

## 摘要

人工智慧盛行在我們的時代，啟能悄悄從它身邊走過, 偶然的接觸到人工智慧，竟然完全無招架之力的被那科技深深吸引，開啟短暫但我相信還有續集的探索之旅。

一、卷積神經網路(Convolutional Neural Network, CNN)：一種用於圖像辨識和分類的深度學習模型。MNIST 手寫數字辨識和 Cifar-10 圖片判斷都是基於卷積神經網路實現的。

二、YOLO 技術：這是一種物件檢測技術，可以實現快速而準確地檢測圖像中的物體。可製作判別貓狗位址和自動瞄準外掛都是基於YOLO實現的。

三、3D虛擬人物控制：MediaPipe 是 Google Research 開發的一個跨平台的機器學習框架，可以實現從相機、影像等輸入中擷取關鍵信息並進行處理。進一步使用 MediaPipe 模型操控3D虛擬人物。

四、神經風格轉換(Neural Style Transfer)：將一張圖像的風格轉移到另一張圖像上的技術。

五、生成對抗網路(Generative Adversarial Network, GAN)：這是一種利用兩個神經網路模型相互競爭來生成逼真的圖像的技術。GAN 的變體 CycleGAN 可以更加快速地實現風格轉換，GAN 這種技術在許多領域都有廣泛的應用。

## 前言

### 一、研究動機

世代的轉換代表的是成長、變動，更是恐慌。農業時代進階到工業時代、工業時代進階到AI時代，都代表著一群企業及人的沒落，同時也創造了另一批企業及人的興起，而現在正值AI時代要進階到人工智慧的時代，不想被這波洪流淹沒，則要加緊腳步迎頭趕上，也正是時機讓我們有個創造的舞台！處在這樣的時間及轉捩點，對我來說是非常興奮的。

### 二、研究目的

- (一) 研究人工智慧所擁有的有趣技術
- (二) 辨識系統的應用
- (三) 虛擬人物的控制
- (四) AI繪圖原理

### 三、人工智慧 (Artificial Intelligence, AI)

1956年，人工智慧被確立為一門學科，半世紀間經過許多起起落落。如今電腦的運算能力約為 30 年前的 100 萬倍，且近10幾年大數據的快速發展，人工智慧重新過來，許多先進的機器學習技術成功應用於社會中的許多問題。

### 四、機器學習 (Machine Learning, ML)

而在AI底下有個分支，也就是這次的主題機器學習。從 1980 開始蓬勃興起。機器學習之所以能興起，也歸功於硬體儲存成本下降、運算能力增強、大數據的發展。而機器學習中又有4類的學習方式，分別為監督學習、半監督學習、無監督學習、強化學習。

### 五、深度學習 (Deep Learning, DL)

深度學習又是機器學習的分支，深度學習能自動提取資料特徵，其能力遠遠甩開其它演算法。深度學習參考人腦神經概念，用程式還原神經網路的構造，人工神經網路架構分為 輸入層(input layer)、隱藏層(hidden layer)、輸出層(output layer)。輸入層是資料進入系統的入口，而隱藏層是處理資訊的地方，隱藏層從輸入層或其他隱藏層取得輸入。人工神經網路可以有大量的隱藏層。每個隱藏層分析前一層的輸出，進一步處理，並將其傳遞給下一層，重複直到輸出層，而最終的計算結果就會顯現在輸出層，也就是預測結果。

### 六、神經網路種類

(一) 前饋神經網路 (Feedforward Neural Network, FNN)  
是最古老的神經網路之一，最簡單的神經網路模型，資料經由輸入層通過隱藏層到輸出層，神經元之間沒有連接迴路存在。

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

卷積神經網路通常用於圖片辨識，模仿人類大腦的認知方式，觀察由細微的事物到整體特色。卷積層 (Convolution Layer) 使權重的減少、池化層 (Pooling layer) 壓縮圖片，以此更高效率的判斷圖片。

