中華民國第 62 屆中小學科學展覽會

作品說明書

高級中等學校組　電腦與資訊學科

探究精神獎

052502

基於影像風格轉換之資料增強

學校名稱：國立新竹科學園區實驗高級中等學校

作者：	指導老師：
高二 曾亭翰	廖祐梓

關鍵詞：影像風格轉換、物件辨識模型、資料增強

摘要

本研究利用影像風格轉換模型作為資料增強的方法，將原始資料集中的影像轉換成不同風格，以增加訓練資料的多樣性。為了驗證使用此方法的有效性，本研究進行了兩個領域的實驗，分別是：(1)物件辨識模型在不同天氣時間的街景圖上的表現，(2)胃腸癌中的腫瘤檢測模型在不同染色結果的組織圖上的表現。結果證實使用風格轉換模型生成的影像訓練的模型的準確率在某些情況下有顯著的提升。此方法的優點在於能夠快速產生多樣風格的資料，由於是對影像的風格做轉換，影像的內容沒有改變，因此能夠沿用原有的標籤，同時節省了蒐集及為新影像標籤的人力及時間。

壹、 研究動機

我觀察到目前我們在收集用來訓練模型的訓練資料時經常會出現資料不平衡的狀況，例如在收集自駕車的物件辨識模型的訓練資料時，收集到的影像多數為在晴天下拍攝的，較少影像為雨天或有霧等等情形。而對於醫學影像，經常會將要辨識的組織染色，如蘇木精-伊紅染色，又稱「H&E染色」（hematoxylin and eosin stain、H&E stain），是組織學最常用的染色方法之一。由於不同的醫療單位使用的染劑會有些微不同，導致染色出來的結果也有所不同。以上兩種例子都會導致模型在訓練時看似有很高的準確率，然而當模型在現實世界進行辨識時，如果出現與當初訓練資料分佈相差較大的情形，模型就會產生誤判。

如果想要解決上述問題就需要為模型收集更完善的資料集，但收集以及為這些新的資料標籤將會耗費大量的人力。因此我希望能找到一種能夠自動產生不同風格訓練資料的資料增強方式，節省不必要的人力花費。

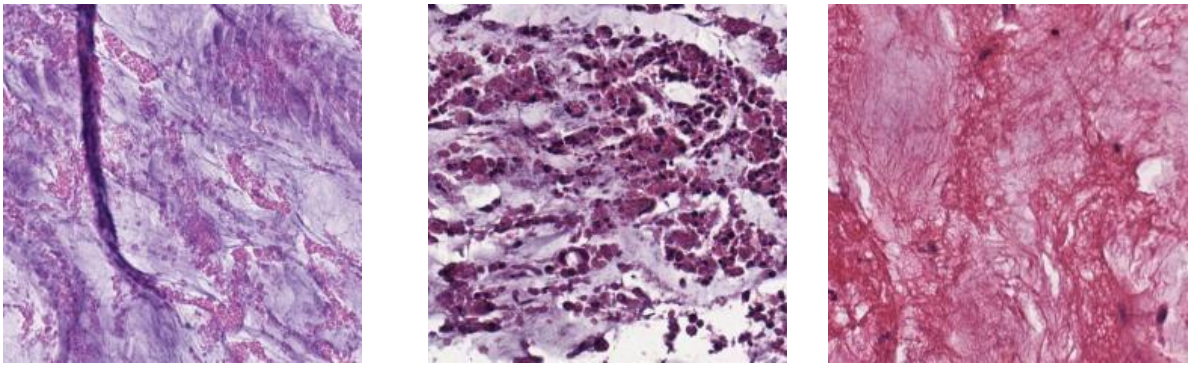


圖1： 來自不同醫療機構的H&E組織染色染色圖

貳、 研究目的

本研究目的主要聚焦在運用影像風格轉換模型的技術去做資料增強，在保留原影像內容物的情況下將其轉換成不同的風格。此方法能快速地透過模型產生不同風格的資料，藉由重複利用原資料的標籤，讓同一份訓練資料的價值增加，省去收集新的資料以及為新資料標籤的人力及時間。

參、研究設備與器材

一、軟體環境：

1. Google Colab
2. Visual Studio Code + Anaconda

二、硬體環境：

1. Google Colab上的虛擬機
2. MSI GF65 9SD(筆電)：

cpu：I7-9750H

gpu：1660Ti

肆、 研究過程與方法

一、 文獻探討：

本研究提出一基於影像風格轉換之資料增強方法，此章節將介紹在電腦視覺中常見的資料增強方法，以供比對。

資料增強(Data Augmentation)是指通過有限的的資料集產生更多等價資料來擴充訓練資料集的技術。神經網路需要大量且多樣的訓練資料以避免過擬和，然而許多現實生活中的場景無法快速取得這麼完整的資料集，因此需要藉著資料增強的輔助來提高模型的表現。

電腦視覺的資料增強方法大致可分成兩類：

1. 基於影像處理：

此類的資料增強主要著重在對原影像做幾何修改，常見的例子如旋轉、隨機裁切、顏色通道的轉換。

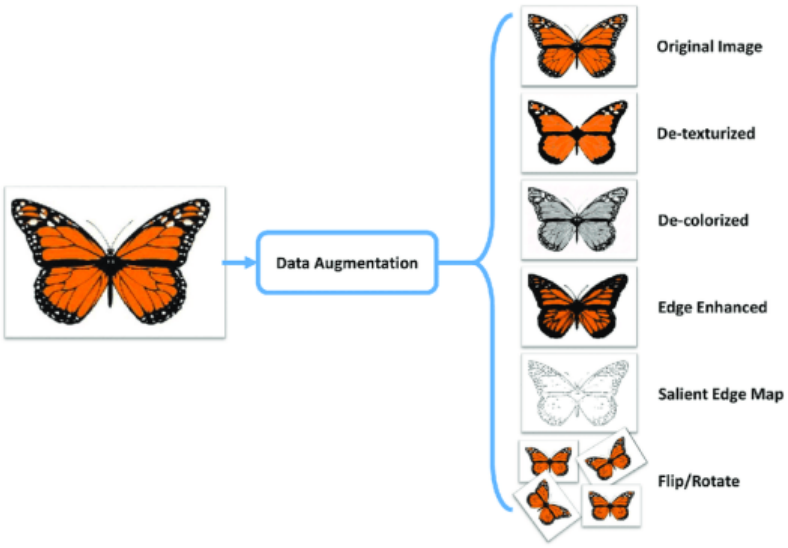


圖 2：常見的影像處理資料增強 [5.]

2. 基於深度學習：

此類的資料增強運用深度學習模型對原影像修改，常見的例子如特徵空間增強、對抗生成、影像風格轉換。

圖 3：基於風格轉換的資料增強

本研究分別使用指定風格轉換之資料增強於不同天氣時間下的街景圖資料集，與隨機風格轉換之資料增強於來自不同研究機構的蘇木精和伊紅 (H&E) 染色的人類結直腸癌 (CRC) 和正常組織的組織學圖像。

二、 街景圖資料增強：

(一)、 研究架構及流程

街景圖資料增強的研究流程如下圖所示，本章節將針對流程中各部分進行詳細介紹

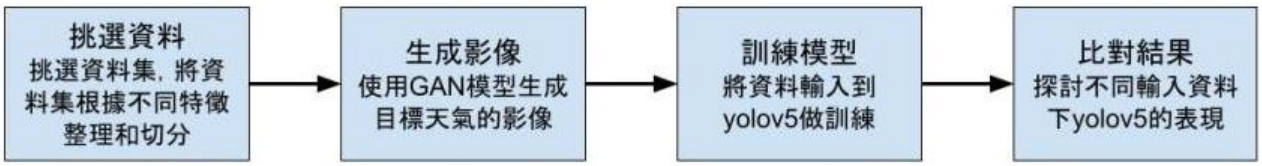


圖 4：實驗流程圖

1. 挑選資料：

(1.) BDD100K：

本研究使用BDD100K作為街景圖資料來源，BDD100K是UC Berkeley於2018年公布的駕駛視訊資料集，能用來做圖像標記，車道檢測，可駕駛區域分割，道路對象檢測，語義分割，實例分割，多對象檢測跟踪，多對象分割跟踪，領域自適應和模仿學習等任務。本研究使用其中的道路對象檢測資料集，由於此資料集有提供拍攝當下的天氣及時間，相較其他街景圖資料集更適合作為本研究的資料來源。

(2.) 篩選資料：

使用python將資料根據時間、天氣、影像中物體(道路對象)分類

<1.> 時間、天氣

晴天定義：weather == "clear" and timeofday == "daytime"

雨天定義：weather == "rainy" and timeofday == "daytime"

夜晚定義：weather == "clear" and timeofday == "night"

Frame attributes

```
- weather: "rainy|snowy|clear|overcast|undefined|partly cloudy|foggy"
- scene: "tunnel|residential|parking lot|undefined|city street|gas stations|highway|"
- timeofday: "daytime|night|dawn/dusk|undefined"
```

圖 5：BDD100K影像標籤

<2.> 影像中物體

將不同天氣和時間的影像分開之後，發現每個類別的數量分布十分不平均，為了避免模型發生過擬和，因此將原本的10類減少成數量最多的三類，分別是car，traffic sign，traffic light。

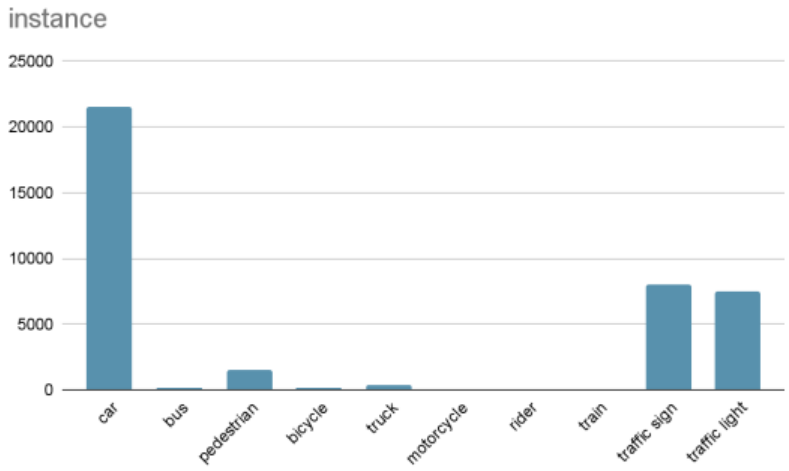


圖 6：2000張night影像中各個類別的數量

最終使用的影像為符合天氣及時間條件且該影像中有出現car，traffic light，traffic sign任一個類別之影像

情況	晴天	雨天	夜晚
影像數量	2750張	2521張	2750張
car	30442	23679	24988
traffic sign	9536	9037	8337
traffic light	6068	8072	7664

