**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。

1. **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.

1. **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）

一張含有 文字 的圖片

自動產生的描述一張含有 文字, 信箋, 信封 的圖片

自動產生的描述

* 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

1. **Test result**

Validation set: 0.95 accuracy

Test set: 0.93 accuracy

1. **Demo result**

Accuracy約為0.9。

1. **Discussion**

我們有嘗試使用不同的backbone architechture，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。

1. **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 concate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x0相加，得到output x1。

2. 加入text:

(a). 將text經過預處理(tile、expand\_dim)，然後和x1 concate起來。

(b). 經過一層convolutional layer和一層normalizaiton，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。

1. **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"。

1. **Discussion:**

在最一開始的嘗試中，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。

1. **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。