

ADLxMLDS hw4 報告

b03902083 陳力

Model description

G 的模型事先將頭部敘述用事先訓練好的 50 維 Glove embedding 配上 100 維的 Gaussian noise 當作 Generator 網路的輸入。G 網路首先將輸入線性轉換成 $256 \times 4 \times 4$ 再經過 4 層的 deconvolution layer，filter 數量分別是 128, 64, 32 與 3，kernel 大小都是 5，stride 大小則是 2。這樣子最後的輸出大小 $3 \times 64 \times 64$ ，經過一些簡單處理就可輸出成圖片。

D 的模型則是將圖片與敘述當成輸入，輸出兩個數值代表是否滿足圖片是真實且符合敘述的機率。D 網路首先將輸入經過 4 層的 convolution layer，filter 數量分別是 32, 64, 128, 256，kernel 大小都是 5，stride 都是 2。這樣處理後會得到 $256 \times 4 \times 4$ 的輸出，然後將敘述經過線性轉換變成 256 維的向量 \vec{t} ，然後把 \vec{t} 複製 $4 \times 4 = 16$ 份，跟前面 CNN 的結果接起來變成 $512 \times 4 \times 4$ 的 tensor，在經過一層 kernel 大小為 1 的 convolution layer 而成 $128 \times 4 \times 4$ 的 tensor；拉平後過一層 fully connected layer 到 2 個輸出。

以上兩個網路在每一層 convolution/deconvolution 後都會做一次 batch normalization。

在訓練的過程中，G 網路的訓練目標是要最大化 $\log D(G(\vec{z}), \vec{z})_1$ ，也就是增加 D 網路判斷 G 網路生出來的東西跟真實世界相符合的機率。

而 D 網路的目標則是要最大化

$\log D(x, z)_1 + (\log D(x, z')_0 + \log D(G(z), z)_0 + \log D(x', z)_0)/3$ ，也就是想要增加正確判斷 1. 訓練資料裡圖片與敘述 2. 真實圖片與錯誤敘述 3. G 網路生成圖片與正確敘述 與 4. 錯誤真實圖片與正確敘述的機率。

我使用的是 Adam 優化法搭配 0.0002 的學習率，每兩次 G 網路更新才會去做一次 D 網路更新。

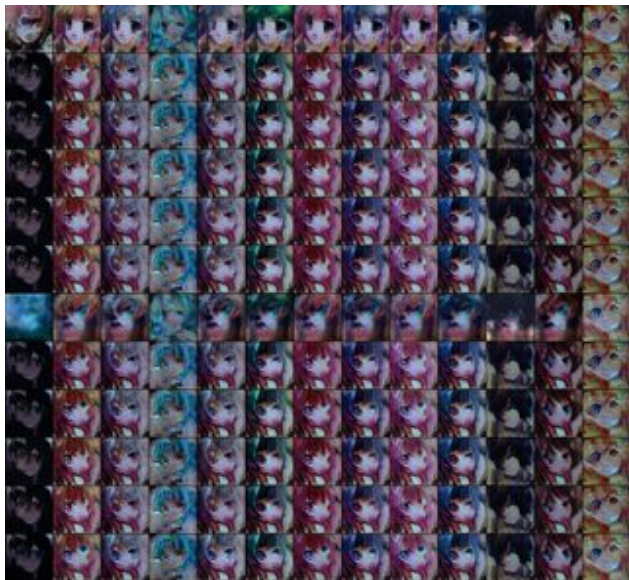
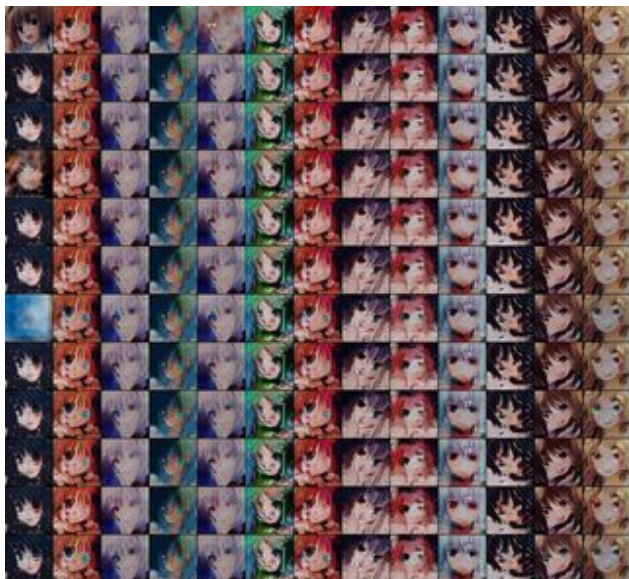
Performance Improvement

