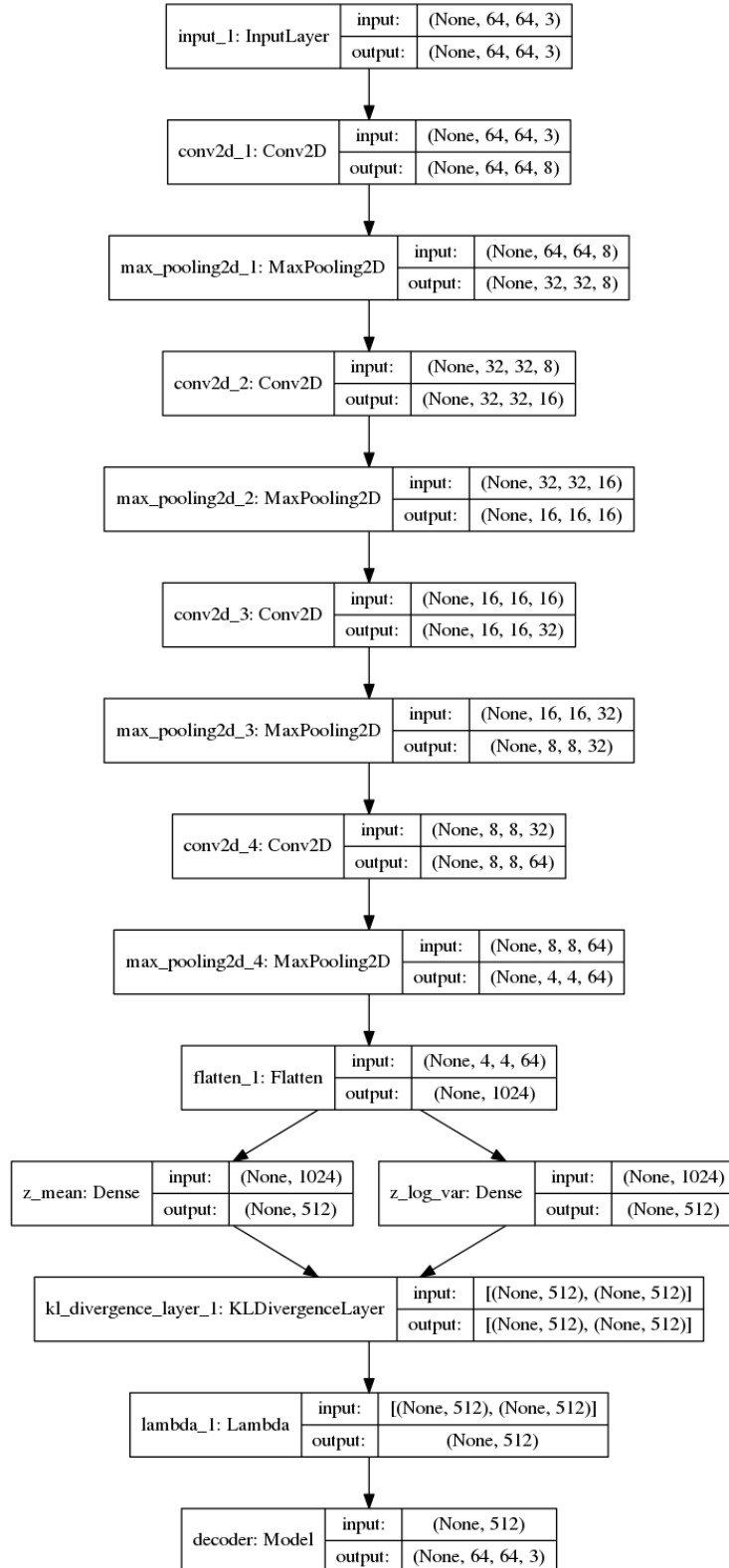


1. Variational Auto Encoder

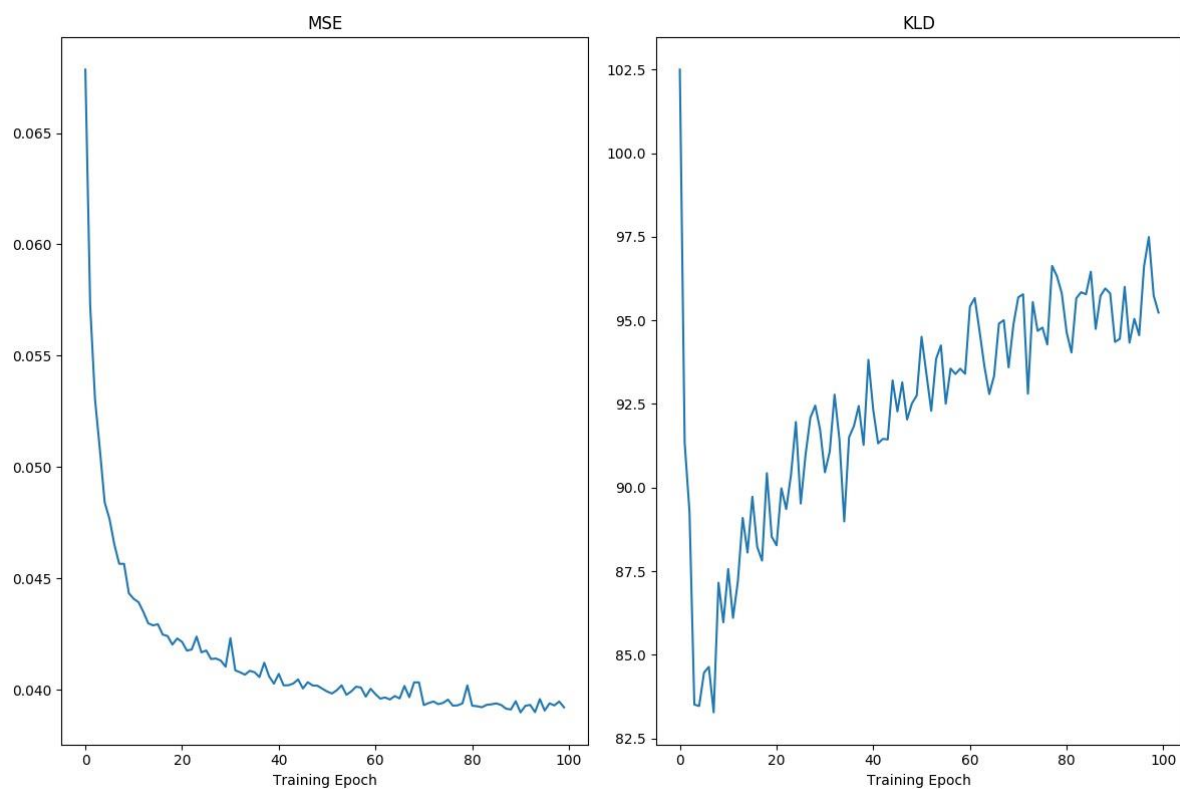
a. Architecture

VAE 的結構如下圖，KL 的 coefficient 設為 $1e-4$ ，因為發現 $1e-5$ 似乎產生的圖片完全沒有臉的樣子才將之調大一個量級。



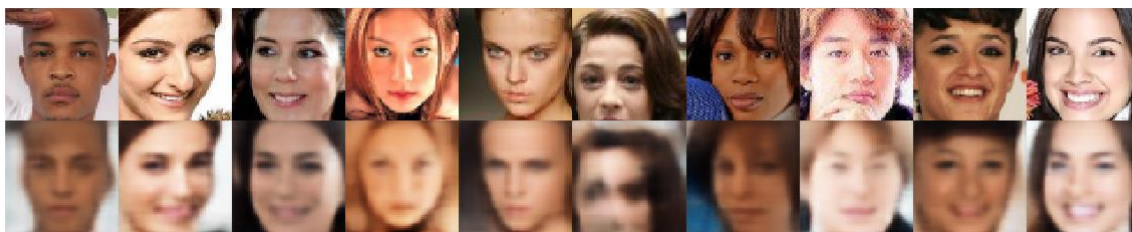
