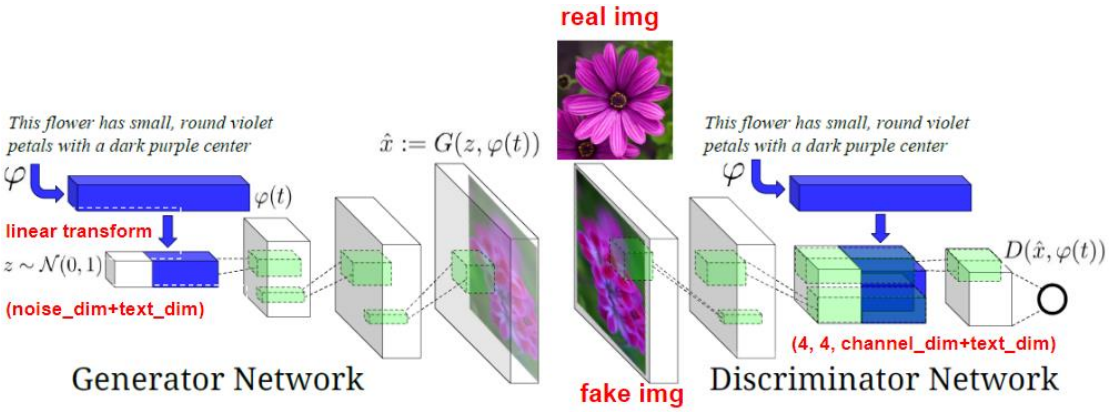
ADL HW4: Generative Adversarial Networks

R06725035 陳廷易

Model description

Model structure

本次作業依據助教所提供的論文 *Generative Adversarial Text to Image Synthesis* 來對 deep convolutional generative adversarial network 進行修改與實作。



DCGAN亦即將原先GAN中的discriminator與generator抽換成兩個CNN。在discriminator方面，利用stride來取代pooling，activation function為使用Leaky Relu(alpha=0.2)；在generator方面則將pooling拿掉，改為使用transposed convolutional layer，activation function為Relu，最後一層則為hyperbolic tangent。而Input時皆有使用batch normalization以避免collapse。

- Image size = 64*64
- Optimizer = adam
- Momentum = 0.5
- Learning rate = 0.0002
- Updates between Discriminator and Generator = 1:1
- Normal distribution, noise dim(z) = 100
- Batch size = 256
- Epoch = 100

Objective function for G

利用一般生成對抗網路之目標函數：

$$\min\left\{E_{y\sim p_y\ ,z\sim p_z(z)}[-\log D(G(z|y))]\right\}$$

※註: y 代表文字向量、z 則代表 noise

Objective function for D

而 Discriminator loss function 則包含了(real image, right text)、(real image, wrong text)、(fake image, right text)三種 case，將此三種的 cross entropy 進行加權平均。

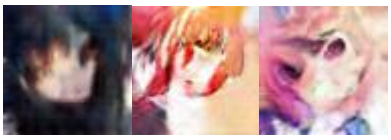
$$\min\Big\{-\Big[E_{x,y\sim P_{data}\ (x|y)}[\log D(x|y)]+E_{x\sim P_{data}\ (x|y),\hat{y}\sim P_y\ ,y\neq\hat{y}}[\log(1-D(x|\hat{y}))]$$

$$+E_{y\sim P_{data}\ (x|y),\hat{x}\sim P_x\ ,x\neq\hat{x}}[\log(1-D(\hat{x}|y))]+E_{y\sim P_y\ ,z\sim P_z(z)}[\log(1-D(G(z|y)))]\Big]\Big\}$$

※註: y 代表文字向量、z 代表 noise、x 代表圖片

Performance Improvement

Data Preprocessing

原先採用全部的 label 與 image 進行訓練，然而訓練結果搭配頭髮與眼睛 condition 時，成像卻會有點模糊。 

雖有可能是因為本身 dataset 當中就含有許多有問題的圖片，使 model 不易學到該呈現的樣子。另一個可能原因就是因為 tag 的數目種類太多，致使 model 無法真正專注在頭髮與眼睛顏色的部分。

因此改將所有 tag 進行過濾，只留下符合助教所給的條件，具頭髮顏色或是眼睛顏色的 tag。若 tag 只含其中一種則將另一條件給予 unk，而若 tag 間有衝突的情況便捨棄該對應的 image。

Data Augmentation

經過 data processing 將許多 image 及雜訊 tag 濾掉，雖使 model 較能夠根據 condition 產生圖片，但也因訓練資料大減少，使效果較為差強人意。

