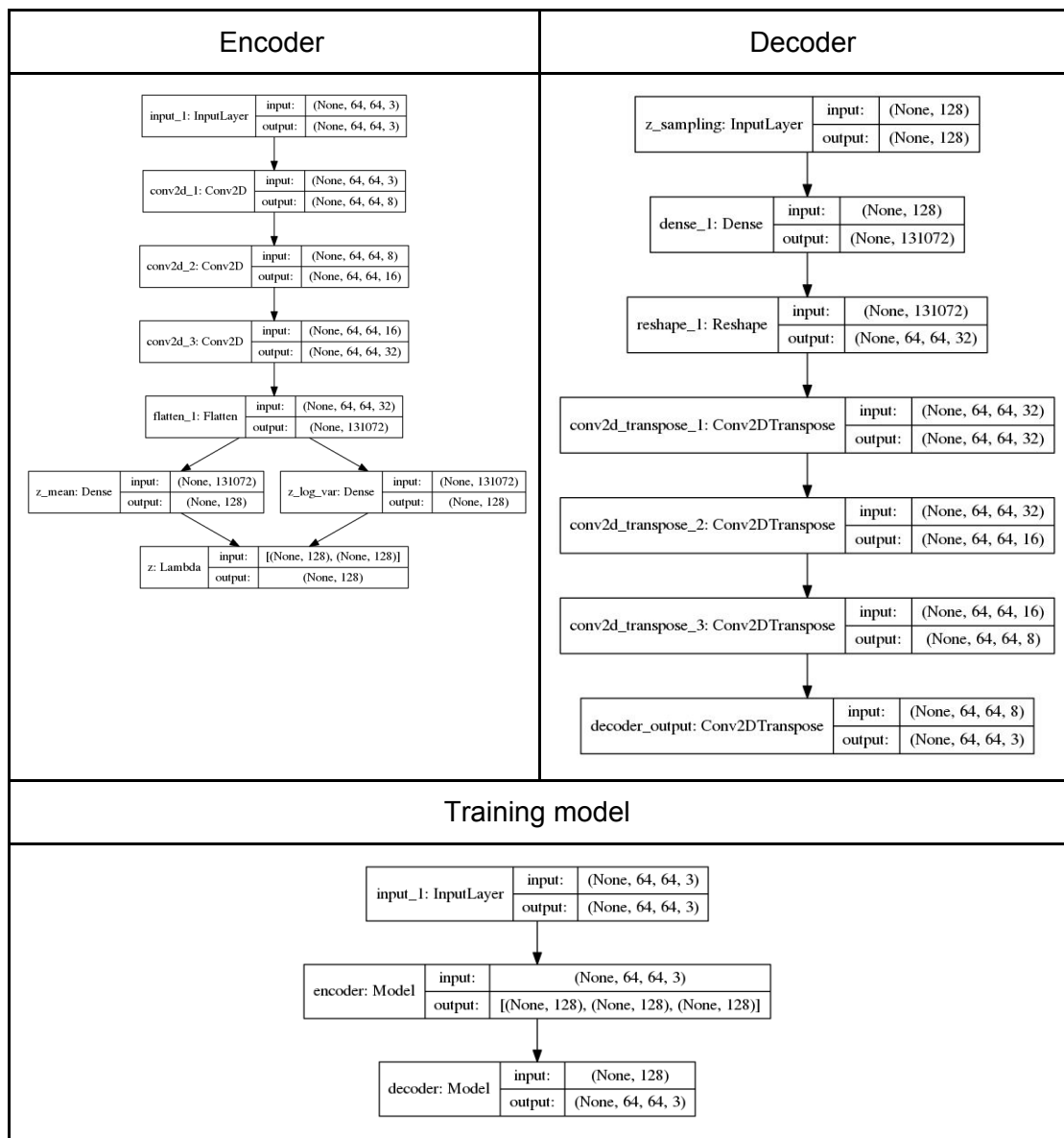


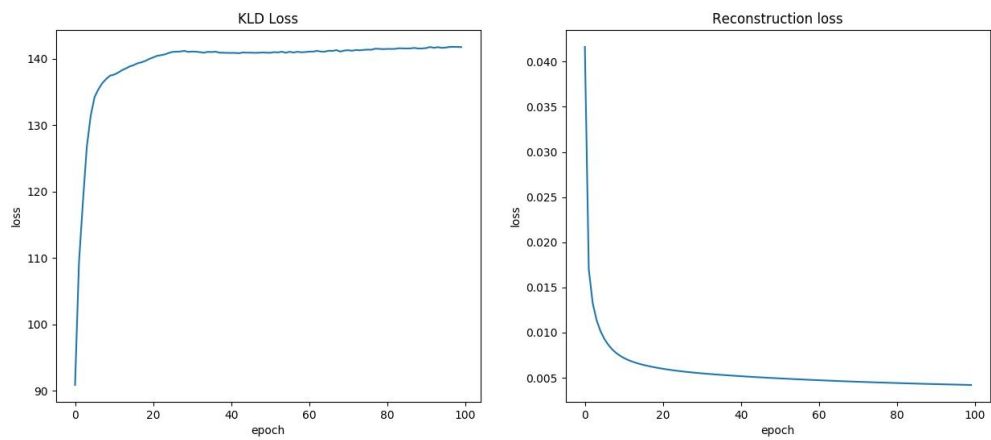
Problem 1. VAE

1. Architecture & Implementation:

- Encoder:** 將normalized 圖片通過 3層 CNN後, Faltten 後接上一層 256 node 數的 Fully connected layer (FC) 作為特徵向量的平均值, 再從 Flatten 中接上另一層 FC 作為特徵向量取 log 後的變異數。最後為了能夠使整個架構可以利用 back propargation 做 gradient descent, 必需將 gaussian sample 部份變成 Lambda layer 接上平均值和變異數層。
- Decoder:** 將 256 維的特徵向量通過 node 數為64*64*32 的 FC, 並 Reshape 成 64,64,32 的三維陣列, 經過3 個 Conv2DTranspose 後變成 64*64*3的陣列, 即可視為一張 64*64的 RGB 圖片。
- Training Process:** 將圖片 normalize 成 [-1,1] 後經過 Encoder 得到特徵向量的平均值、log 變異數、sample 特徵向量, 從平均值及log 變異數求 KL Divergance, 並將 sample 特徵通過decoder 所得的圖片與原圖算 mean square error





