

# Lab 4 結報

姓名：林業語

學號：109611036

## 1. 請簡述 data augmentation 的目的與作用

**Data augmentation** 是指在機器學習中使用各種技術，如旋轉、翻轉、縮放、剪裁、添加雜訊等方法，通過對現有訓練數據進行處理，生成更多且多樣化的訓練數據，以增強模型的泛化能力和預測能力。

其目的是為了解決以下三個問題：

**訓練資料量不夠**：對於某些複雜的機器學習任務，擁有足夠的訓練數據對於訓練良好的模型是非常重要的。但是在實際應用中，獲取足夠的訓練數據往往是困難的，因此可以通過增加訓練數據的方法來解決這個問題。

**訓練資料不夠全面**：即使有足夠的訓練數據，有時也會出現資料集中資料特徵太過相似的情況，導致模型學習到的特徵不夠全面，從而影響模型的泛化能力。透過 **data augmentation** 可以生成多樣化的訓練數據，從而增加資料的多樣性，使模型能夠學習更全面的特徵。

**避免過度擬合**：當模型在訓練過程中過度擬合訓練數據時，通常是由於模型複雜度太高，而訓練數據太少或者太單一所導致的。增加訓練數據的方法可以通過減少模型複雜度的方式來解決過度擬合的問題。因此，在進行 **data augmentation** 的同時，通常也需要適當的減少模型複雜度，從而避免過度擬合。

## 2. 請簡述 AE-GAN 的功用

**AE-GAN**（**Adversarial Autoencoder Generative Adversarial Network**）是一種基於自編碼器（**Autoencoder**）和生成對抗網絡（**GAN**）的深度學習模型。它的主要功用是生成與訓練數據相似但不完全相同的新數據，以增加數據集的多樣性。

**AE-GAN** 通過兩個部分組成，一個是自編碼器部分，用於將原始數據壓縮成潛在空間編碼，然後進行解碼以重建原始數據；另一個是生成對抗網絡部分，用於從潛在空間編碼中生成新的數據樣本，通過生成器和鑑別器的對抗訓練，使生成的數據盡可能地逼真。

具體來說，AE-GAN 的運作方式如下：

自編碼器部分：將原始數據經過壓縮編碼，得到潛在空間編碼，然後再進行解碼以重建原始數據。

生成對抗網絡部分：生成器接收潛在空間編碼並生成一個新的數據樣本，鑑別器則試圖識別出生成的數據是否真實的。

對抗訓練：通過生成器和鑑別器的對抗訓練，不斷優化兩者之間的競爭關係，使生成器生成的數據盡可能逼真，而鑑別器則能夠準確區分真實數據和生成數據。

### 3. 心得

這次 Lab 結合了 Lab2 和 Lab3 的內容，我們使用 Lab3 的 generator(訓練到 loss 接近-1)

```
epochs: 20  
dis_loss: 0.0009300112724304199  
gen_loss: -0.9996315240859985
```

以產生虛擬的資料，並與舊有的資料作合併，再用 lab2 中的 CNN+DNN 訓練模型

```
Epoch 30/30  
23/23 [=====] - 0s 20ms/step - loss: 0.0127 - accuracy: 0.9958 - val_loss: 4.2540e-06 - val_accuracy: 1.0000  
Model: "model_17"
```

整體來說，這次實驗不算太難，因為是使用前兩次 Lab 的內容進行合併與修改。

對我而言，這次 Lab 讓我學到了如何在訓練資料量不足的情況下，使用數據擴增(Data Augmentation)的方法來擴展訓練集