為了提升清晰度，便依據助教的 tip，在不使用額外 dataset 的情況下，利用 scikit-image 對圖片進行旋轉正負二十度，彌補原先的不足。

此外，因後來 deadline 向後延了一個禮拜，便使我能有些時間可以稍微整理一下 dataset。將一些沒有頭像的 image 剔除或是將缺少眼睛頭髮的 tag 補上，雖因時間不多只能完成一部分，但從結果來看的確有稍微提升 image 的正確度與清晰度。

Batch Normalization

在實驗過程中也發現 batch normalization 對一般 DCGAN 來說相當重要，在 model 架構一樣的情況下，若未進行 batch normalization 幾乎很難 train 起來。但如果是在實作 improved WGAN 時，則會變成反過來，不應該在 discriminator 使用 batch normalization，才可以得到比較好的效果。

Discriminator Output

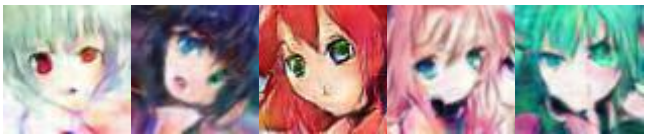
使用助教所建議的 training tips，只有(real image, right text)才是正確的結果，而將(fake image, right text)、(real image, wrong text)及(wrong image, right text)都讓 discriminator 都判定為錯誤的結果。將此三項的 loss 相加予 optimizer 進行 minimize。

Noise

在實作時也發現 noise 雖然可以使所產生出來的圖片可以有比較大的差異，但同時也會提升產生出壞掉 image 的可能。

Experiment settings and Observation

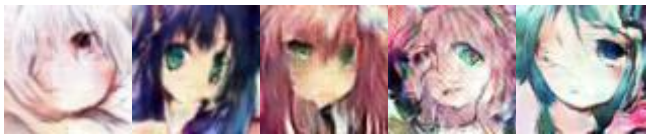
經前面所述的 model 與實驗、改進以後，所產生最好結果的圖片約莫如下：



Wasserstein GAN

在 WGAN 的部分，改成要 minimize earth mover distance，因此 discriminator 的 output 就從原本的 probability 變成衡量距離是否變近，有遠近關係也就可用來衡量 performance 是否進步。除了對 weight 進行 clipping 以外，也將原先 discriminator 最後輸出的 sigmoid 去除，改成最後一層的捲積層直接輸出，當成是衡量的分數。

而 discriminator 的 objective function 也和原先的 DCGAN 雷同，但並非採用 cross entropy，而是改將 positive case 的分數檢調 negative case 的分數。在這邊 optimizer 採用的是 RMSProp，每個 batch 會更新 discriminator 五次再更新 generator 一次。以下為產生的結果圖片：



Improved W-GAN

而 improved WGAN 則希望 norm 越接近 1 越好，使大於小於 1 都會有 penalty。原本 weight clipping 會使 discriminator 的 distribution 大多是集中在 clipping 的地方，使找到的 function 不太容易複雜，然而 improved WGAN 就會有比較正常的 distribution。Improved WGAN 的輸出與目標函數皆與 WGAN 相同，只是會另外加上 penalty (實驗 scale=10)。Optimizer 則使用 Adam，每個 batch 也是先更新五次 discriminator 參數再更新 generator 參數一次。產生結果如下：

在此次 Comics Generation 中，以所給的 condition 正確度而言，這幾種 model 幾乎不分軒輊。不過在圖片清晰度上，或許是因為參數跟 model 尚未調校到最佳的緣故，目前以 DCGAN 的 image 稍微銳利一點。

Bonus—Style Transfer

Animation to Sketch—Cycle GAN

此次 style transfer 要做的是將助教所給的 dataset 轉為素描的圖案，而所利用到的便是老師上課所講述的 Cycle GAN，素描的 dataset 是使用 CUHK Face Sketch Database。在 model 方面，主要是參考 *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks* 這篇論文進行修改與實作，架構與參數使用上基本上雷同。利用 cycle GAN 重建的特性，藉由 dual learning 以達到 unpaired data 的非監督式學習。

所使用的 Loss function 有三類：前兩類為真的 sample 與假的 sample 之 loss，只是方向不同兩者的真假 sample 互調，利用 mean square error 來計算。而第三類為重建誤差，希望還原的圖片能與原本真實的圖片越近越好。Optimizer 使用的是 Adam，learning rate 為 2e-4。

經過相當長的 training 時間，最終結果呈現如下：

可看出所繪出的素描圖皆相當逼真，也確認過並無與素描 dataset 有重複的狀況。但或許因為在 sketch dataset 當中大多為中年男性，且畫風一致性過高，致使很難利用素描來詮釋動畫風格的樣貌，也就不易保留特徵細節。

如果是將素描畫轉為動畫風格，基本上是 train 不太起來，結果大部分是會壞掉的：