

## HW6 Report

### 1. Describe the difference between WGAN and GAN, list at least two differences:

#### Differences:

1. 從 Loss function 的角度對 GAN 做改進優化，使模型即使在 Fully Connected Layer 也能得到很好的表現結果，由於 Cross Entropy 的 Loss Function 不適合衡量具有不相交部分的分佈之間的距離，所以提出了 Wassertein 距離來衡量數據分佈和真實數據分佈之間的距離，從理論上解決訓練不穩定的問題，不再需要小心平衡 Generator 和 Discriminator 的訓練程度。
2. 解決 Collapse Mode 的問題（生成器傾向於生成一些有把握但相似的圖片，而不敢嘗試生成沒把握的新圖片，從而產生樣本缺乏多樣性的問題），使得生成器生成的結果更加多樣化。

因此 WGAN 透過改進以下五點，對 GAN 模型做更進一步的優化：

1. 去除 Discriminator 最後一層的 Sigmoid
2. Generator 和 Discriminator 的 Loss Function 不含有 log
3. 在梯度更新後的權重強制限制於 $[-0.01, 0.01]$ 範圍內，滿足 Lipschitz 連續性條件（在樣本空間中，要求判別器函數的梯度值不大於一個有限的常數  $k$ ）
4. 採用其他 SGD，或是 RMSprop 優化器，不再使用基於動量的優化算法（例如：momentum 或 Adam）
5. 提出了一個具體的評估指標來評估模型訓練的好壞

## 2. Please plot the “Gradient norm” result.

- Use training dataset, set the number of discriminator layer to 4 (minimum requirement)
- Plot two setting: i. weight clipping ii. gradient penalty
- Y-axis: gradient norm (log scale), X-axis: discriminator layer number (from low to high)

Blue Line: WGAN

Orange Line: WGAN-GP

