

# Final Project

Team 4

組員：108062313 黃允暘、108062311 杜方辰、108062122 顏訓哲 108062336 陳柏元

## Classifier：

### 1. Methodology

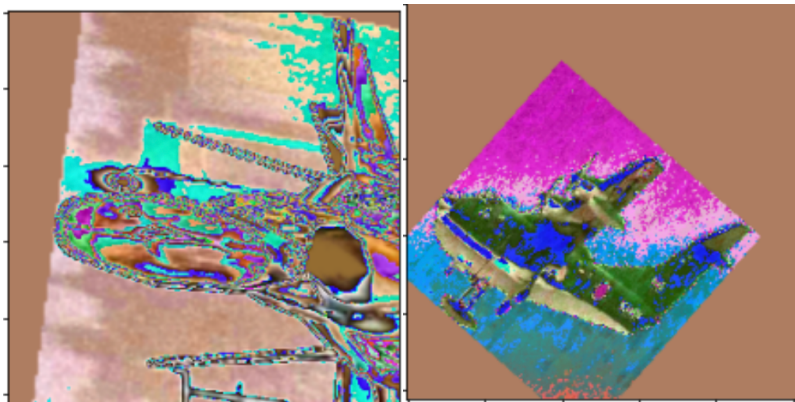
關於 model，我們使用的是 Efficientnet-b6，因為他是目前 image classify score 最高的分數，並且 model size 也算是中小型，訓練速度較其他例如：Resnet 快速許多。

### 2. Data analysis:

- 計算所有 data 的 mean, std 用來 normalization.
- 計算每個種類的數量，才能平均分在 training set, validation set, testing set.

### 3. How to train out model

- Optimizer 使用 Adam，learning rate 設為 0.00001。loss function 使用 Categorical Cross Entropy。Metric 則使用 accuracy。
- 我們將 dataset 中的 0.2 切割出來作為 validation set，0.1 作為 testing set。
- 接著我們嘗試加上許多 heavy 的 augmentation，因為資料不多，所以需要 augmentation 來增加多樣性。(ColorJitter, GaussianBlur, rotate, RandomAffine, RandomInvert, RandomAdjustSharpness, RandomCrop 等等)



- 此外，我們也有動態的調整 learning rate。當 validation set 的 ACC 連續 60 個 epoch 沒有升高，我們就把 learning rate 以 factor = 0.5 調低，並且儲存最好的 model。
- Total epoch: 100 epochs

### 4. Test result

Validation set: 0.95 accuracy

Test set: 0.93 accuracy

### 5. Demo result

Accuracy 約為 0.9。

## 6. Discussion

我們有嘗試使用不同的 backbone architecture，像是是 EffiecientNet-B0、EffiecientNetV2-B2、DenseNet121、RegNet 等等，不過最後測出來，最好的結果為 Efficientnet-b6，所以我們最後選擇採用 Efficientnet-b6。

# Generator

## 1. Methodology

關於 model，我們主要參考了「Improved Training of Wasserstein GANs」這篇論文，其作者為 Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville。我們使用了 conditional GAN，搭配上 Improved WGAN 來實作這一次的 project，也取得不錯的成效。

## 2. Training model:

- **Condition :**

Basic 的部分，我們採用 conditional GAN，並搭配 improved WGAN 的演算法去實作，我們將 manufacturer 作為 condition 的條件輸入。因此可以藉由輸入不同的品種，generate 出不同的飛機。

- **Text encoder :**

因為 label 是個各廠牌的飛機名稱，是一個字串，因此我們採用 BERT 這個 pretrained model 來當作我們的 text encoder，用來取得 embedding，並當作 condition。