表 1：本研究用到的全部影像資料。car、traffic sign、traffic light的數值代表此類別的數量

2. 生成影像：

(1.) 對抗生成網路 Generated Adversarial Network（GAN）：

生成對抗網路簡稱GAN，是屬於機器學習中無監督學習的一種方式。對抗生成網路同時訓練兩個模型：一個生成模型 G 用於捕捉真實樣本資料的分布，從潛在空間中(如高斯分布、均勻分布等等）生

成資料，目標是生成出的樣本跟真實樣本越像越好。另一個判別模型 D 用於估計一個樣本來自真實資料的機率，目標是正確判斷一個輸入樣本是來自真實樣本還是生成模型生成的假樣本。藉由輪流優化兩個模型的損失函數，讓生成模型與判別模型不斷的相互對抗，在合適的時間停止訓練後，就能得到一個能生成出與真實樣本分布相近的生成模型。

可將上面的敘述用下列公式與圖表示：

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log 1 - D(G(z))]$$

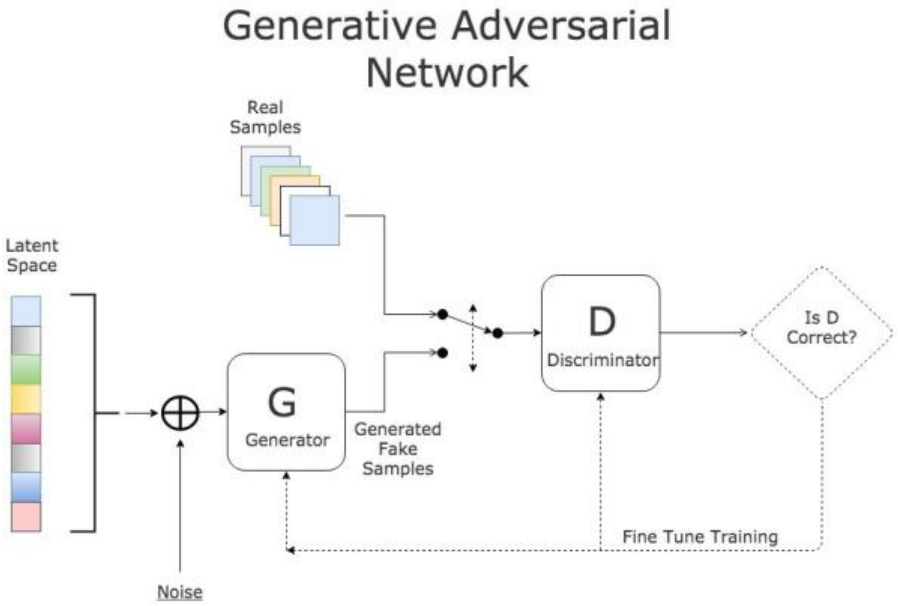


圖 7：生成對抗網路模型示意圖 [7.]

(2.) 風格轉換 Style Transfer：

風格轉換是指在保留影像內容結構(content)的前提下，將影像的風格(style)轉換成其他風格。通常透過卷積神經網路擷取原影像的content和目標風格影像的style，再透過生成模型逐步混合出一張保留原影像content又擁有風格影像style的影像。

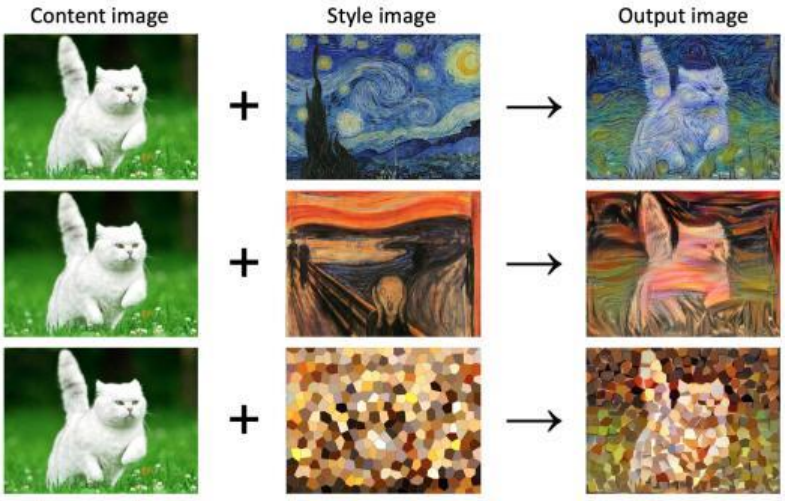


圖 8：影像風格轉換示意圖 [6.]

(3.) TSIT：

本研究使用TSIT做為風格轉換的模型，將晴朗白天的街景圖轉換成雨天與夜晚的街景圖。TSIT是一個簡單且通用的圖像到圖像轉換的模型，此模型可處理任意風格轉換，包含自然影像、現實世界場景、藝術繪畫。通過輸入一張內容影像及一張風格影像，就能生成出一張保有內容影像內容但風格為風格影像的影像。

<1.> TSIT模型架構：

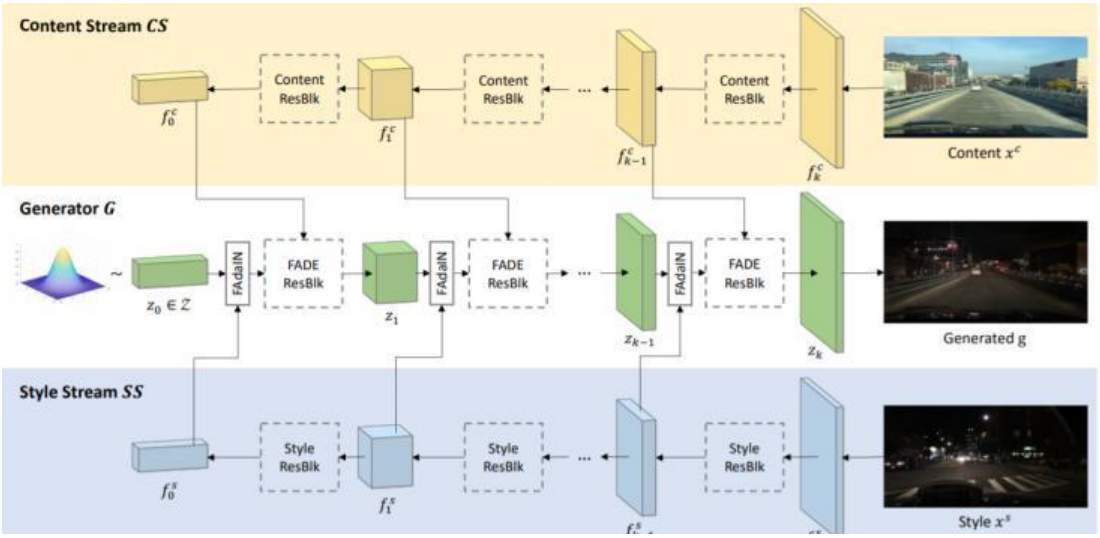


圖 9：TSIT的模型架構圖，圖中省略discriminator

TSIT模型架構主要由四個元件所構成：

- content stream：

  由標準殘差塊(standard residual blocks)組成，在content stream中稱為content residual block。如圖 7 所示，每個殘差塊有三個卷積層，其中一個是為了跳躍學習。activation fuction使用的是Leaky ReLU。content stream的目標是擷取原影像的content特徵並作為對應的generator特徵轉換層的輸入。

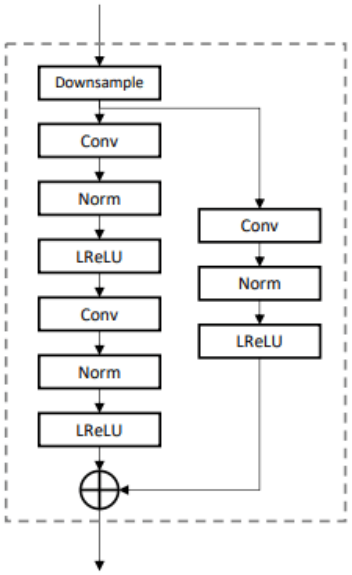


圖 10：content 和 style residual block架構

- style stream：

  style stream的架構設計與content stream一致，目標是擷取風格影像的style特徵並作為對應的generator特徵轉換層的輸入。

- generator：

  generator的架構與content/style stream完全相反，這是為了要讓generator接收到的特徵始終與此特徵在原影像中的層級相同。generator的初始輸入是一個從高斯分布中sample出來的noise，通過從content/style stream中抽取到不

同層次的特徵，不斷動態的修改影像，最後產生輸出即為風格轉換後的影像。

- discriminator：

  TSIT模型中總共使用了三個multi-scale patch-base discriminator，patch-base training能讓discriminator分別在大尺度作global information和小尺度作local detail的捕捉，通過在生成影像的不同尺度檢視其與原影像的差異，generator能生成出更精細的影像。

<2.> 特徵轉換：

TSIT使用兩個模塊做特徵轉換：

- Feature adaptive denormalization(FADE)：

  FADE模塊的主要功能是提取content影像在不同層級的特徵訊息，提取的過程是element-wise，確保generator在生成影像時有正確保留content影像中的景物。

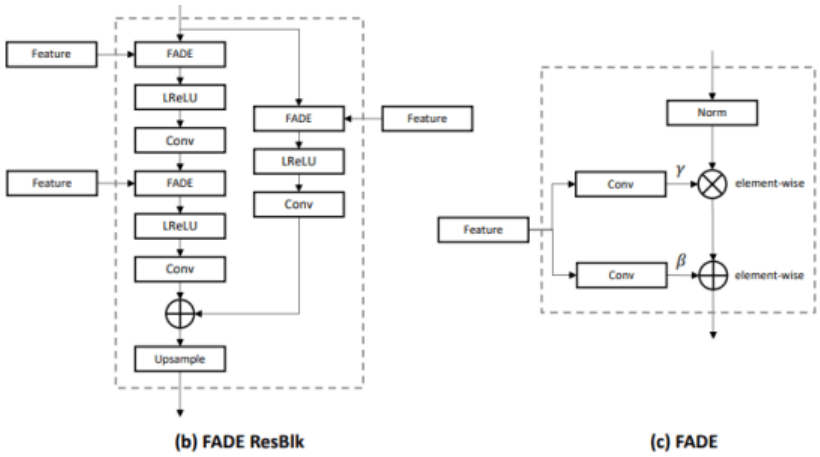


圖 11：FADE模塊架構

- Feature adaptive instance normalization (FAdaIN)：

  FAdaIN模塊的主要功能是提取style影像的特徵訊息，並動態的調整不同層級下的style特徵表現，使generator在生成影像時有附加上style影像的style。