1. 卷積層 (Convolution Layer)

將輸入的圖像劃分為若干個矩形區域，對每個子區域以相同權重運算，最後加上激勵函數。神經元運算中無須每個輸入都要一個權重，我們稱 共享權值 (Shared weights)，可大幅減少權重數量，藉此減少運算時間。

2. 池化層 (Pooling Layer)

一個壓縮圖片並保留重要資訊的方法，取圖片範圍內最高或平均當做輸出，常用的有最大池化 (Max pooling) 與 平均池化 (Average pooling)。

3. 扁平層 (Flatten Layer)

將多維的輸入壓扁為一維輸出，常用在從卷積層到全連接層的過渡。

4. 全連接層 (Fully Connected Layer)

連接最基本的神經網路。

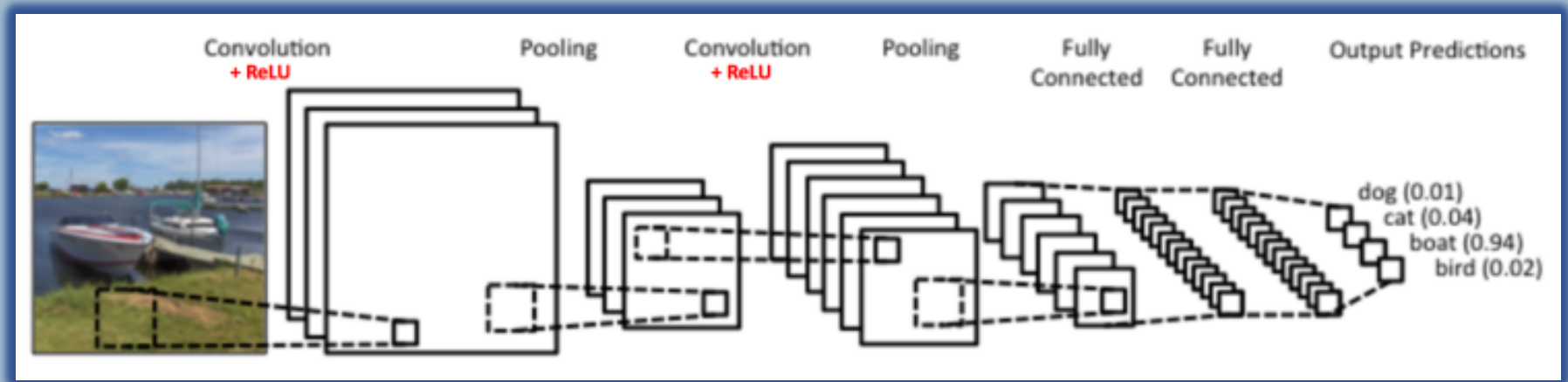


圖1 卷積神經網路架構

(三) 遞迴神經網路 (Recurrent Neural Network, RNN)

最常被用來處理時間和序列相關的問題。與使用前饋類神經網路不同的是，循環類神經網路具備前一層事件的「記憶」，並附加到目前層的輸出內容。

(四) 長短期記憶網路 (Long Short-Term Memory, LSTM)

是進階的遞迴神經網路，解決許多問題。長短期記憶網路會透過三個控制閥(輸入閥、遺忘閥、輸出閥)來決定將什麼資料保存(記憶)下來，而什麼記憶又該捨棄(遺忘)。看似不錯但也因為家人入了許多內容導致參數變多，訓練難度提升了不少。

(五) 生成對抗網絡 (Generative Adversarial Network, GAN)

生成對抗網路是種非監督式學習，主要是兩個相互競爭的神經網路 生成網路 (Generative Network) 與 判別網路 (Discriminative Network)。生成網路生成圖片，目標騙過判別網路，判別網路判斷是否與資料相同，目標提升鑑定水準，這樣一來一回的對抗促使兩邊互相成長。