- **Generator :**

我們的架構如下：

1. 先將 text embedding flatten 並傳入一層 dense layer，然後和 noise concat 起來後再傳入一層 dense layer，normalize 並 reshape 後產生第一個 output x0。
2. 之後，將 x0 傳入三層 convolutional layer，中間穿插 batch normalization，產生第二個 output x1。
3. 接著，將 x0、x1 相加，經過一層 convolutional transpose layer 並 normalization，產生第三個 output x2。
4. 把 x2 傳入三層 convolutional layer，中間穿插 batch normalization，產生第四個 output x。
5. 把 x、x2 相加，經過兩層 convolutional transpose layer 並穿插 normalization，產生最終的 logits。
6. 最後用 tanh 產生最終 output。

此外，在上述架構中，有用到 activation function 的部份我們全部是採用 relu，並將 relu 的 alpha 值設為 0.2。

- **Discriminator :**

我們的架構如下：

1. 先單獨處理圖片：
  - (a). 把圖片經過一個 convolutional layer 後，再經過三次 convolutional layer 和 batch normalization 的穿插，得到第一個 output x0。
  - (b). 將 x0 再經過三次 convolutional layer 和 batch normalization 的穿插，得到

output  $x$ 。

(c). 把  $x$  和  $x_0$  相加，得到 output  $x_1$ 。

2. 加入 text:

(a). 將 text 經過預處理(tile、expand\_dim)，然後和  $x_1$  concat 起來。

(b). 經過一層 convolutional layer 和一層 normalization，產生最終的 logits。

3. 最後用 tanh 產生最終 output。

- **Loss function and optimization :**

loss 的計算方式，我們並不是單純將 condition 加入 GAN，而是有使用類似 improved WGAN 的方式來計算。和 improved WGAN 相同的地方是，我們有用到 gradient penalty 來限制 interpolate 的 gradient 不能大於 1。

以下是我們 loss 計算方式的特點:

1. 採用 improved WGAN 的策略，在 fake image 和 real image 之間以隨機比例取得 interpolate。

2. 除了(real\_image, text)、(fake\_image, text)的 pair 外，多加了第 3 個 pair: (interpolate, text)

3. 用第三個 pair 丟入 discriminator，產生的 logits 和 interpolate 計算 gradient 作為限制的 penalty。

4. 最後，和 improved WGAN 一樣，在 loss 的計算中加上 penalty。

5. optimizer 的部分使用 Adam，generator 部分的 learning rate 設為  $1e-4$ ，discriminator 部分的 learning rate 設為  $1e-4$ 。

### 3. Test result / Demo result



上圖是 basic+bonus part 的結果，將 manufacturer 設為 conditional control，我們訓練了 2000 個 epoch。將品牌的名字，如: "Airbus"，丟入 bert，將會得到一個 embedding，

再將這個 `embedding` 丟到 `generator` 中，即可以得到該品牌的飛機圖片。圖中的 8 個 row 分別為生成出來 8 個不同品牌的飛機。分別是 "Airbus"、"ATR"、"Antonov"、"British"、"Beechcraft"、"Lockheed"、"Douglas"。

#### 4. Discussion:

在最先開始的嘗試中，我們只有嘗試兩種 pair: (`real_image`, `text`)、(`fake_image`, `text`)，雖然產生出來的圖片本身是呈現得還不錯的，飛機的形狀也不會有模糊，不過文字與圖片對不起來。因此必須要加入第三個 pair，強迫模型學習 `text`，否則 `generator` 會和 `discriminator` 串通好兩個都不學 `text`。

#### 5. Conclusion:

在這次的 `final project` 中，我們成功的利用 `Car196 dataset` 以及應用現有的一些 `classifier architecture`，並參考近期的 `paper`，調整實作出擁有還算不錯的 `accuracy` 的車子廠牌 `classifier`。

在 `generator` 的部分，除了實作 `conditional GAN` 之外，我們也參考了 `improved WGAN`，大大的提升 `generator` 的效果。我們也達成 `bonus part`，根據廠牌生產出不同的飛機圖片！感謝老師及助教這學期的付出，著實讓學習到許多知識。