<3.> 目標函數：

通過輪流優化generator和discriminator的目標函數，讓兩個模型不斷對抗，最終就能得到我們想要的generator。

- generator目標函數：

$$\mathcal{L}_G = -\mathbb{E}\left[D\left(g\right)\right] + \lambda_P \mathcal{L}_P\left(g, x^c\right) + \lambda_{FM} \mathcal{L}_{FM}\left(g, x^s\right)$$

generator的核心概念是希望$D\left(g\right)$的值越大越好，表示discriminator無法正確判斷生成的影像是真實資料還是生成資料。

- discriminator目標函數：

$$\mathcal{L}_D = -\mathbb{E}\left[\min\left(-1 + D\left(x^s\right), 0\right)\right] - \mathbb{E}\left[\min\left(-1 - D\left(g\right), 0\right)\right]$$

discriminator的核心概念是希望$D\left(g\right)$的值越小且$D\left(x^s\right)$的值越大，表示discriminator能正確判斷生成的影像是真實資料還是生成資料。

- 符號解釋：

$D()$為discriminator的判斷函數，其值為介於0到1之間的數值，代表輸入影像是真實影像的機率。g為generator生成之影像，z_0, x^c, x^s分別為初始輸入的noise、content影像、style影像。$\mathcal{L}_P$為計算content影像x^c與生成影像g之間的特徵差異之函數，特徵擷取使用VGG-19網路。$\mathcal{L}_{FM}$是discriminator在不同層次中檢視style影像x^s與生成影像g的特徵差異之函數。λ_P和λ_{FM}是這兩個函數的對應權重。

(4.) 生成模型實驗設計：

本研究總共訓練了兩個不同配置的TSIT模型，分別使用 50 張style image和 200 張style image做訓練，目的是想探討style image數量多寡對最後yolo模型的辨識度的影響，如下表：

name	TSIT50	TSIT200
content image數量	2750 晴天	2750 晴天
style image數量	50 雨天 50 夜晚	200 雨天 200 夜晚

表 2：生成模型設計

(5.) 模型訓練結果：

以下結果皆為模型訓練到20 epoch時的表現。

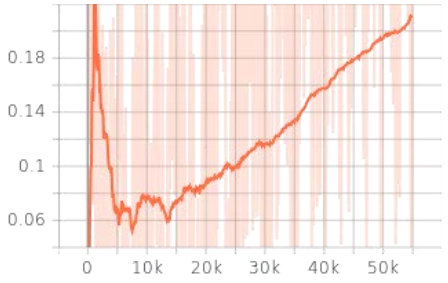
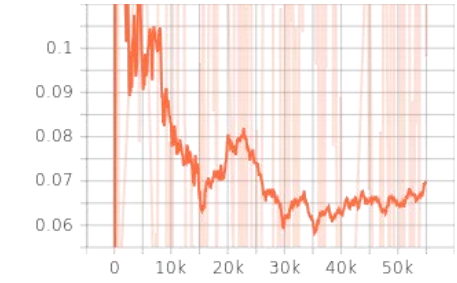
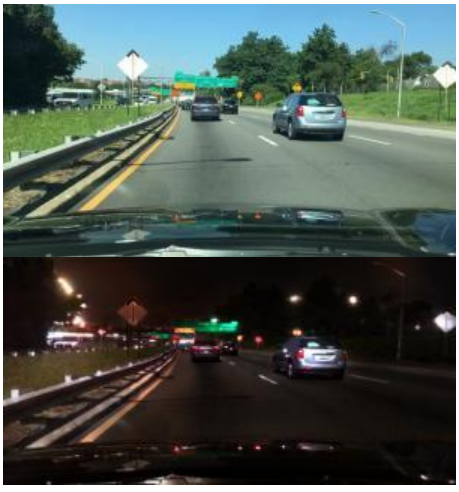
name	TSIT50	TSIT200
generator loss		
結果圖片		

表 3：生成模型訓練結果

3. 訓練模型：

(1.) 物件辨識 Object Detection：

物件辨識是電腦視覺中常見的任務，通常會在模型中搭配卷積神經網路抽取輸入影像的特徵，目標是在影像或影片中定位物體。

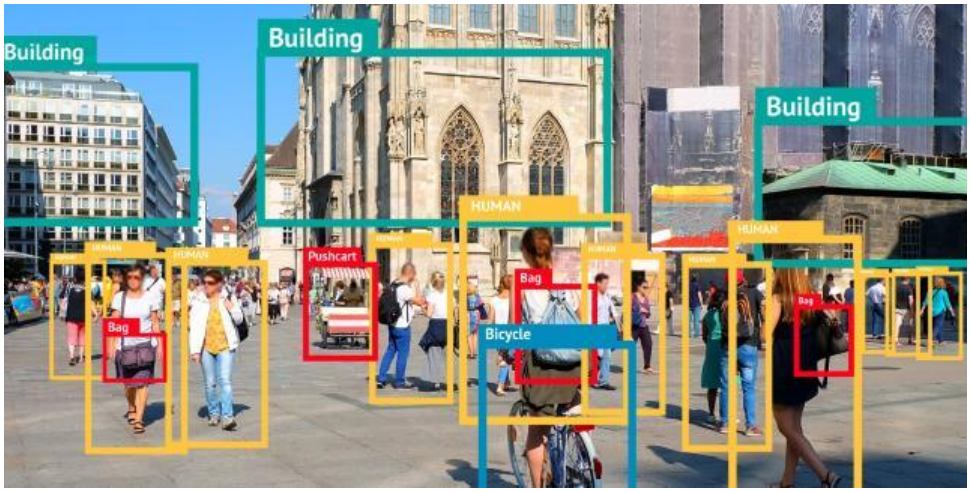


圖 12：物件辨識示意圖 [9.]

(2.) yolov5：

yolov5是一個one stage的物件辨識架構，其中包含四個版本，本研究使用的是yolov5s，為四個版本中參數最少的，訓練花費的時間也是四個模型中最快的。使用yolov5是因為one stage物件辨識模型只需要對圖片作一次 CNN架構便能夠判斷圖形內的物體位置與類別，辨識速度較多數two stage物件辨識模型快，能夠接收fps(frame per second)較高的影片作為輸入，在現實生活中的應用場景較多。

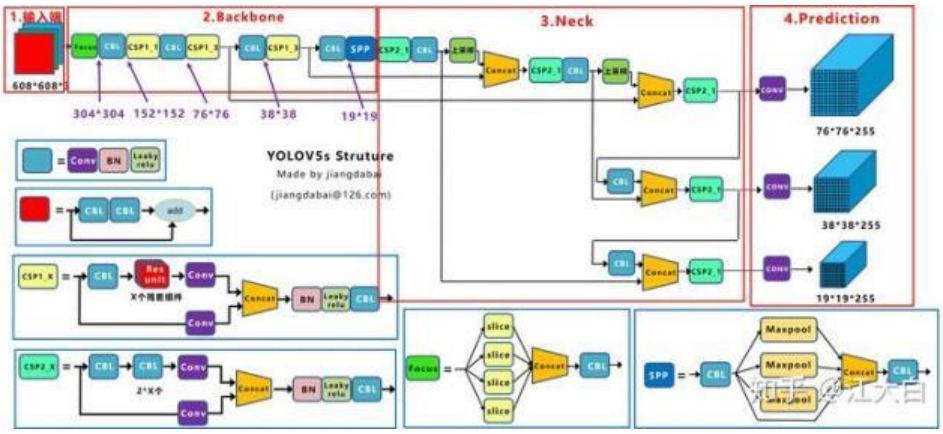


圖 13：yolov5架構 [7.]

(3.) 模型設計：

本研究為了驗證使用風格轉換生成影像對物件辨識的幫助，設計了11個不同訓練資料的yolov5模型：

name	base50	base200	transform50	transform200	supervise	art
晴天訓練	1925	1925	1925	1925	1925	1925
目標風格訓練	35	140	1960	2065	1925	1925
總訓練張數	1960	2065	3885	3990	3850	3850
晴天驗證	275	275	275	275	275	275
目標風格驗證	5	20	280	295	275	275
總驗證張數	280	295	555	570	550	550

表 4：不同模型的資料配置

資料說明：

訓練資料與驗證資料與測試資料的比例為7:1:2，其中由於雨天的資料數量較少(可見表 1)，其測試資料只有321張(為了要使其訓練資料和驗證資料與晴天和夜晚數目相同)。

<1.> base50：

模擬現實生活中資料分布極度不平均的狀況，使用2750張真實晴天街景圖與50張真實目標風格街景圖，為transform50的對照組。

<2.> base200：

模擬現實生活中資料分布極度不平均的狀況，使用2750張真實晴天街景圖與200張真實目標風格街景圖，為transform200的對照組。

<3.> transform50：

以2750張真實晴天街景圖作為TSIT模型的content，50張真實目標風格街景圖作為TSIT模型的style，生成出2750張目標風格街景圖，將以上三種資料均作為yolo模型的輸入。

<4.> transform200：

以2750張真實晴天街景圖作為TSIT模型的content，200張真實目標風格街景圖作為TSIT模型的style，生成出2750張目標風格街景圖，將以上三種資料均作為yolo模型的輸入。

<5.> supervise：

模擬現實生活中的理想情況，即資料分布十分平均。使用2750張真實晴天街景圖與2750張真實目標風格街景圖。

<6.> art：

以2750張晴天影像作為藝術風格轉換模型的輸入，生成2750張增強影像，再將晴天與增強影像作為yolo模型的輸入。

(4.) 前置作業：

<1.> 將BDD100K label轉成yolo形式：

由於BDD100K影像標籤是使用Scalabel Format形式，而yolov5使用的是yolo獨有的txt形式，因此在訓練前須先將用到的label轉成yolo的標籤格式。本研究使用python打開BDD100K label的json檔，判斷其中yolo需要的資料，再寫入到txt檔中。其中BDD10

0K紀錄bounding box的方式和yolo不同，需先將bounding box的數值轉換成yolo的形式才可寫入。

<2.> 準備train.txt、val.txt、test.txt：

yolov5在讀取影像時使用的是存放影像路徑的txt檔，本研究使用python進到label的json檔中，篩選出符合條件的影像名稱，再將影像名稱和路徑寫入txt檔中。

<3.> 網路結構配置：

本研究使用yolov5s作為模型，訓練之前需先修改yolov5s.yaml，將其中預設的類別參數改成3，分別是car、traffic light、traffic sign。