2. Plot learning curve



3. Plot 10 testing images

Reconstruction MSEloss of testing images = 0.0059

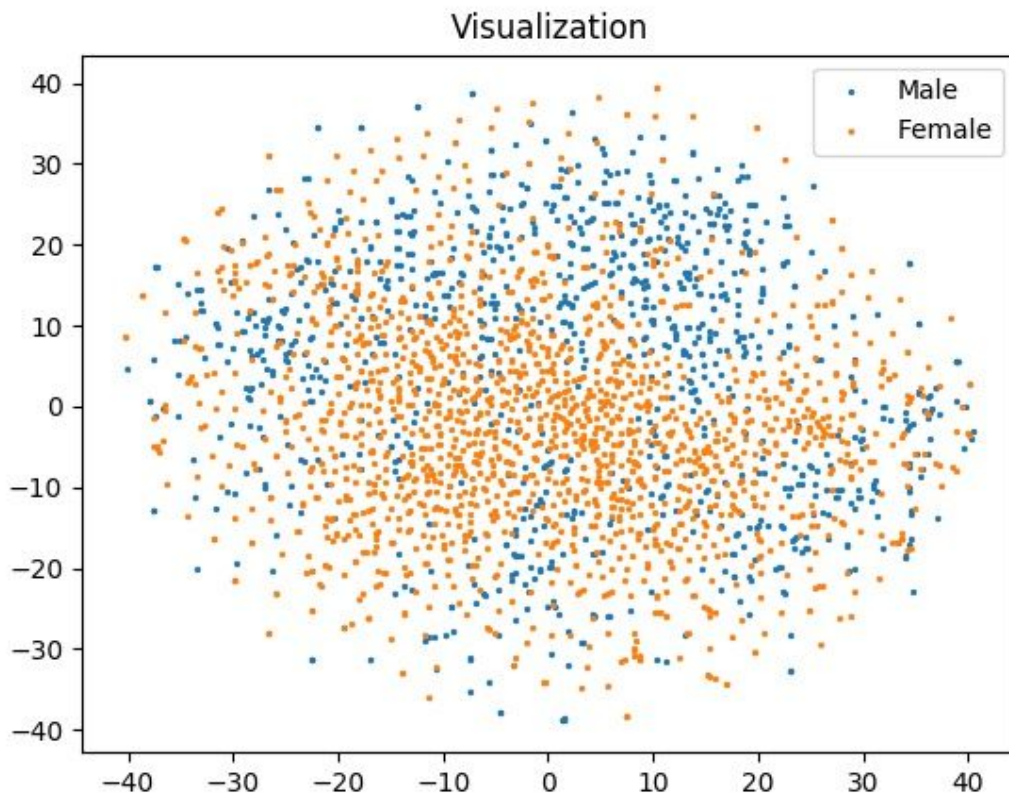
Image	
Reconstruction	

4. Plot 32 random generated images



5. Visualization

由下圖可知，降維後男性與女性的分佈雖然有些混雜，但仍可發現男性偏於周圍，女性偏於中央。



6. Discuss

VAE 的好處是可以從 loss 判斷出 model 是否有train 成功，實作方面也相對簡單。然而，壞處是由於 KLD 和 MSE 是一種 trade off，換句話說，當 reconstruction 太好時，generated 的圖片就沒有那麼漂亮。