為了提升出來的成果，我所調整的參數是 D 網路與 G 網路更新的頻率。我試過 1. 2G1D (兩次 G 網路更新再做一次 D 網路更新) 2. 1G1D 3. 4G1D 4. 1G2D。

輸出成圖片時，我會先將整張圖片 scale 到 0~1 之間，然後為了增強顏色，會每個 pixel 都變成平方，這樣越小的數值會被縮得更小，增強對比。

Experiments

以下是對於各種更新頻率跑了 500 個 epoch 的結果。每個 column 都是同一種髮色對應不同眼睛顏色的生成結果。

D2G1**D1G1****D1G2****D1G4**

用肉眼觀察發現 G 網路更新頻率越高，生成出來的圖片完整度越好。

相對於頭髮，眼睛的條件相當不明顯。

以下是 D1G2 每隔 100 epoch 的輸出結果：

10



120

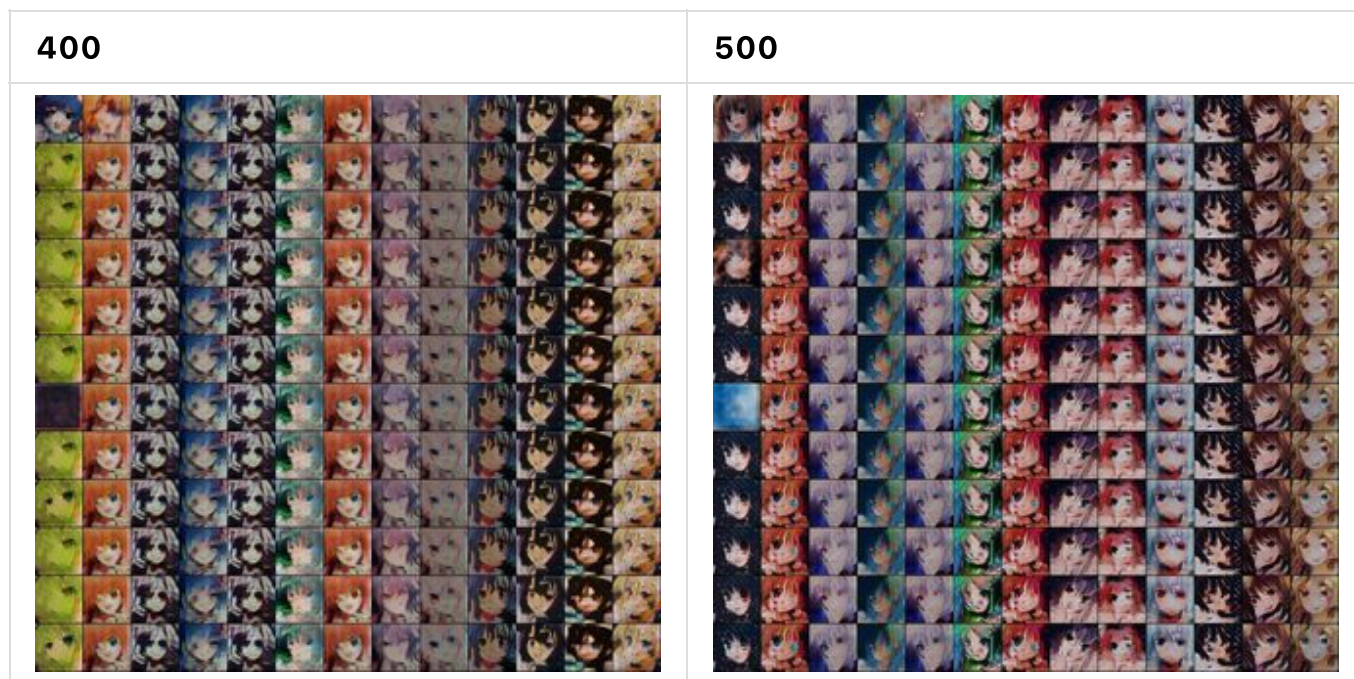