(5.) 訓練過程：

本研究訓練的每個yolov5模型初始超參數皆使用yolov5在coco資料集上預訓練的超參數，並訓練300 epoch。每個模型的結果都與圖 14十分相近。因為使用預訓練權重，且模型架構是參數最少的yolov5s，loss值的收斂十分快速且mAP也能快速升高，在實際應用上能加速模型的落地。

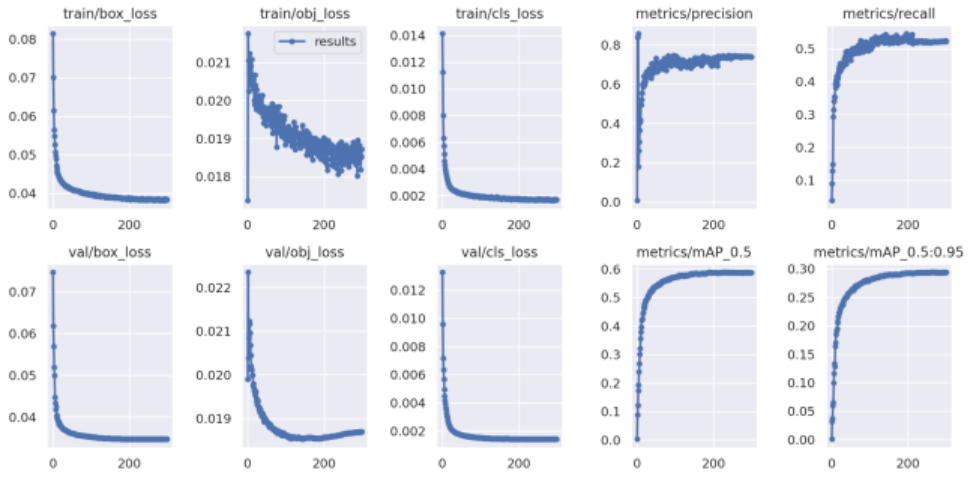
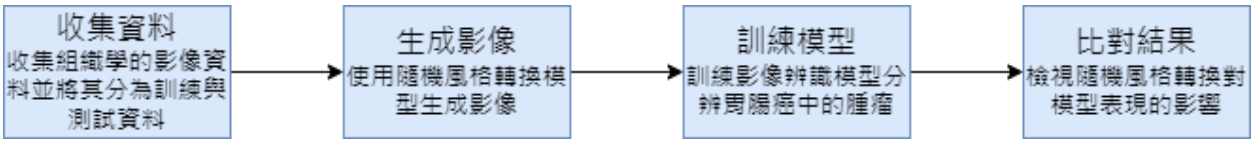


圖 14：sunny rainy supervise模型的損失函數和mAP趨勢

三、 組織學影像資料增強：

(一)、研究架構及流程：



1. 資料來源：

(1) NCT-CRC-HE-100K資料集：

- 這是一個由 100,000 張非重複影像，分別來自人類結腸直腸癌 (CRC) 和正常組織的蘇木精和伊紅 (H&E) 染色組織學圖像資料集。
- 組織類別為：脂肪 (ADI)、背景 (BACK)、碎片 (DEB)、淋巴細胞 (LYM)、粘液 (MUC)、平滑肌 (MUS)、正常結腸粘膜 (NORM)、癌症相關基質 (STR)、結直腸腺癌上皮（TUM）。
- 這些圖像是從 NCT Biobank 的福爾馬林固定石蠟包埋 (FFPE) 和 UMM 的病理學檔案樣本中手動提取的 N=86 H&E 染色的人類癌症組織切片 。組織樣本包含 CRC 原發性腫瘤切片和來自 CRC 肝轉移灶的腫瘤組織。

(2) 重新分類資料:

最後從100,000張影像中挑選出11977張，並將原本的9類減少為3類：

- ADIMUC - 脂肪組織和粘液，即鬆散的非腫瘤組織
- STRMUS - 基質和肌肉，即緻密的非腫瘤組織
- TUMSTU - 大腸癌上皮組織和胃癌上皮組織，即腫瘤組織

此為初步的將原始資料集簡化，後面在訓練分類模型時會將資料進一步做刪減和分成訓練集與測試集。

2. 生成影像：

(1.) 隨機非組織風格轉換：

此方法將組織影像圖轉換成藝術風格的影像，目的為探討將影像轉成與組織影像無關的是否會影響模型的表現。此風格轉換模型的風格嵌

入是由多元常態分佈中隨機取樣，而此多元常態分佈的平均值與共變異數與Painter By Numbers (PBN)資料集的風格嵌入的分佈一致，因此從此分佈中採樣也就等同於從PBN資料集中隨機選一張圖並計算其風格嵌入，只是使用這種方式能夠大幅降低運算量，並且每次生成新圖也不需要從整個資料集中採樣，只需從已經計算好的分佈中隨機取樣即可。

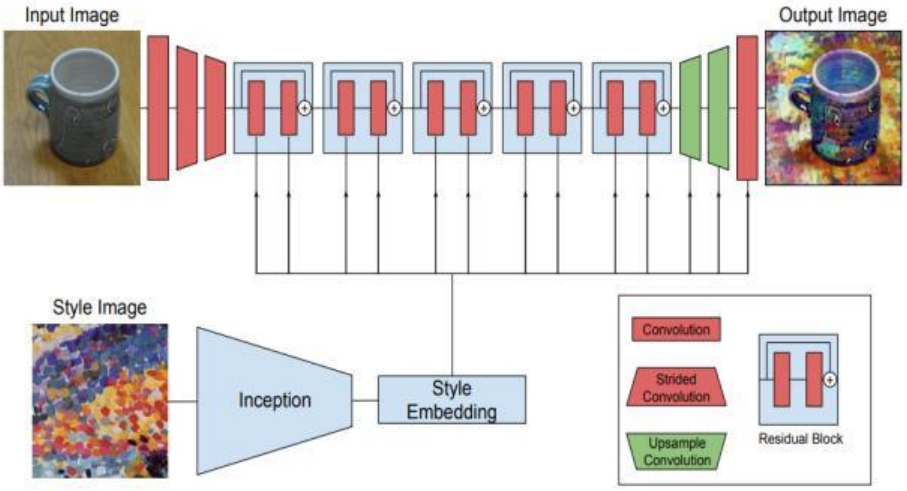
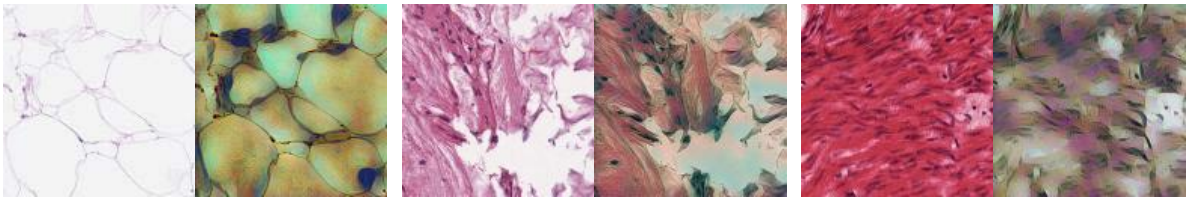
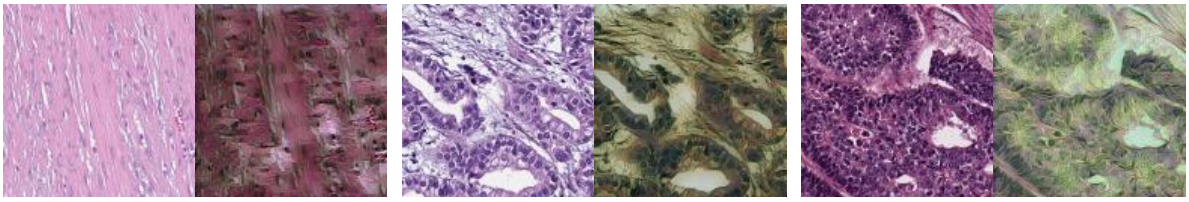


圖15： 隨機非組織風格轉換之架構圖

隨機非組織風格轉換結果(左：原圖，右：轉換結果)



脂肪組織(ADI)　　黏液(MUC)　　肌肉(MUS)

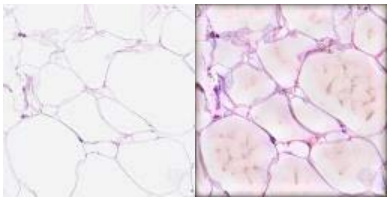


基質(STR)　　胃癌上皮組織(STU)　　大腸癌上皮組織(TMU)

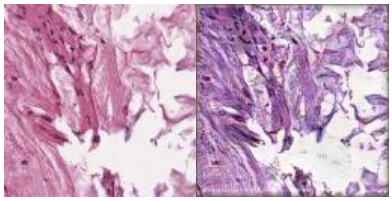
(2.) 隨機組織風格轉換：

此方法將組織影像圖轉換成資料集中其他組織影像圖的風格，目的為探討此方法是否能增加訓練資料的多樣性，使模型不會在特定的染色方式上出現過擬合的現象。使用與TSIT相同的方式，通過輸入一張內容影像及一張風格影像(隨機從資料集中選取的其他組織影像做為風格)，生成出一張保有內容影像內容但風格為風格影像的影像。

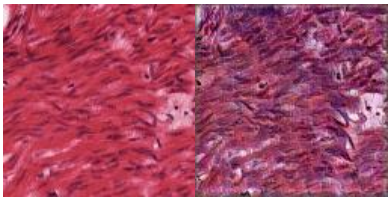
隨機組織風格轉換結果(左：原圖，右：轉換結果)



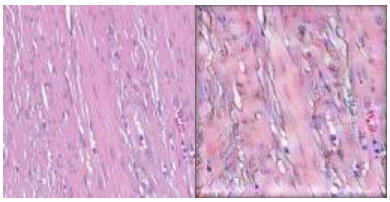
脂肪組織(ADI)



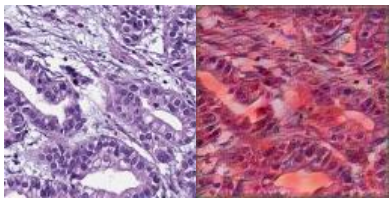
黏液(MUC)



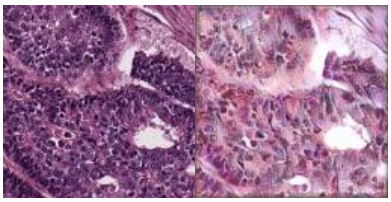
肌肉(MUS)



基質(STR)



胃癌上皮組織(STU)



大腸癌上皮組織(TMU)