Problem 2. GAN

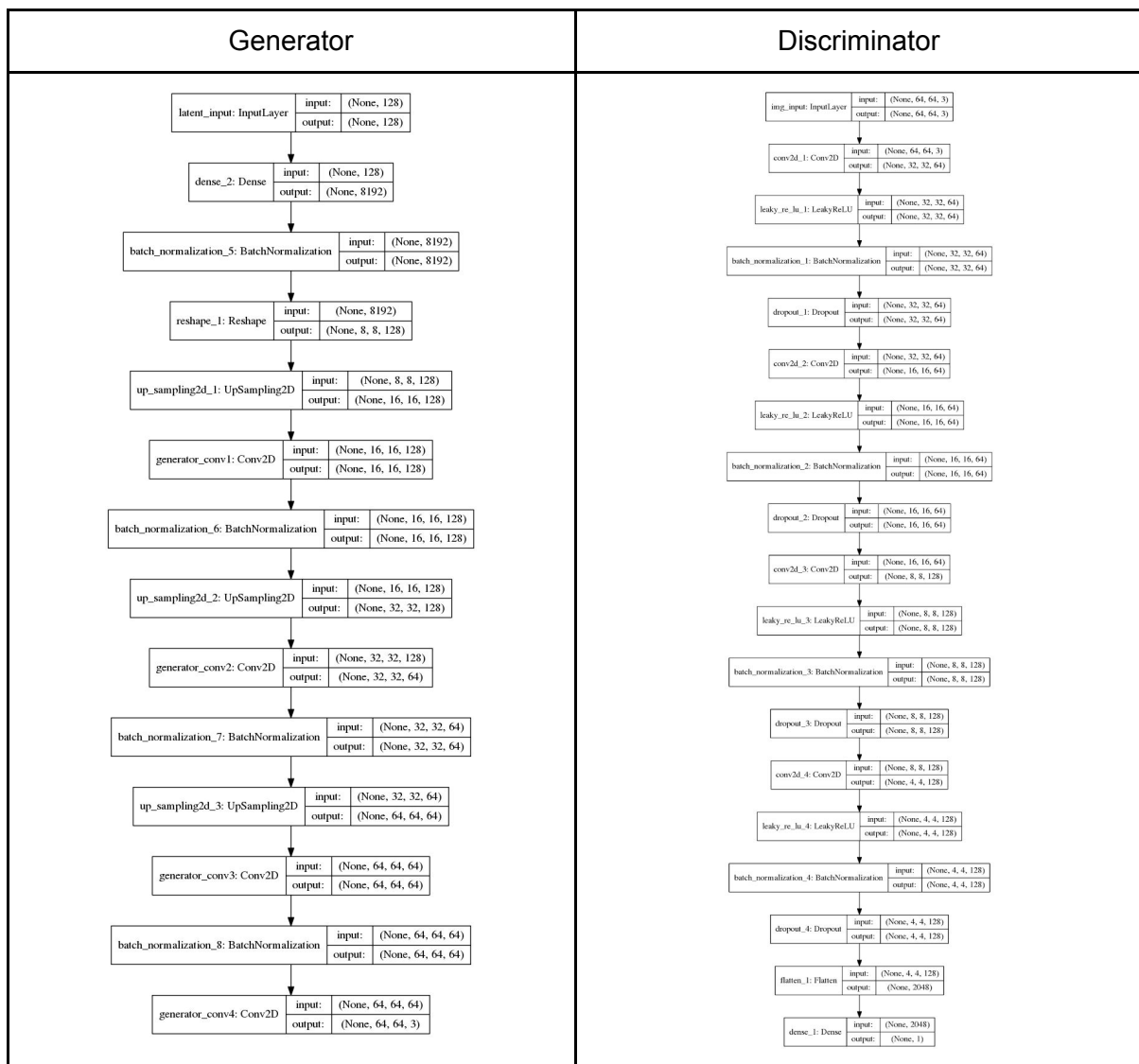
1. Architecture & Implementation

a. Generator:

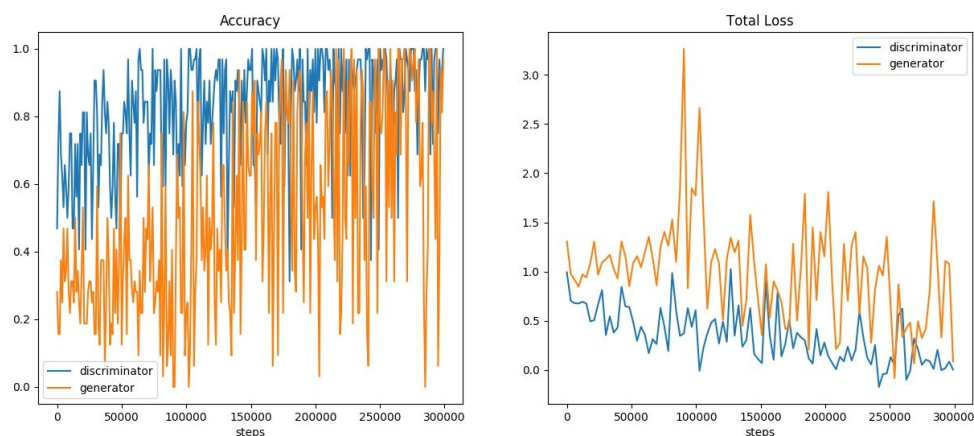
先將 128 維的特徵向量通過 $8 \times 8 \times 128$ node 數的 Dense，並 Reshape 成 $(8, 8, 128)$ 的三維陣列。之後通過 Upsampling 及 Conv2D 將前兩維放大至 $(64, 64)$ channel 為 3，中間每個 layer 都是用 relu 作為 activation，並加上 batch Normalization。最後一層則是 tanh 讓輸出介於 -1, 1 之間。相當於 normalized 的 RGB 圖片。

b. Discriminator:

將 normalized 圖片通過 4 層 CNN，其中 activation 為 LeakyReLU，並加入 Batch Normalization 以及 Dropout。Flatten 後通過一個 node 的 Dense，activation 為 sigmoid，期望將真實的圖片判斷為 1，假的圖片為 0。



2. Plot the learning curve and briefly explain



由learn curve 可知，generator 和 discriminator 的 loss 及 accuracy 的趨勢大約一致。舉例來說，當 discriminator 的 loss 下降時，generator 也同時會有下降的趨勢。而從 loss 圖可以看出來兩者的 loss 的確有稍微下降的傾向，因此可以確定這組 discriminator 及 generator 是有透過競爭來 training 的。

3. Plot 32 random generated images



4. Discuss

GAN 雖然理論很厲害，實作上卻很困難，由其不穩定的情況很容易讓 model train 壞掉。比如發生 gradient vanishing 及 mode collapse 的現象。gradient vanishing 主要是 generator 或 discriminator 其中一方太強，造成另一方的參數無法更新。mode collapse 則是 generator 只學到部份的 data 分佈，造成產生的圖片長相相似。因此，我嘗試實作 WGAN，並搭配 label smoothing，但以這次的 task 而言，效果並沒有比較好，甚至更糟。