### 七、YOLOv8

YOLO (You Only Look Once) 是一種物件偵測方法，目前共推出8個版本。YOLO 的主要優勢是其快速的運算速度，能夠及時處理圖像。YOLOv8剛好在2023登陸，既然是最新版本，運算成本應該較低，因此選用 YOLOv8。

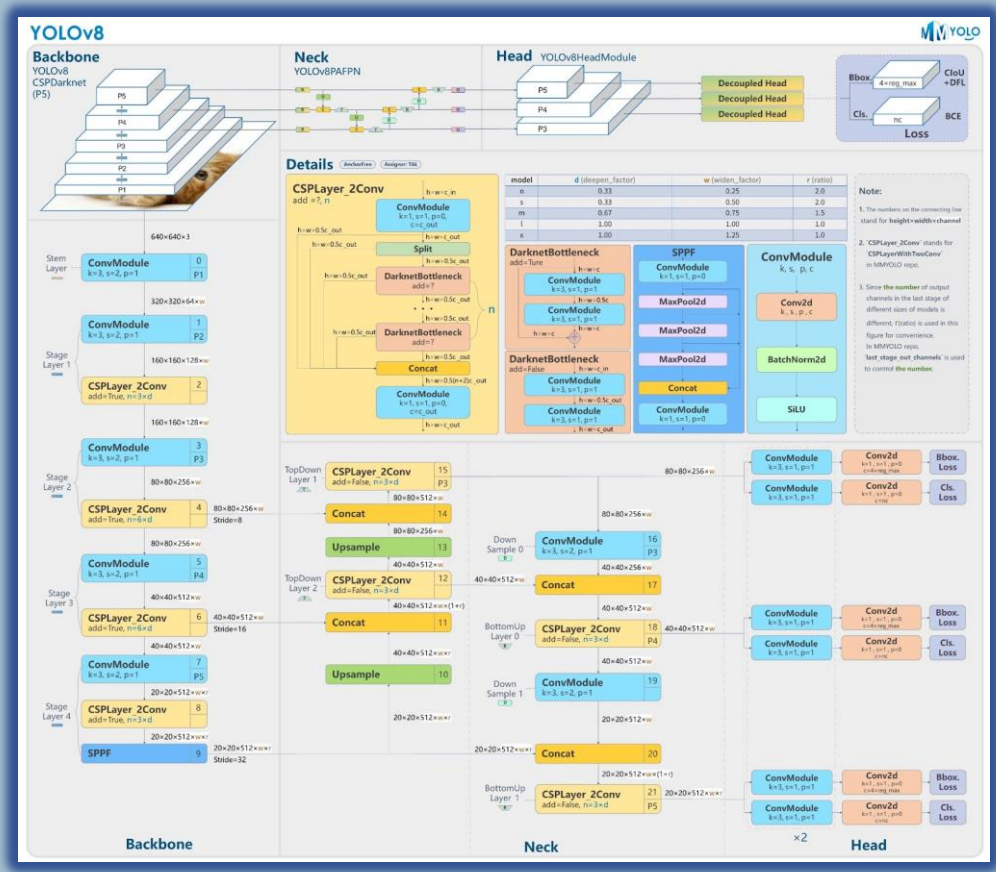


圖2 YOLOv8 架構

## 研究設備及器材

### 一、硬體設備

(一) 桌上型電腦

- 作業系統：Windows 10
- CPU：Intel Core i7-12700K
- GPU：NVIDIA GeForce RTX 3060
- 記憶體：32 GB

(二) Logitech C310 HD 網路攝影機

### 二、軟體工具

- (一) Python 3.9：程式語言
- (二) C#：程式語言
- (三) Unity：遊戲引擎
- (四) CSGO：射擊遊戲(測試用)
- (五) Anaconda：虛擬環境
- (六) Kaggle：數據建模和數據分析競賽平台
- (七) Roboflow：線上圖片標註