3. 訓練腫瘤組織辨識模型：

(1.) VGG16：

本研究使用在ImageNet上預訓練好的VGG16模型去做遷移學習。VGG 16 是一種簡單且廣泛使用的捲積神經網絡(CNN) 架構。 VGG16 架構由牛津大學的 Karen Simonyan 和 Andrew Zisserman 於 2014 年提出。16 表示這個架構有16層隱藏層。本研究將其用於建立一個能正確分類腫瘤組織與非腫瘤組織的影像分類模型。

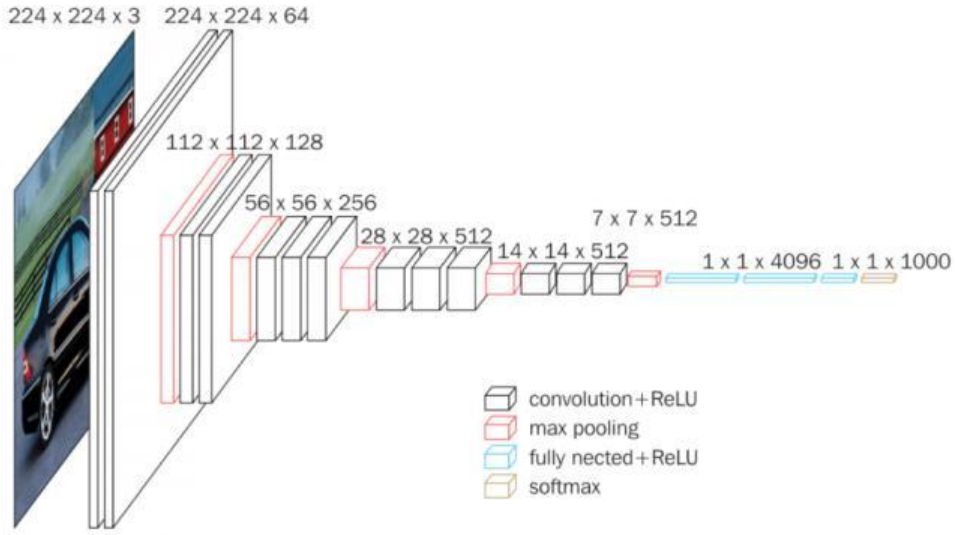


圖16：VGG16架構

(2.) 模型設計：

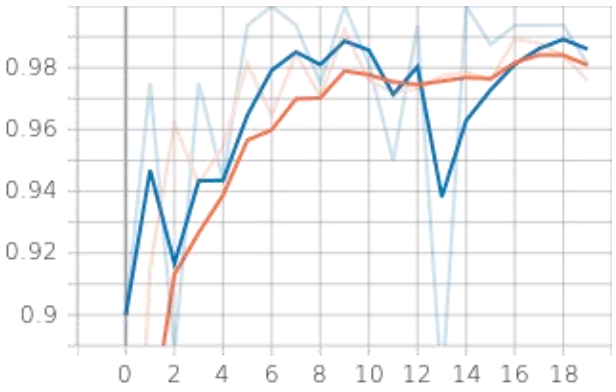
本研究共訓練了兩種配置(影像張數)的影像辨識模型，一種只使用真實影像訓練，而另一種加入350張風格轉換生成的影像。共三個模型分為：全部使用真實影像、加入非組織風格轉換、加入組織風格轉換

類別	ADI	MUC	STR	MUS	TUM	STU
訓練資料	245+245	245+245	245+245	245+245	245+245	245+245
驗證資料	35+35	35+35	35+35	35+35	35+35	35+35
測試資料	70+70	70+70	70+70	70+70	70+70	70+70

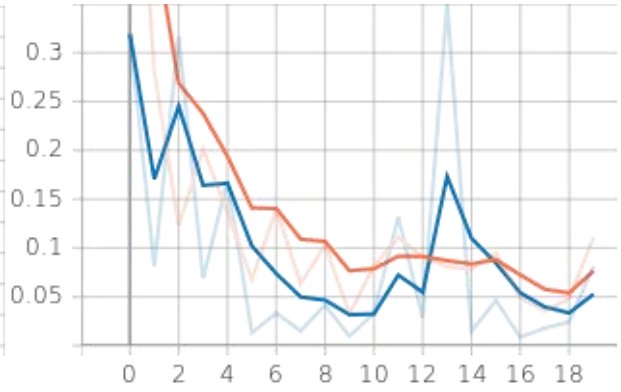
表5：影像張數配置(紅色數字為加入的增強影像數量)

(3.) 訓練結果：

由於避免篇幅過長，且其他兩個模型皆為相近的趨勢，本部分以使用真實影像做訓練資料的模型作為代表。



訓練模型的準確值 train val



訓練模型的損失值 train val

伍、 研究結果與討論

一、街景圖資料增強：

實驗數據解釋：

以下表格中的數值均為mAP。mAP是mean of Average Precision的縮寫，即平均精確度（average precision）的平均（mean），是衡量物件辨識模型的標準，數值越高表示模型表現越好。

$$Precision = \frac{TP}{TP + FP}$$

TruePositive(TP)：是目標類別且模型判斷為目標類別

FalsePositive(FP)：不是目標類別且模型判斷為目標類別

(一) yolov5模型的實驗結果：

1. 訓練資料使用晴天與夜晚：

(1.) 在晴天測試資料上的表現：

Class	art	base50	base200	transform 50	transform200	supervise
all	0.597	0.565	0.567	0.61	0.608	0.576
car	0.718	0.704	0.704	0.722	0.725	0.705
traffic sign	0.543	0.492	0.495	0.57	0.562	0.505

traffic light	0.532	0.498	0.5	0.538	0.539	0.517

表6：實驗結果mAP數值表格

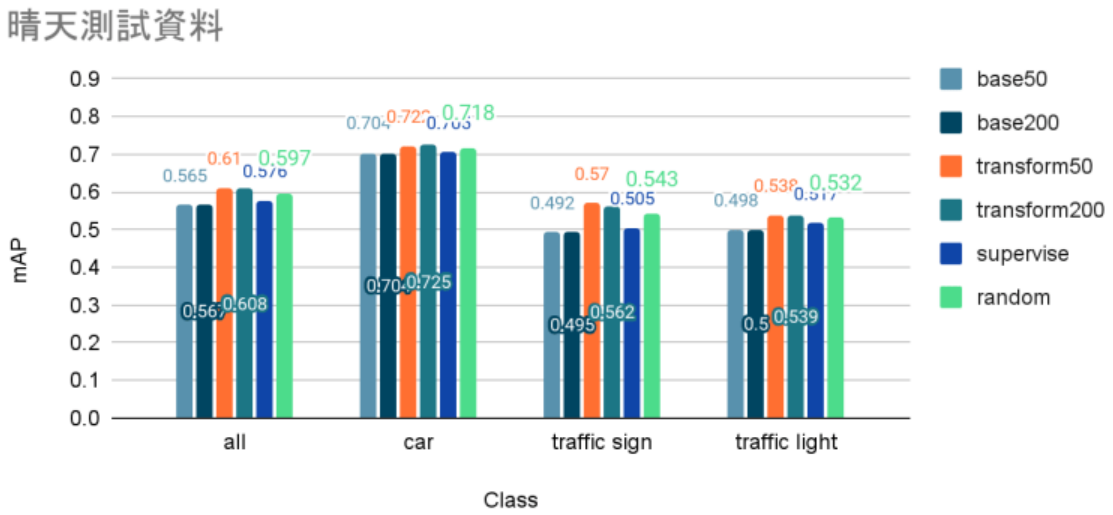


圖 17：實驗結果數值長條圖

(2.) 在夜晚測試資料上的表現：

Class	art	base50	base200	transform50	transform200	supervise
all	0.45	0.454	0.495	0.442	0.456	0.581
car	0.605	0.612	0.645	0.626	0.626	0.704
traffic sign	0.418	0.413	0.457	0.425	0.428	0.544
traffic light	0.327	0.338	0.385	0.275	0.314	0.496

表 7：實驗結果mAP數值表格

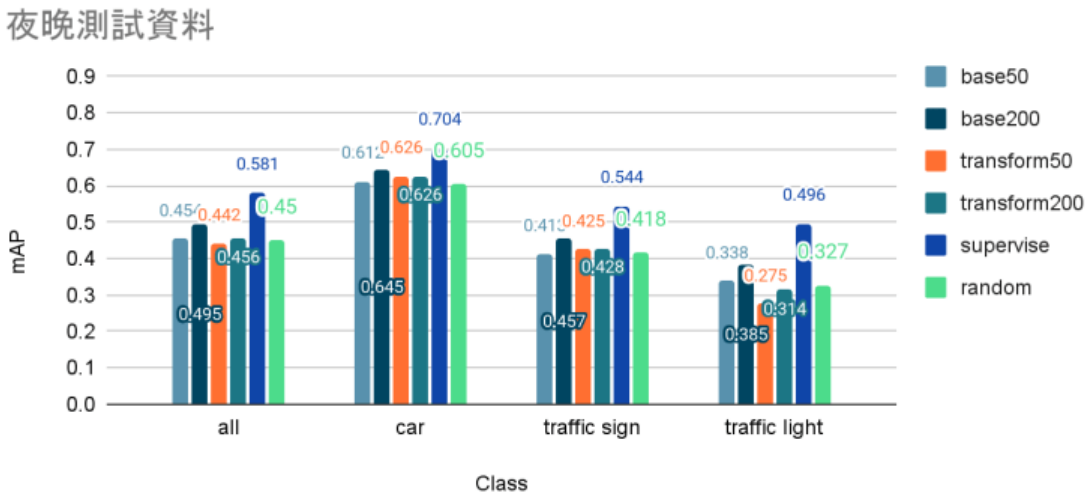


圖 18：實驗結果mAP數值長條圖

2. 訓練資料使用晴天與雨天：

(1.) 在晴天測試資料上的表現：

Class	art	base50	base200	transform 50	transform200	supervise
all	0.597	0.572	0.571	0.642	0.637	0.596
car	0.718	0.705	0.705	0.744	0.742	0.717
traffic sig n	0.543	0.493	0.493	0.610	0.609	0.529
traffic lig ht	0.532	0.516	0.515	0.573	0.56	0.543