200



300



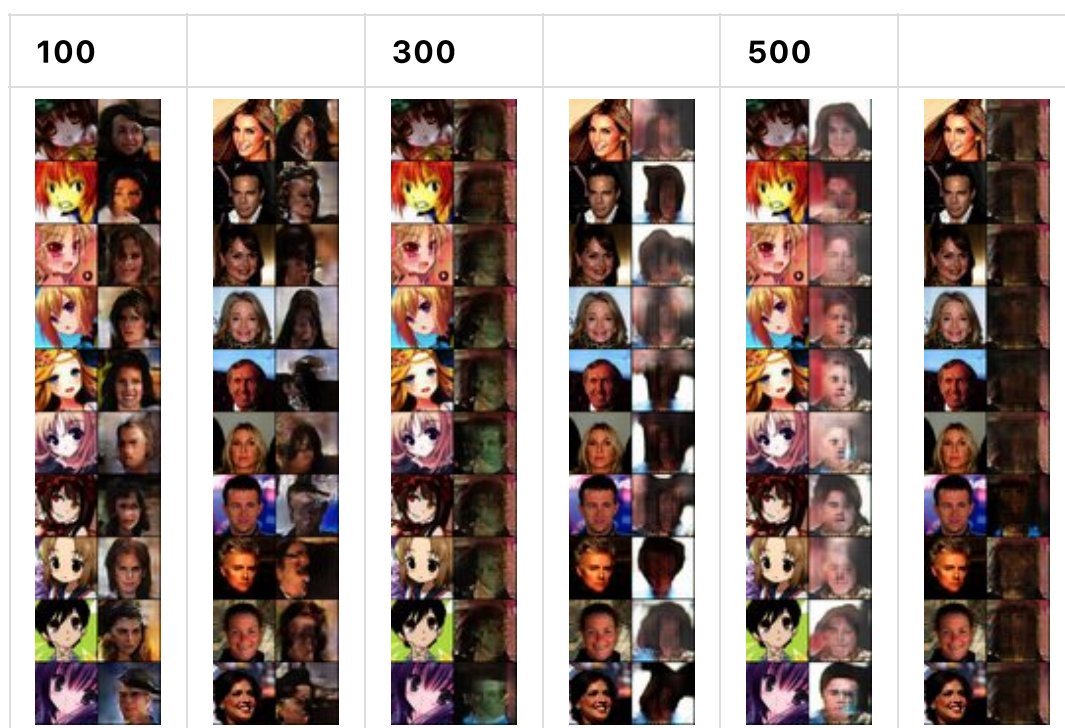


大概 100 個 epoch 後就已經生出很有動漫頭像風格的圖片了，而 200 個 epoch 後就頭髮顏色已經差不多出來了。

Bonus

我在 bonus 部份實作的是使用 cycle GAN 做動漫頭像與 celeba 真人臉孔之間的風格轉換。

以下是分別跑了 100, 300 與 500 個 epoch 的轉換結果



Reference

Generative Adversarial Text to Image Synthesis (<https://arxiv.org/pdf/1605.05396.pdf>)

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

(<https://arxiv.org/pdf/1703.10593.pdf>)