## 機器辨識

大家想到大數據總是又愛又恨，愛的是他的功能強大，只要有足夠的數據，能在幾毫秒得到所需答案，比過去數萬人、同時作業數年的效果更快更精準，相對的也可能涉及到隱私問題。機器視覺中的人臉辨識已是最基礎的，各種生物、物品的辨識所造成的人力精簡及時間的省節，都可藉由深度學習來達成，且準確度更不會有人為的疏失！

### (一) MNIST、Cifar-10 辨識

MNIST 是一種黑白的手寫數字資料集，分類為 10 類 (0~9)，在踏入CNN這領域時，大多數人都會使用這資料集當作練習，可以將解決 MNIST 視為深度學習的 "Hello World!"。

Cifar-10 資料集內含 10 種類別的圖片，分別是飛機、汽車、鳥、貓、鹿、狗、青蛙、馬、船、卡車，且屬於彩色圖片。

#### 1. 實際執行

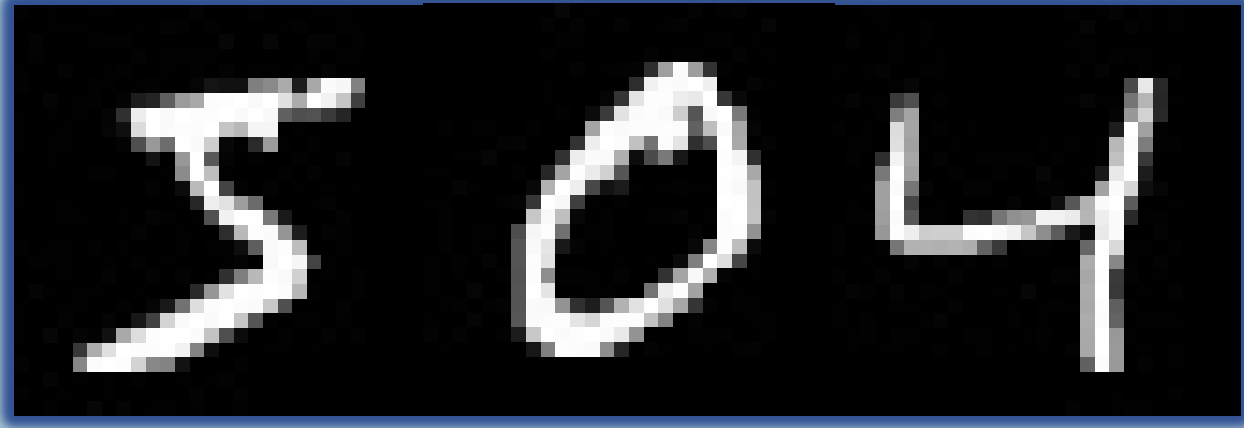


圖3 偵測為5、0、4

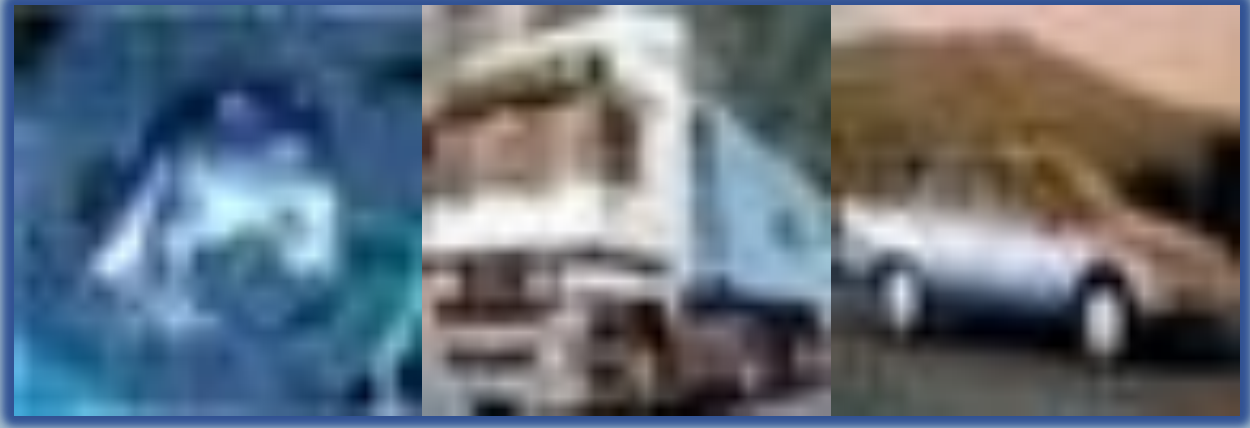


圖4 偵測為青蛙、卡車、汽車

#### 2. 結論

第一次辨識成功後，我非常驚訝只用數學運算就可辨識物件，難怪現在深度學習如此有人氣，想不到未來這技術能有多強大，真的很佩服想到深度學習的人。