表 8：實驗結果mAP數值表格

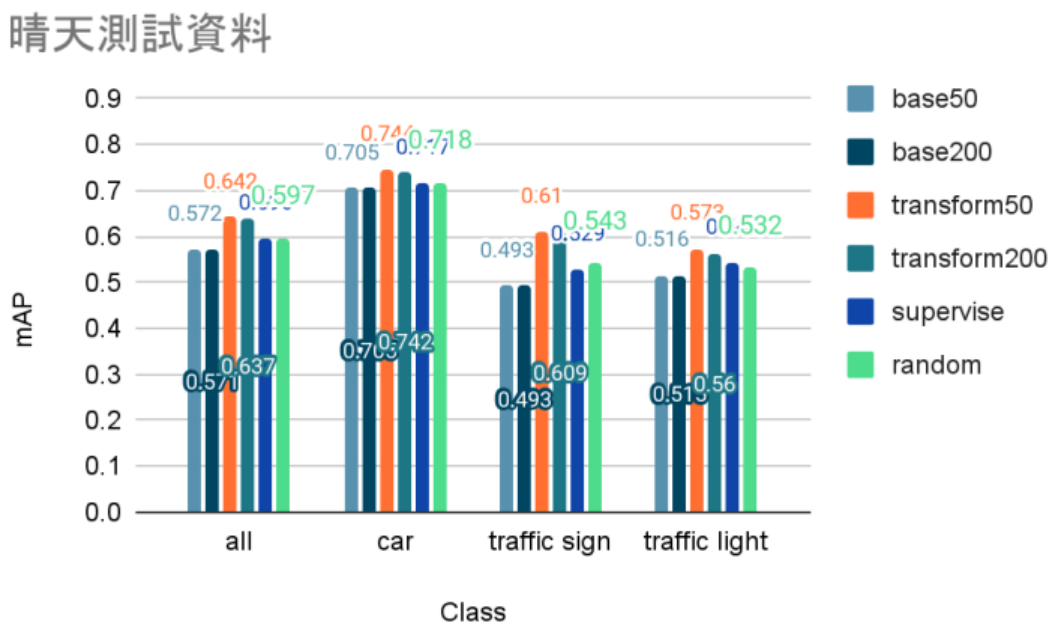


圖 19：實驗結果mAP數值長條圖

(2.) 在雨天測試資料上的表現：

Class	art	base50	base200	transform 50	transform200	supervise
all	0.526	0.522	0.523	0.528	0.533	0.585
car	0.708	0.711	0.718	0.714	0.721	0.745
traffic sig n	0.456	0.457	0.453	0.466	0.469	0.539
traffic lig ht	0.413	0.397	0.399	0.404	0.408	0.471

表 9：實驗結果mAP數值表格

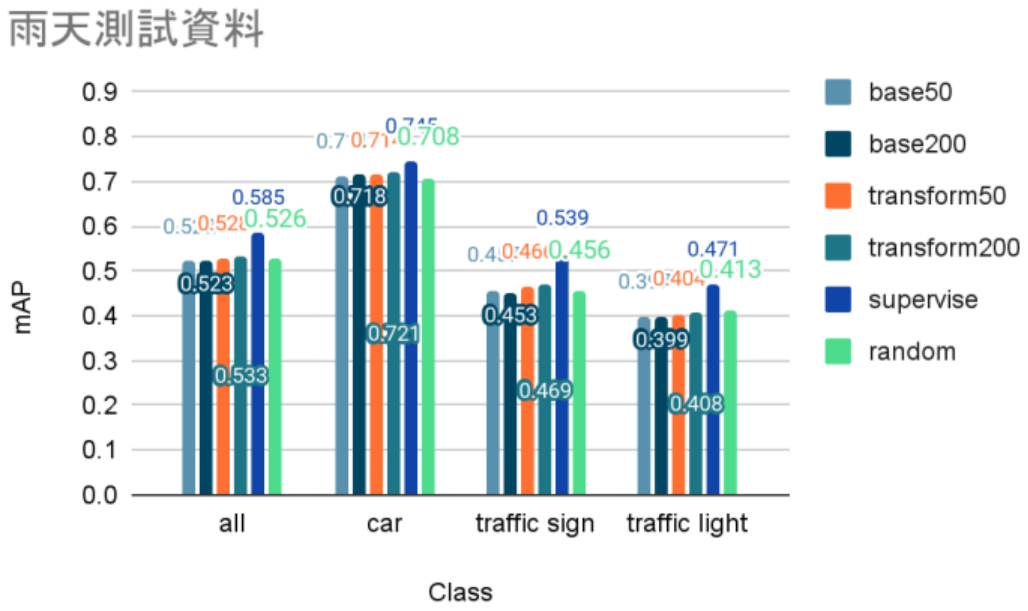


圖 20：實驗結果mAP數值長條圖

(二) 探討使用風格轉換影像訓練對yolov5模型map的影響：

從圖、表 5和圖、表 7中可以得知，在晴天的測試資料上，使用本研究提出之方法的模型均取得最高的準確度。而從圖、表 6和圖、表 8中可以得知在目標風格的測試資料上，使用真實資料訓練的supervise準確度最高，而使用本研究提出之方法的模型次之，準確度最低的模型為模擬現實生活中資料不平均狀況之模型。由實驗結果可知，本研究提出之資料增強方法能改善資料風格不平均時的模型準確度。

(三) 探討style image數量對yolov5模型map的影響：

從圖、表 6和圖、表 8中可以得知，使用200張style image的模型在目標天氣的測試資料上的表現均好於使用50張style image的模型，因此可以推測多樣化的style image對於生成訓練資料有幫助。

(四) 探討真實資料數量對yolov5模型map的影響：

根據實驗結果，將真實目標天氣的訓練資料從50張增加到200張，在目標天氣的測試資料上就能有顯著的進步，且使用最多真實影像的supervise準確度又比使用200張的模型高，由此可知增加真實資料能大幅提升模型的表現。

(五) 探討夜晚與雨天的差異：

在探討(二)的時候發現，使用本研究提出之方法的模型在雨天的測試資料上表現較在夜晚的測試資料上佳，於是進一步地做了實驗去探討不同模型在**使用風格轉換生成的影像**上的表現，想要藉此找出雨天生成影像跟夜晚生成影像的不同之處。

1. 使用**TSIT生成的夜晚影像**作為測試資料：

Class	transform200	supervise
all	0.496	0.398
car	0.662	0.59
traffic sign	0.424	0.317
traffic light	0.403	0.287

表 10：實驗結果mAP數值表格

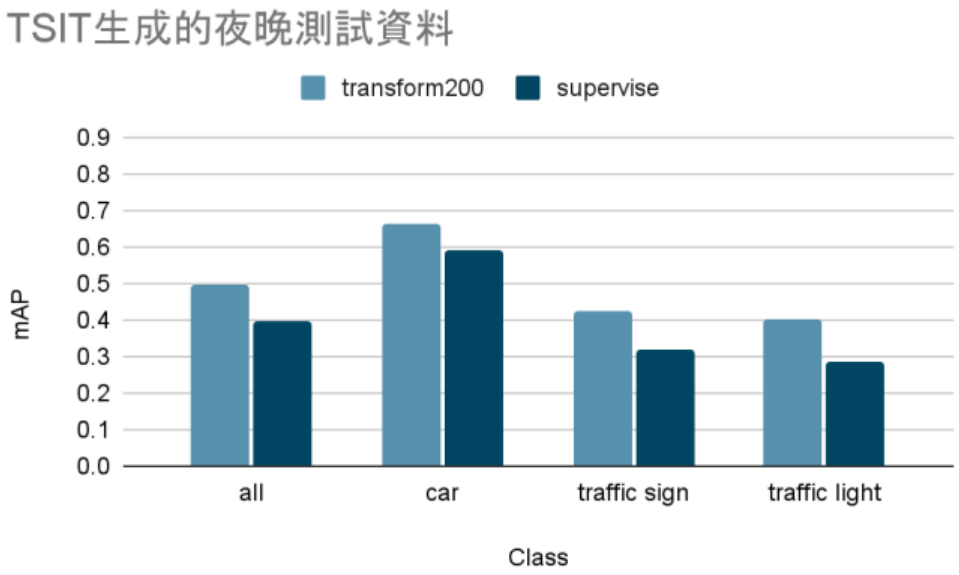


圖 21：實驗結果mAP數值長條圖

2. 使用**TSIT生成的雨天影像**作為測試資料：

Class	transform200	supervise
all	0.59	0.558
car	0.71	0.685
traffic sign	0.527	0.495
traffic light	0.535	0.496

表 11：實驗結果mAP數值表格

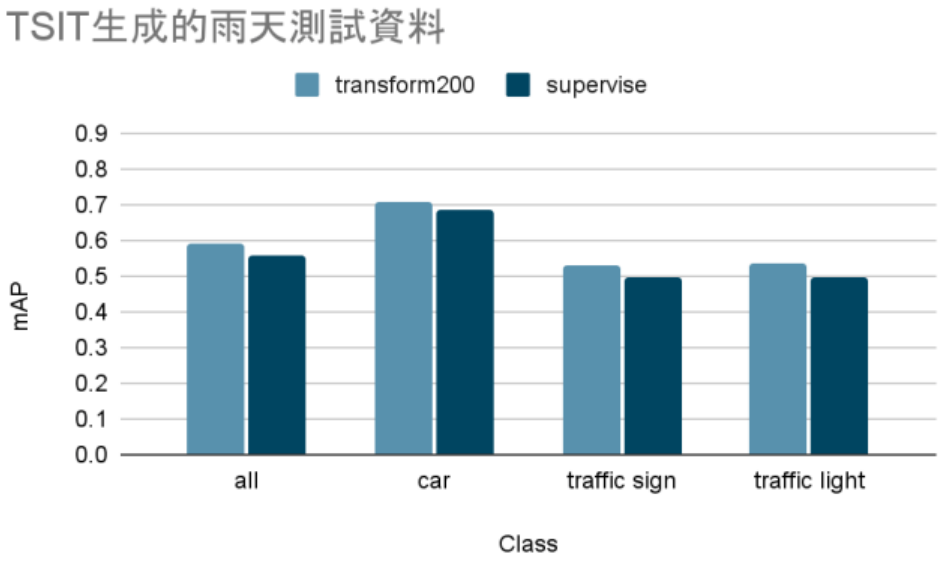


圖 22：實驗結果mAP數值長條圖

根據實驗結果可以發現，supervise和transform200在雨天的生成資料上的表現差距較小，supervise和transform200在夜晚的生成資料上的表現差距較大。由於supervise完全是由真實影像作訓練，因此supervise的表現越好表示生成的影像越接近真實影像。而造成雨天的轉換好於夜晚的轉換的原因，推測是由於晴天的街景圖與雨天的街景圖在風格上較為接近，在轉換的過程中較不會丟失content訊息。

(六) 探討style image數量對TSIT模型的影響：

在訓練TSIT模型的時候發現，只使用50張style image的generator loss會慢慢升高，無法收斂，而使用200張style image的generator loss會逐漸下降，能成功收斂。