5. Compare the difference between images generated by VAE and GAN

VAE 由於是用 MSE 做為 loss function，所產生的影像比較模糊，但人臉的輪廓通常都還保持得住。反觀 GAN，雖然影像似乎比較清晰，但壞掉的影像就真的很慘，連輪廓都看不出來，甚至爛掉的圖片就算多 train 幾個 epoch 也不會有進步的趨勢。

Problem 3. ACGAN

1. Architecture & Implementation

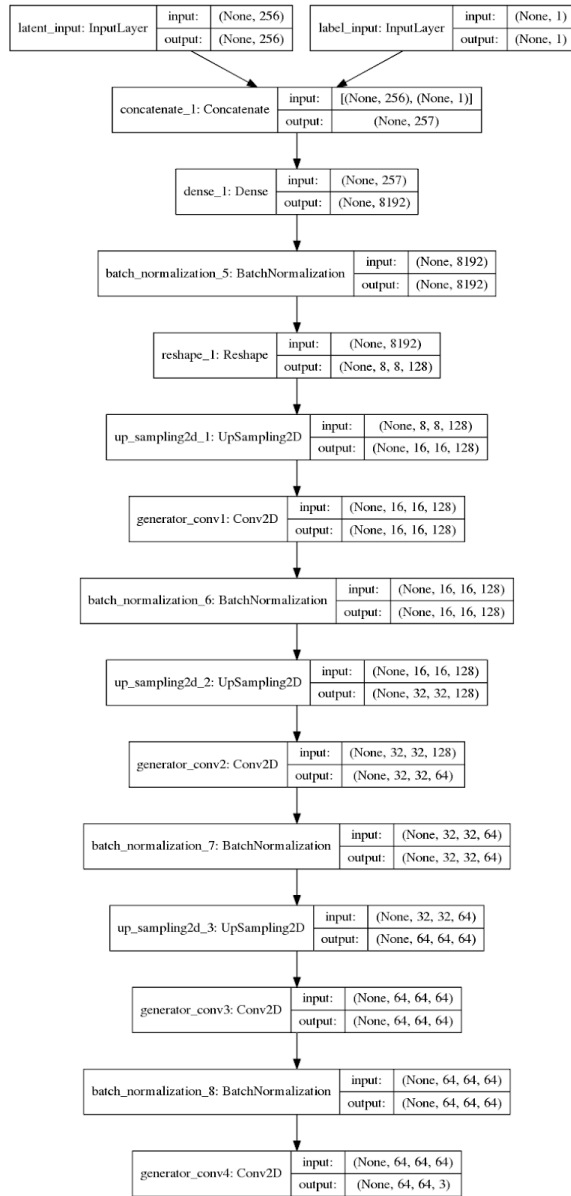
a. Generator:

將 256 維的特徵向量和一個 attribute 合併後經過 Dense 並 Reshape 成 (8,8,128) 的三維陣列，再經過 3 個 Conv2D 與 Upsampling 得到 (64,64,3) 輸出的，如同上述的 GAN 一樣，之間使用的 activation 為 relu，並加入 Batch Normalization。最後的 Output 則是使用 tanh 將數值限制在 $[-1,1]$ 。