### (二) 貓狗偵測

試驗 YOLOv8 技術，採用 yolov8l-seg 訓練模型，較講求精準。

#### 1. 實際執行



圖5 貓狗

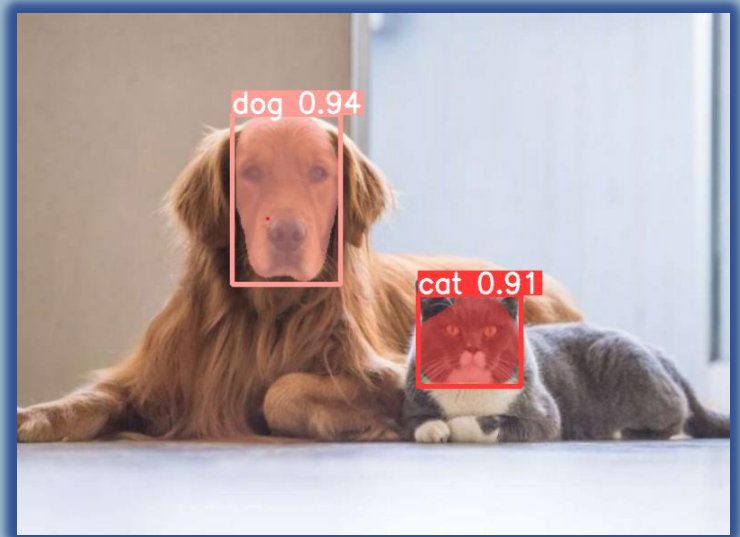


圖6 貓狗偵測

#### 2. 結論

製作資料集時手真的差點抽筋，非常痛苦的三小時，其餘都相對簡單。意外的是，只需120張圖片就可做出正確率如此高的辨識模型，讚嘆 YOLO。

### (三) CSGO 自動瞄準

在逛 Roboflow 網站時，意外找到 CSGO(射擊遊戲) 人物偵測的資料集，馬上意識到這資料集是為了外掛而生。使用 YOLO 預測人物位置，再移動鼠標至人物，加點隨機數，這不就是個無法防範的外掛(一般外掛以破解遊戲內部資料為主)。

#### 1. 實際執行

準備就緒後，開一場電腦場(只有我一個玩家)，不出意外的話就要出意外了。偵測不給力、時間差、容易偵測到隊友、鼠標移動，種種問題浮現，主要還是偵測太慢且不准。輸入到偵測完成，敵人早已移動到下個位置了，如果能一直偵測到，這問題就可以用預測的方式解決，但這偵測模型只要距離稍微遠一些，就完全偵測不到了。



