是異常時,就可以判定圖片有被修改過。

噪點識別技術

使用相機拍照時,照片上一定會有密密麻麻地小顆粒,尤其是夜晚拍照時更為明顯,這些顆粒就是所謂的噪點,噪點是由感光元件接收粗糙的部分所形成的,噪點存在著特定的分布,如果圖片上某一個部分是複製過來或是修改過就會很明顯,這種異常人眼是很難看出來的,但對機器來說卻十分容易看出有沒有修改過。

JPG 資料壓縮

JPG 操作模式

在資料壓縮（data compression）方面，JPG 有如下的二種操作模式:

一、非失真壓縮（lossless compression）：圖形檔案經過壓縮後，能完全還原成原來的檔案，不會損失任何資訊。但是此種模式的資料壓縮效果不佳，一般在 JPG 的應用中不會採取此種方式。

二、失真壓縮（lossy compression）：圖形檔案經過壓縮後，無法完全還原成原來的檔案,因此會有一些資訊的損失。不過由於人類眼睛無法察覺圖形(或相片）中色彩、亮度、與解析度等方面非常細微的差異，因此在色彩、亮度、與解析度上的某些資訊損失對人眼而言，是沒有影響或可被容忍的。一般而言，10 倍的壓縮比（即壓縮成原本檔案大小的 1/10），對圖形（或相片）的觀賞效果不會有任何的影響，因此這是最常被採用的操作模式。

因為壓縮比與圖形（或相片）的觀賞效果是反比的關係，即壓縮比越高，圖形的觀賞效果越差，不過實際情況經常因不同的圖形而有很大的差異，因此 JPG 的處理軟體也有提供一個參數，可以讓使用者自行來調整壓縮比，以便在檔案大小和圖形觀賞效果間取得較佳的平衡點。

JPG 圖像有損壓縮流程：

一、顏色轉換:由於 JPEG 只支持 YUV 顏色模式的數據結構，而不支持 RGB 圖像數據結構，所以在將彩色圖像進行壓縮之前，必須先對顏色模式進行數據轉換。

二、DCT 變換:DCT（Discrete ConsineTransform）是將圖像信號在頻率域上進行變換，分離出高頻和低頻信息的處理過程。然後再對圖像的高頻部分（即圖像細節）進行壓縮，以達到壓縮圖像數據的目的。

三、量化:由於在後面編碼過程中使用的碼本都是整數，因此需要對變換後的頻率係數進行量化，將之轉換爲整數。由於進行數據量化後，矩陣中的數據都是近似值，和原始圖像數據之間有了差異，這一差異是造成圖像壓縮後失真的主要原因。

四、編碼:從前面過程我們可以看到，顏色轉換完成到編碼之前，圖像並沒有得到進一步的壓縮，DCT 變換和量化可以說是爲編碼階段做準備。

編碼採用兩種機制:一、是 0 值的行程長度編碼; 二、是熵編碼（EntropyCoding）。

RLE(行程長度編碼):

RLE(行程長度編碼)壓縮算法是 Windows 系統中使用的一種圖像檔案壓縮方法, 其基本思想是: 將一掃描行中顏色值相同的相鄰像素用兩個位元組來表示, 第一個位元組是一個計數值, 用於指定像素重複的次數; 第二個位元組是具體像素的值。主要通過壓縮除掉數據中的冗餘位元組或位元組中的冗餘位,從而達到減少檔案所占空間的目的。

熵編碼法:

一種主要類型的熵編碼方式是對輸入的每一個符號,建立並分配一個唯一的字首碼,然後,通過將每個固定長度的輸入符號替換成相應的可變長度字首無關輸出碼字替換,從而達到壓縮資料的目的。每個碼字的長度近似與概率的負對數成比例。因此，最常見的符號使用最短的碼。

(三)研究方法:

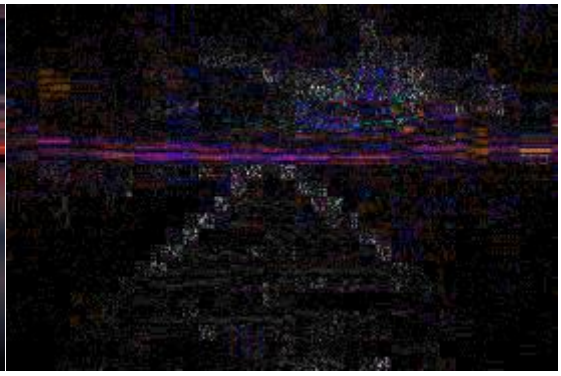
在判別影像上 我們使用 Dr. Neal Krawetz 提出的錯誤級別分析 ELA ，ELA 是一種取證方法 用於識別具有不同壓縮級別的圖像部分，該技術可用於確定圖片是