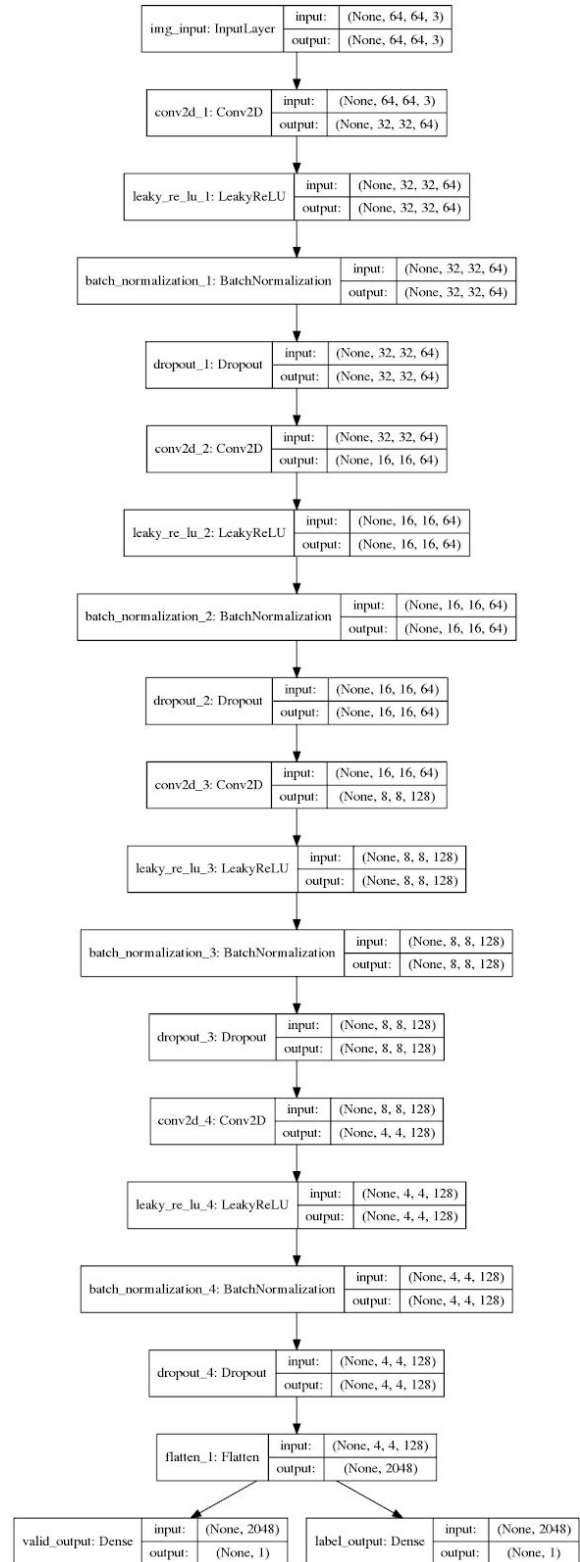
b. Discriminator:

輸入一張 (64,64,3) 的 normalized 圖片，經過 Conv2D (activation 為 Leaky ReLU)、Batch Normalization 及 Dropout 後，Flatten 並經過兩個 node 為 1 的 Dense，分別判斷圖片的真偽以及 attribute。

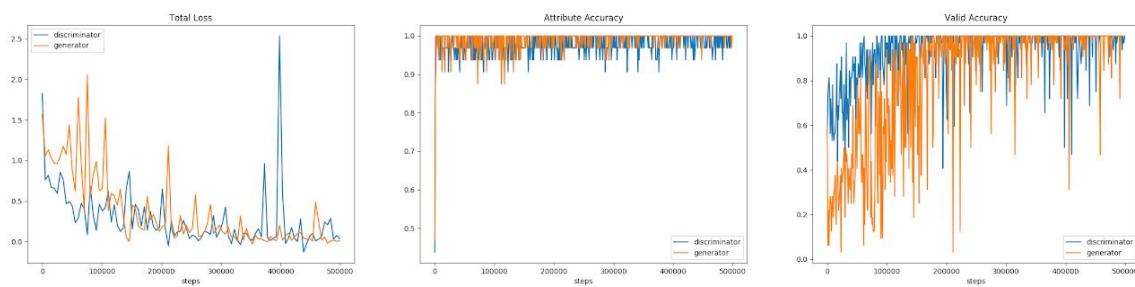
Generator



Discriminator




2. Plot the learning curve and briefly explain



由learn curve 可知，如同GAN一樣，generator 和 discriminator 的 loss 及 accuracy 的趨勢大約一致。舉例來說，當 discriminator 的 valid accuracy 上升時，generator 也同時會有上升的趨勢。而從 loss 圖可以看出來兩者的 loss 的確有持續下降的傾向，因此可以確定這組 discriminator 及 generator 是有透過競爭來 training 的。另外，Attribute accuracy 大部份都在 0.8 到 1.0 之間振盪，因此可以知道generator 是有學到 attribute 給的訊息的。

3. Plot 10 pair of random generated images

Not smiling	
Smiling	