圖7 SCGO畫面

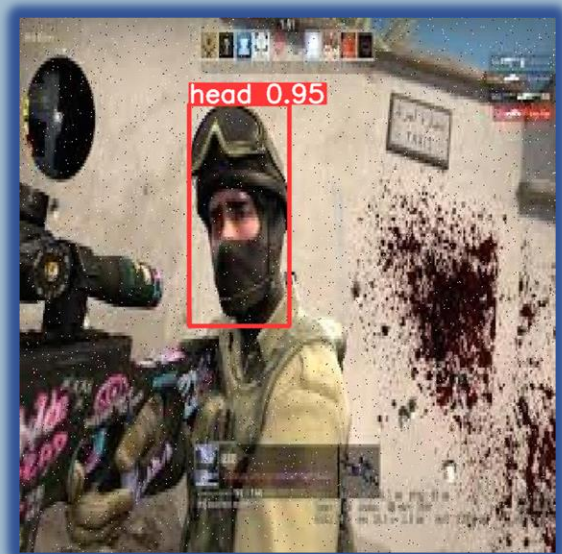


圖8 SCGO偵測

#### 2. 結論

可說是完全的失敗，不過我相信只要優化過偵測模型，還是能成為很好的外掛。但為什麼這種外掛模式不怎麼聽過，很明顯效益不高，與傳統外掛相比，傳統外掛直接竊取內部資料簡單、快速、準確、更強。相反這種外掛還需要看到敵人才能有所反應，各項技能皆差距傳統外掛一大截，唯一好處是難被抓包，不過遊戲端應該也可反用AI抓這種外掛。

## 總結

透過最基礎的卷積神經網路技術的測試以及進一步的應用，可以發現其在辨識與定位方面的應用潛力非常大。除了手寫數字辨識、貓狗辨識等基礎應用外，還可以應用在許多重要的場景中，例如自駕車、軍事、失蹤人口協尋等等。然而，應用卷積神經網路技術時也必須注意遵守相關法律法規以及道德規範，確保技術應用的合法性與合理性。總體而言，卷積神經網路技術的應用前景廣闊，將對未來的生活、工作、社會等方面產生深遠的影響。

## 藝術

透過AI的協助更可擴大藝術家、非藝術家的創意發想，只要想得到的，不只都能創作出來，且可發揮比想到的更大更精準的效果，並引領創作者達到另一個層次與想像！換句話說，也不再需要藝術家了，人人都可以透過此設計就比藝術家更厲害了。

### (一) 神經風格轉換 (Neural Style Transfer, NST)

過去幾個月流行將現實圖片轉換成動漫風，而我也被這技術驚豔到，因此學習神經風格轉換。

#### 1. 實際執行

選擇本人照片當內容圖，而使用 梵谷《星夜》(The Starry Night)、康丁斯基《第七號構圖》(Composition VII)、葛飾北齋《神奈川沖浪裏》(The Great Wave off Kanagawa) 為風格圖。