否已被修改 ELA 以 95% 的壓縮率重新保存原始圖像，並評估與原始圖像的差異。

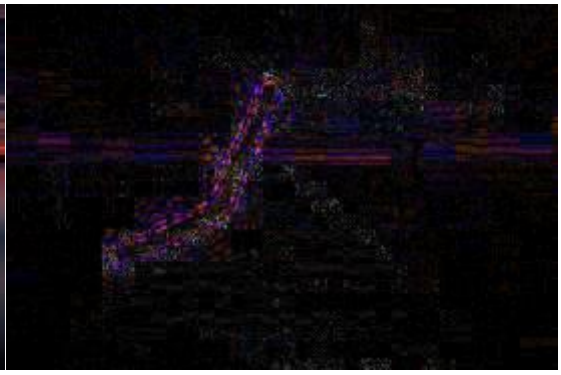
在圖像中顯示為較暗的區域，純色呈現出良好的壓縮水平，具有較小的 ELA 值，也通常代表沒有被修改過的部分。 ELA 突出圖像中的更改部分，給予較高的 ELA 值並在圖像中呈現亮色。

適當的使用 ELA 分析可以輕鬆發現圖像修改(包括縮放、裁剪和重新保存操作)，但 ELA 分析取決於圖像的質量，多次重新保存的圖片是無效的。 如果圖像被多次重新保存，那麼它可能有一個最小的錯誤等級，更多的重新保存不會改變 ELA 的圖像。 ELA 將返回黑色圖像，並且不會檢測到任何修改。 該技術在發現使用 Photoshop 或 Gimp 等工具方面非常有效。 僅使用這些應用程序保存圖片，用戶就會在圖像中得到更高級別的潛在錯誤，之後我們將使用 ELA 得出的圖片進行 CNN 分析。

當圖像被改變時，特定部分的壓縮比會相對於其他部分發生變化，（ELA）和神經網路進行結合，來判定特定圖檔(JPG)有沒有被改變。原始圖片會給定 ELA 值，對於後續針對圖片的操作，會使整體 ELA 值慢慢降低。淺色 ELA 的值會比較高，而深色 ELA 較低，通常是背景原色，因此如果圖片中的某一部分與周邊有極為顯著的誤差等級差別，圖片就可以判定為修改過。



原始圖片　　經過 EVL 轉換後的圖片

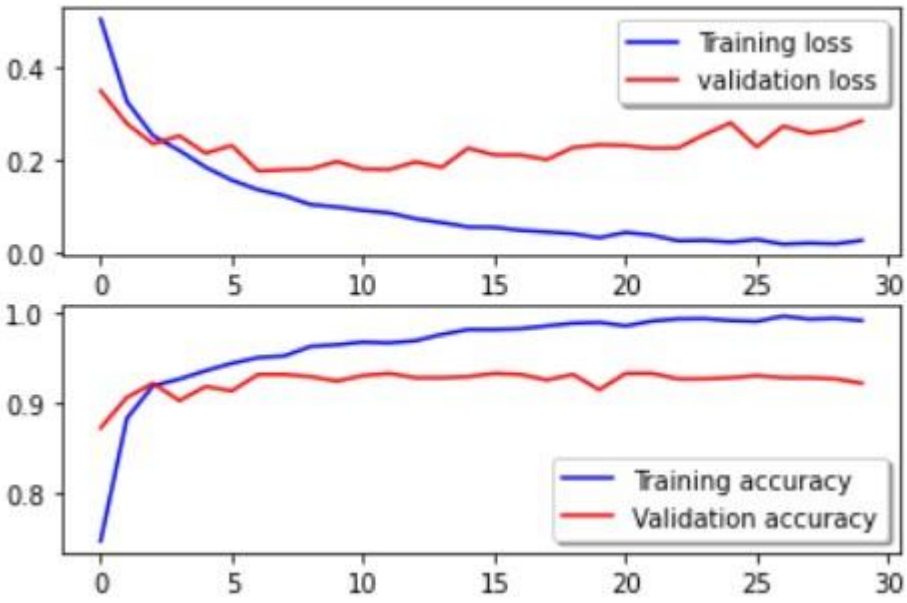


圖片修圖一次，明顯看到右圖，整體顏色降低，修改後的地方 EVL 值變高

把四維資訊(多張圖片、圖片的寬和高、圖片的 RGB)變成一資訊(圖片數量 *128*128*3) 核心(kernel size)使用 5 作為大小，經由 30 個輪資料迭代(epochs=30)，在訓練過程中每次選擇 32 批量(batch size)來進行處理，使用的損失函數為(loss)為 binary_crossentropy，而當超過兩個資料迭代(patience)精準度沒有進步，提早結束退出。