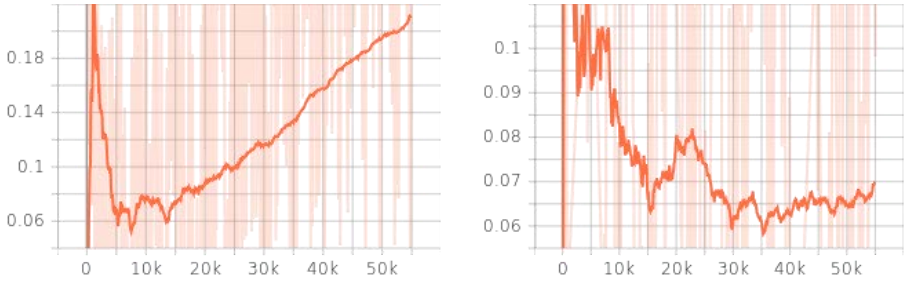


圖 23：TSIT50(左)和TSIT200(右)的generator loss

(七) 探討TSIT模型生成的影像與實際影像的FID分數：

Fréchet 初始距離 (FID) 是一種用於評估由生成模型（如生成對抗網絡 (GAN)）創建的圖像質量的指標。

1. TSIT生成的夜晚影像與真實夜晚影像的距離較真實晴天影像接近，因此可知yolo v5模型在夜晚表現較差的原因並不是影像沒有成功轉換成夜晚的風格

2. TSIT生成的雨天影像與真實晴天影像的距離較真實雨天影像接近。與討論(五)的結果不符。最終推測是因為晴天和雨天在風格上本來就較為相近，所以雖然FID分數的結果顯示生成出的雨天與真實的晴天較相似(也就是與真實的雨天較不相似)，還是能滿足在討論(五)中生成的雨天與訓練資料較相似的推論(因為晴天跟雨天本來就很相似)。

	50	200
Transform night True night	84.167	74.720
Transform night True sunny	92.194	97.669
Transform rainy True rainy	77.747	79.380

Transform rainy True sunny	29.675	26.801

表 12：實驗結果FID數值表格

二、組織圖資料增強

(一)探討使用風格轉換資料增強對模型準確率的影響：

模型	真實資料	非組織風格轉換	組織風格轉換
準確率	0.8384	0.8556	0.9031
損失值	0.6540	0.5930	0.4331

表13：使用不同訓練資料得出的準確率

由上表可以得知，不管是使用非組織還是組織風格轉換，都能有效地提升模型辨識的準確率。

(二) 探討training和validation的準確率與testing的落差：

在實驗過程中我發現，三個模型均在驗證集(validation set)上有很高的準確率，然而在testing set上的準確率就有所下降，從此結果可以推論不同的染色結果對於模型的準確率有一定程度的影響(因為testing set中的染色方式與模型在training和validation中學到的有所不同)。

陸、 結論與應用

一、街景圖資料增強：

結論：

1. 此資料增強方法能提升模型在晴天、雨天的表現
2. 目前此資料增強方法尚不適用於夜晚
3. 本研究使用的Style image的數量對於yolov5模型的表現影響不明顯
4. 對於yolov5來說晴天與雨天本來在風格上就較為相似，因此就算FID score算出來生成的雨天與晴天較接近，yolov5依舊能有較好的表現。

5. 從FID score可以得知，transform模型在夜晚表現差的原因不是影像風格轉換失敗

二、組織學影像資料增強：

結論：

1. 使用非組織和組織風格轉換皆能提升模型準確率
2. 組織影像風格轉換的提升效果較非組織影像風格轉換明顯

三、 現實應用：

對於現實生活中許多需要在戶外運作的模型來說，訓練資料的風格是影響模型表現的重要因素，例如季節、時間、光照強度、地點等等，然而收集一個完整且多元的資料集需要花費大量的時間和人力，如果需要快速地讓模型開始運作，就可透過本研究提出之方法做資料增強，後續再不斷更新真實資料到資料集中，這樣在模型運作前期比起原本直接用未增強過的資料訓練出來的模型能有更高的準確率。

可能運用的領域：

自動駕駛、農作物辨識、行人辨識、空照圖的物件辨識、戶外野生動植物辨識、海洋生態辨識

四、未來展望：

1. 短期目標：
   (1.) 選擇一個現實生活中因為訓練資料風格缺乏，導致在特定場景下準確率不佳的問題，並用本研究提出之方法改善。
   (2.) 修改TSIT模型架構使其生成出來的影像符合yolov5的backbone
   (3.) 尋找生成資料與真實資料在訓練時的最佳比例，或是挑選代表性的style image
2. 長期目標：
   (1.) 使用不同的影像風格轉換模型
   (2.) 優化此資料增強方法的易用性，如整合成package供他人使用

柒、 參考文獻及附錄

[1.] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, Trevor Darrell(2018). BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. arXiv:1805.04687 [cs.CV]

[2.] Liming Jiang, Changxu Zhang, Mingyang Huang, Chunxiao Liu, Jianping Shi, Chen Change Loy(2020). TSIT: A Simple and Versatile Framework for Image-to-Image Translation. arXiv:2007.12072 [cs.CV]

[3.] Glenn Jocher; Alex Stoken; Ayush Chaurasia; Jirka Borovec; NanoCode012; TaoXie; Yonghye Kwon; Kalen Michael; Liu Changyu; Jiacong Fang; Abhiram V; Laughing; tkianai; yxNONG; Piotr Skalski; Adam Hogan; Jebastin Nadar; imyhxy; Lorenzo Mammana; AlexWang1900;Cristi Fati; Diego Montes; Jan Hajek; Laurentiu Diaconu; Mai Thanh Minh; Marc; albinxavi; fatih; oleg;
wanghaoyang0106(2020). yolov5. https://github.com/ultralytics/yolov5

[4.] Philip T. Jackson, Amir Atapour-Abarghouei, Stephen Bonner, Toby Breckon, Boguslaw Obara(2018). Style Augmentation: Data Augmentation via Style Randomization. arXiv:1809.05375 [cs.CV]

[5.] Ahmad, Jamil; Muhammad, Khan; Baik, Sung Wook (2017): Data augmentation using semantic-preserving transformation for SBIR.. PLOS ONE. Figure. https://doi.org/10.1371/journal.pone.0183838.g002

[6.] 影像風格轉換示意圖：https://mitmedialab.github.io/ai-creativity-workshop/curriculum.html

[7.] yolov5架構：https://zhuanlan.zhihu.com/p/172121380

[8.] 生成對抗網路模型示意圖： https://www.kdnuggets.com/2017/01/generative-adversarial-networks-hot-topic-machine-learning.html

[9.] 物件辨識示意圖：https://bitmovin.com/object-detection/

[10.] Rikiya Yamashita, Jin Long, Snikitha Banda, Jeanne Shen, Daniel L. Rubin(2021): Learning domain-agnostic visual representation for computational pathology using medically-irrelevant style transfer augmentation
arXiv:2102.01678

【評語】052502

1. 本作品實作影響風格轉換模型作為資料增強的方法並應用於兩個實驗：(1)不同天氣時間的街景圖 and (2) 胃腸癌腫瘤檢測模型。

2. 作品有完整實驗並探討風格轉換生成的訓練影像對於準確率的影響。

3. 作品說明書大多篇幅都在介紹所使用到的模型，建議可多一些跟主題有關的討論，包括在所選擇的兩個應用情境，風格轉換作為資料增強是否會比其他資料增強方法更好？既然有兩個特定的應用，建議可針對這兩個應用比較不同資料增強方法的結果。

作品簡報

基於影像風格轉換之資料增強

Data Augmentation Based on Style Transfer

Introduction

- 訓練電腦視覺的模型需要花費大量的人力與時間收集與標籤資料
- 許多模型被訓練在有bias的資料集上，導致模型表現不穩定

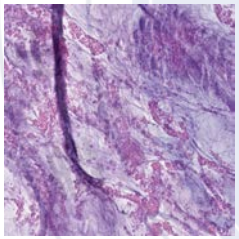
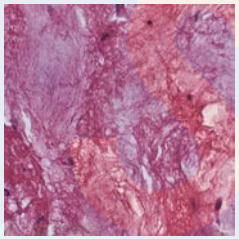
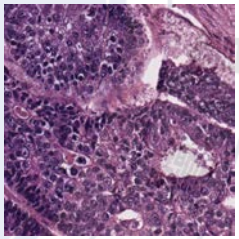
問題一：自駕車在不同天氣時間下的辨識率不穩定

由於當初收集資料時多數訓練影像為晴天，導致模型在雨天與夜晚等其他情況的準確率下降

不同訓練資料在不同場景下的準確度 (mAP)

訓練資料場景	晴天	晴天+夜晚	晴天	晴天+雨天
mAP	0.454	0.581	0.522	0.585
	測試場景：夜晚		測試場景：雨天	

問題二：醫學影像模型會受到不同醫院染色組織圖的影響

目前多數醫學影像模型皆訓練在單一醫院染色的資料上，較少多樣化的資料

研究目標：

利用影像風格轉換做資料增強，解決訓練資料bias的問題

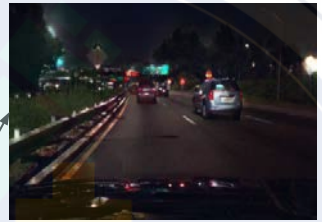
利用影像風格轉換增加資料多樣性 - 自駕車問題

晴天影像

其他風格

影像風格轉換模型

生成新資料



舊資料

新資料與舊資料一同餵入模型訓練

物件辨識模型

風格轉換方式:

1. Random style embedding
2. TSIT(指定風格)

物件辨識模型:

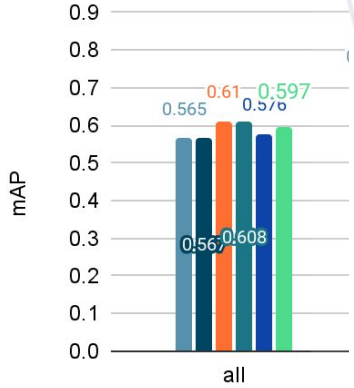
yolov5

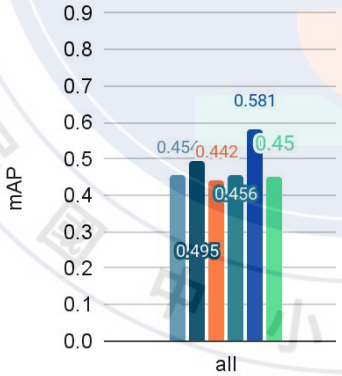
實驗一 - 探討風格轉換影像對yolov5模型map的影響

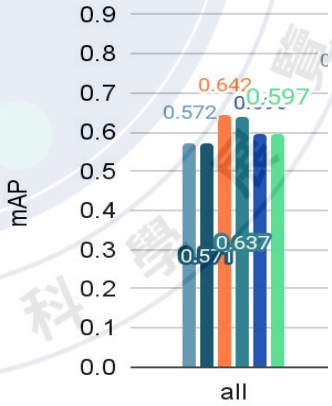
共訓練11個模型

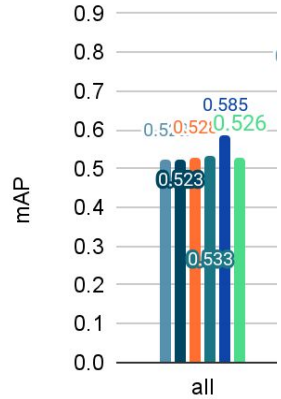
- base50
- base200
- transform50
- transform200
- supervise
- random

	art(random)	base50	base200	tran50	tran200	supervise
晴天	0.597	0.565	0.567	0.61	0.608	0.576
夜晚	0.45	0.454	0.495	0.442	0.456	0.581
晴天	0.597	0.572	0.571	0.642	0.637	0.596
雨天	0.526	0.522	0.523	0.528	0.533	0.585

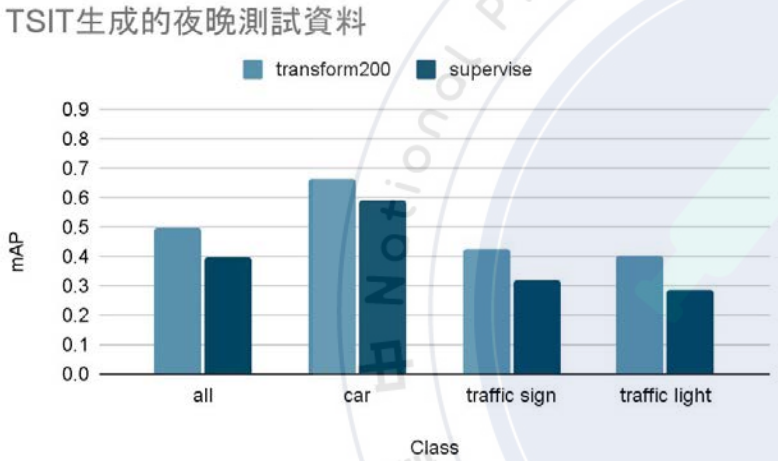




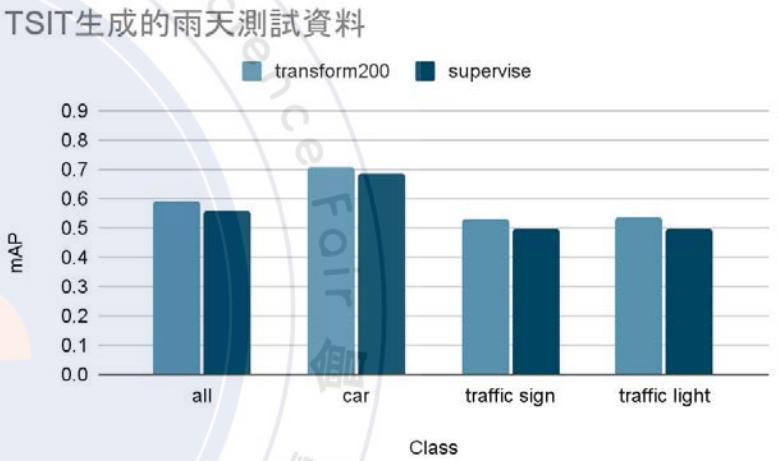




實驗二 - 探討夜晚與雨天的差異



Supervise 與 transform mAP 差距較大



Supervise 與 transform mAP 差距較小

- 由於supervise完全是由真實影像作訓練，因此推測supervise的表現越好表示生成的影像越接近真實影像。
- **推論：模型生成的雨天影像較接近真實雨天影像，夜晚則相反**

實驗三 - 探討TSIT模型生成的影像與實際影像的FID分數

	50	200
Transform night True night	84.167	74.720
Transform night True sunny	92.194	97.669

低

高

	50	200
Transform rainy True rainy	77.747	79.380
Transform rainy True sunny	29.675	26.801

高

低

$$FID = \|\mu_r - \mu_g\|^2 + T_r(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2})$$

FID 是一種評估兩個不同 domain的影像相近程度的指標。數 值越小表示兩個domain越相近

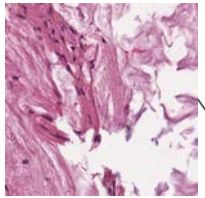
由FID可知:

- 晴天有成功轉成夜晚
- 晴天沒有成功轉成雨天，還是與晴天較相似 (與實驗三的推論不相符)

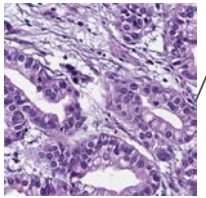
最終推論:晴天和雨天在風格上本來就十分相近，所以 FID score認為晴天沒有成功轉成雨天，而實驗三的推論同時也能成立 (因為本來兩個就很像，所以 supervise跟transform差距會比夜晚小。

利用影像風格轉換增加資料多樣性 - 醫學影像問題

原組織影像

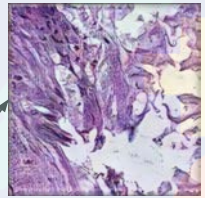


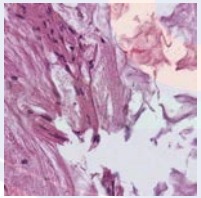
組織風格



影像風格轉換模型

生成新資料





舊資料

新資料與舊資料一同餵入模型訓練

腫瘤辨識模型

風格轉換方式:

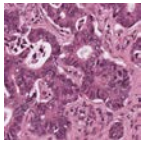
1. Random style embedding
2. adain(指定風格)

腫瘤辨識模型:

vgg16

實驗一 - 探討使用風格轉換資料增強對模型準確率的影響

1. 非組織風格轉換:隨機選取一種藝術風格作為影像風格轉換的風格輸入
2. 組織風格轉換:隨機從醫學影像資料集中選取一張影像



模型	真實資料	非組織風格轉換	組織風格轉換
準確率	0.8384	0.8556	0.9031
損失值	0.6540	0.5930	0.4331

不管是使用非組織還是組織風格轉換，都能有效地提升模型辨識的準確率。

探討training和validation的準確率與testing的落差:

實驗中發現三個模型均在驗證集 (validation set)上有很高的準確率，然而在 testing set 上的準確率就有所下降，從此結果可以推論不同的染色結果對於模型的準確率有一定程度的影響(因為testing set中的染色方式與模型在 training和validation中學到的有所不同)。

結論

自駕車問題

1. 使用風格傳換資料增強能提升模型在晴天、雨天的表現，但尚不適用於夜晚
2. 本研究使用的Style image的數量對於yolov5模型的表現影響不明顯
3. 對於yolov5來說晴天與雨天本來在風格上就較為相似，因此就算FID score算出來生成的雨天與晴天較接近，yolov5依舊能有較好的表現。
4. 從FID score可以得知，transform模型在夜晚表現差的原因不是影像風格轉換失敗
5. 使用隨機藝術風格轉換表現好於不使用資料增強，但差於指定風格轉換

醫學影像問題

1. 使用非組織和組織風格轉換皆能提升模型準確率
2. 組織影像風格轉換的提升效果較非組織影像風格轉換明顯

腫瘤辨識率

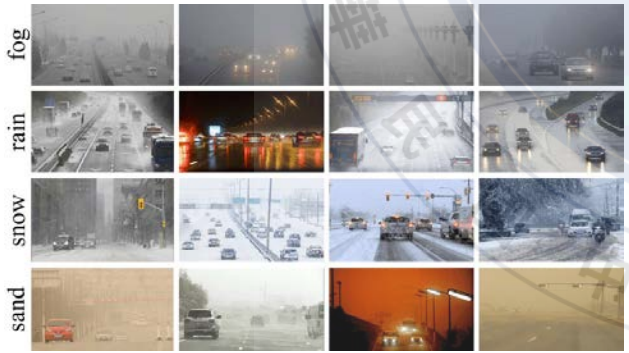
現實應用

應用場景:

1. 辨識物體受到季節、時間、光照強度、地點等影響，需要多樣風格但資料不足的
2. 希望能讓模型快速落地，可先用此方法增加訓練資料，之後再慢慢補全真實訓練資料

可能運用的領域:

自動駕駛、農作物辨識、行人辨識、空照圖的物件辨識、戶外野生動植物辨識、海洋生態辨識



未來展望

短期目標:

1. **將風格轉換與辨識的模型整合在一起訓練，讓兩者能更好的結合**
2. 選擇一個現實生活中因為訓練資料風格缺乏，導致在特定場景下準確率不佳的問題，並用本研究提出之方法改善。
3. 尋找生成資料與真實資料在訓練時的最佳比例，或是挑選代表性的style image

長期目標:

1. 使用不同的影像風格轉換模型
2. 優化此資料增強方法的易用性，如整合成package供他人使用

References

[1.] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, Trevor Darrell(2018). BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. arXiv:1805.04687 [cs.CV]

[2.] Liming Jiang, Changxu Zhang, Mingyang Huang, Chunxiao Liu, Jianping Shi, Chen Change Loy(2020). TSIT: A Simple and Versatile Framework for Image-to-Image Translation. arXiv:2007.12072 [cs.CV]

[3.] Glenn Jocher; Alex Stoken; Ayush Chaurasia; Jirka Borovec; NanoCode012; TaoXie; Yonghye Kwon; Kalen Michael; Liu Changyu; Jiacong Fang; Abhiram V; Laughing; tkianai; yxNONG; Piotr Skalski; Adam Hogan; Jebastin Nadar; imyhxy; Lorenzo Mammana; AlexWang1900;Cristi Fati; Diego Montes; Jan Hajek; Laurentiu Diaconu; Mai Thanh Minh; Marc; albinxavi; fatih; oleg;
wanghaoyang0106(2020). yolov5. https://github.com/ultralytics/yolov5

[4.] Philip T. Jackson, Amir Atapour-Abarghouei, Stephen Bonner, Toby Breckon, Boguslaw Obara(2018). Style Augmentation: Data Augmentation via Style Randomization. arXiv:1809.05375 [cs.CV]

[5.] Ahmad, Jamil; Muhammad, Khan; Baik, Sung Wook (2017): Data augmentation using semantic-preserving transformation for SBIR.. PLOS ONE. Figure. https://doi.org/10.1371/journal.pone.0183838.g002