圖9 風格轉換

#### 2. 結論

神經風格轉換為神經網路影像風格轉換的開山始祖，但每次要生成一張新的風格化影像需要重新訓練一次 VGG-19，效率頗低，且完全不能即時轉換。

### (二) DCGAN

發現神經風格轉換無法實現即時轉換後，嘗試找出何種方式可即時轉換。雖然當時沒找到，但找到生成對抗網絡這有趣技術，只要給無標註的圖片，就可隨機生成出同類別圖片。

#### 1. 實際執行

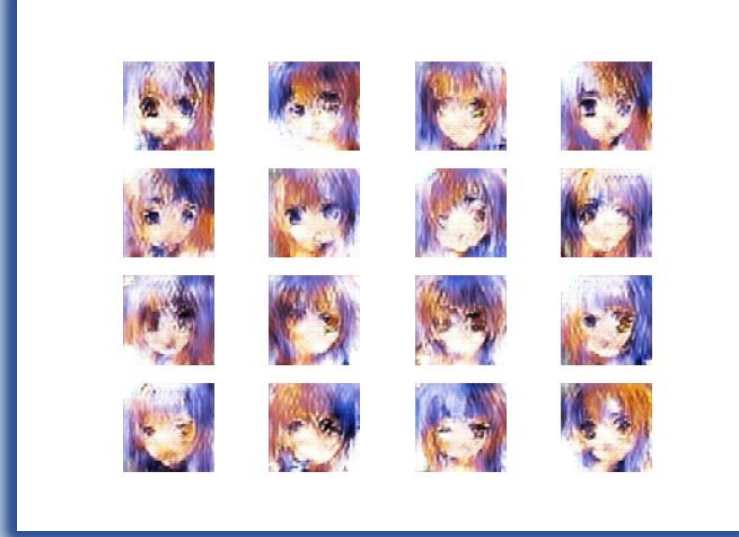


圖10 版本一 大小64\*64

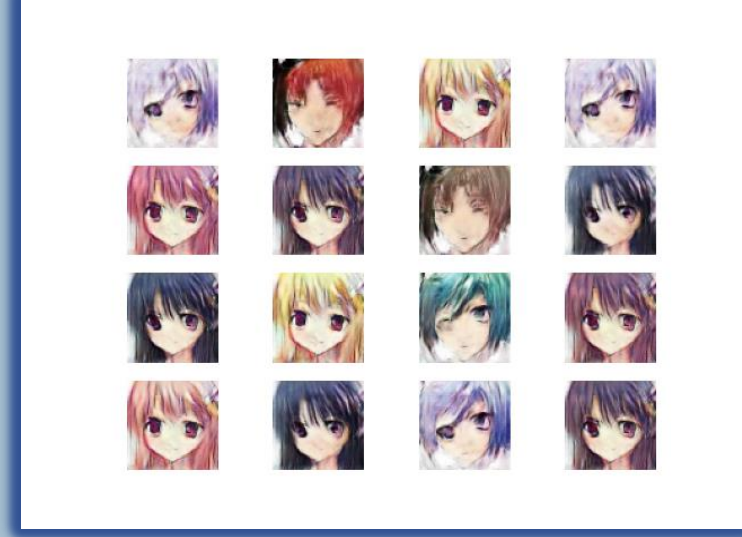


圖11 版本二 大小64\*64



圖12 版本一 大小256\*256



圖13 版本二 大小256\*256

#### 2. 結論

只能說待加強，感覺模型還可再增強、學習率也可再調，生成對抗網絡是由兩個模型互相比較而來，兩個模型要一起成長、進步，不能其中一個的能力強過多於另一個，因此微調其中參數也是很重要的一環。

### (三) CycleGAN

生成對抗網絡有非常多變體，多到以英文字頭+GAN取名的變體都快佔滿了，當中我無意看中 CycleGAN，希望藉此達成即時轉換功能。

#### 1. 實際執行



圖14 現實轉莫內圖片

#### 2. 結論

DCGAN 都如此難訓練了，更別說 CycleGAN 了。原理是很有趣、巧妙，但產生的圖形缺乏多樣化，訓練時損失震盪巨大，成效差強人意，並不是太有效的風格轉換方式。

## 總結

使用 GAN 技術進行影像風格轉換是一種很有前景的應用，它可以讓我們將一張圖片的風格轉換成另外一種風格，例如將印象派畫作的風格應用到一張照片中。儘管成果目前並不十分樂觀，但這個技術仍有很大的發展空間，特別是隨著AI技術的不斷進步，未來我們可能能夠更好地掌握影像的細節，進一步提高影像風格轉換的效果。

這項技術對美化世界有很大的潛力，它可以幫助藝術家以更快、更有效的方式創作出令人讚嘆的作品，同時也可以創造出更多具有價值的商品。除此之外，影像風格轉換還可以應用在影視、遊戲等領域，讓觀眾享受到更多美感和豐富的視覺體驗。總之，這項技術有很大的潛力，值得我們繼續關注和發展。