(四)預期結果

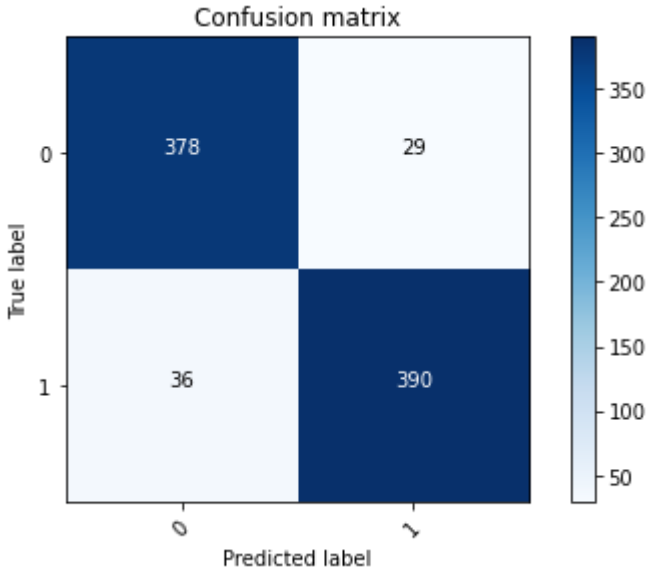
在模型的訓練上我們不希望 他過適(over-fitting)或是 under-fitting 過適的模型在訓練集上表現良好但在測試集上表現不佳，這意味著該模型在涉及新數據時似乎無法泛化 所以要提早結束(early stop) 右圖的上面是訓練和驗證的疏誤 及下面的是訓練和驗證的精準度，從下面的折線圖中，可以看出資料的訓練狀況



透過資料進行判讀，從而得到以下結果和指標

混淆矩陣

從資料集當中隨機選取圖片建立模型，利用 convolutional neural network 建立資料模型，使用 confusion matrix 做影像識別，判斷有修過圖跟沒修過圖的數量，與實際數量做比對，從而判斷系統有沒有得分，最後慢慢的接近真實結果



如圖，深藍色區域為得分，機器將會慢慢將結果全部導向深藍色區塊（預測值等於實際值）

(五)結論

本報告介紹了 deepfake 的危害以及現代其他的影像辨識技術、ELA 使用的 JPG 技術，並針對圖片的內容使用人工智慧進行判讀，且使用一維資料進行神經網路運算的實驗結果。

現在假圖片、假新聞盛行，本報告使用 ELA 搭配神經網路，可以得知資料壓縮的分布，從而讓電腦去判讀圖片是否有被修改過，從而讓錯誤資訊得以識別。資料處裡完畢後，針對效能進行評估。

(六)資料來源

Nor Bakiah Abd Warif; Mohd. Yamani Idna Idris; Ainuddin Wahid Abdul Wahab; Rosli Salleh, An evaluation of Error Level Analysis in image forensics

深偽技術- 維基百科，自由的百科全書

Deepfake 到底是什麼？

JPEG - 圖形檔案壓縮格式 - 國家教育研究院雙語詞彙

Deepfake 技術親手實驗

Deepfake 應用新例！

行程長度編碼(RLE)定義 - 中文百科全書

熵編碼法- 維基百科，自由的百科全書

圖像 JPEG 格式介紹 - 人人焦點

afsalashyana/FakeImageDetection: Fake Image Detection Using Machine Learning (github.com)

casia dataset | Kaggle

bh-usa-07-krawetz-wp.pdf (hackerfactor.com)

Error Level Analysis in Python (github.com)

Error Level Analysis As A Guide Mask For Robust Deepfake Detection__臺灣博碩士論文知識加值系統 (ncl.edu.tw)

qumuase/ELA: ELA 全称：Error Level Analysis ，汉译为“错误级别分析”或者叫“误差分析”。通过检测特定压缩比率重

新绘制图像后造成的误差分布，可用于识别 JPEG 图像的压缩。 (github.com)

#ELA 分析#可能是讓 X 姓男星人設漸崩的有力一錘_效率火箭 - 微文庫 (gushiciku.cn)

Fake Image Detection Using Machine Learning | Yogesh Gaikwad and Kuncheria Kuruvilla - Academia.edu

ELA Photo Forensics - eForensics (eforensicsmag.com)

Spot faked photos using digital forensic techniques | Popular Science (popsci.com)

Clone Detection - an overview | ScienceDirect Topics