b. Learning curve

KLD 以及 MSE 都是一個 epoch 畫一次



c. Reconstruction images

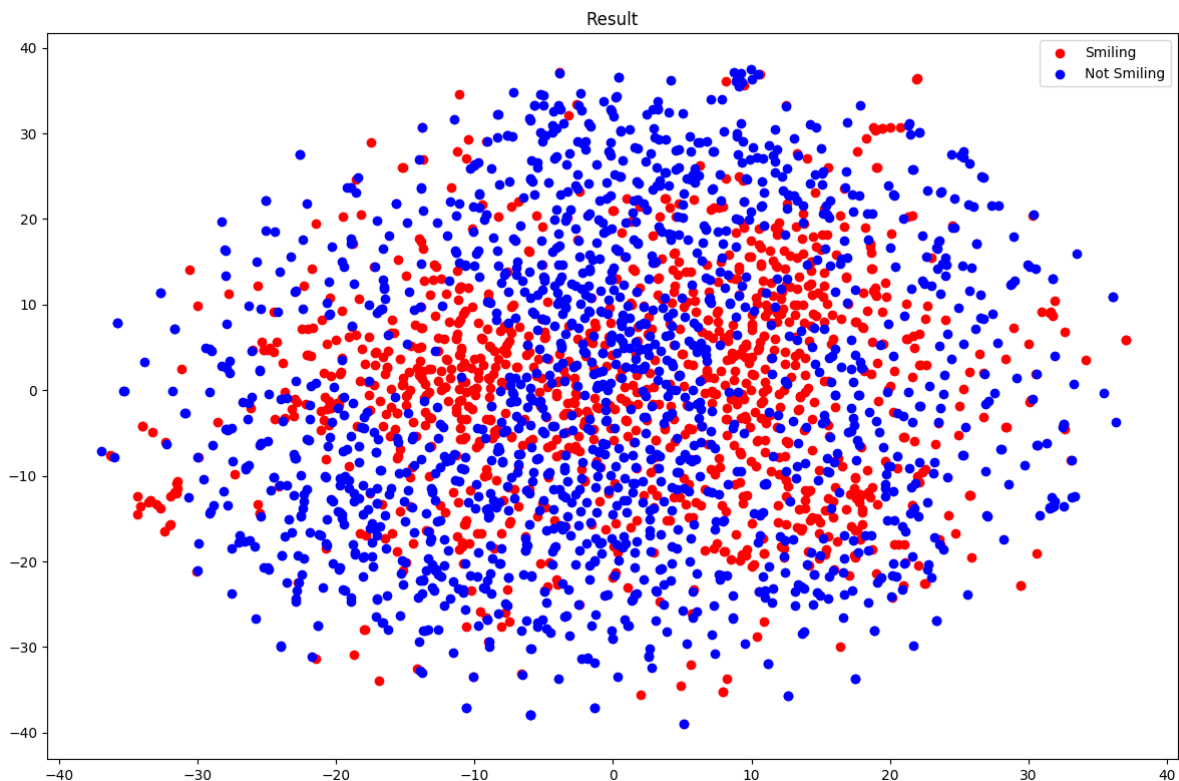
整個 test set 的 MSE 為 0.0392



d. 32 Random generated images



e. Latent space visualization



f. Findings and learnings

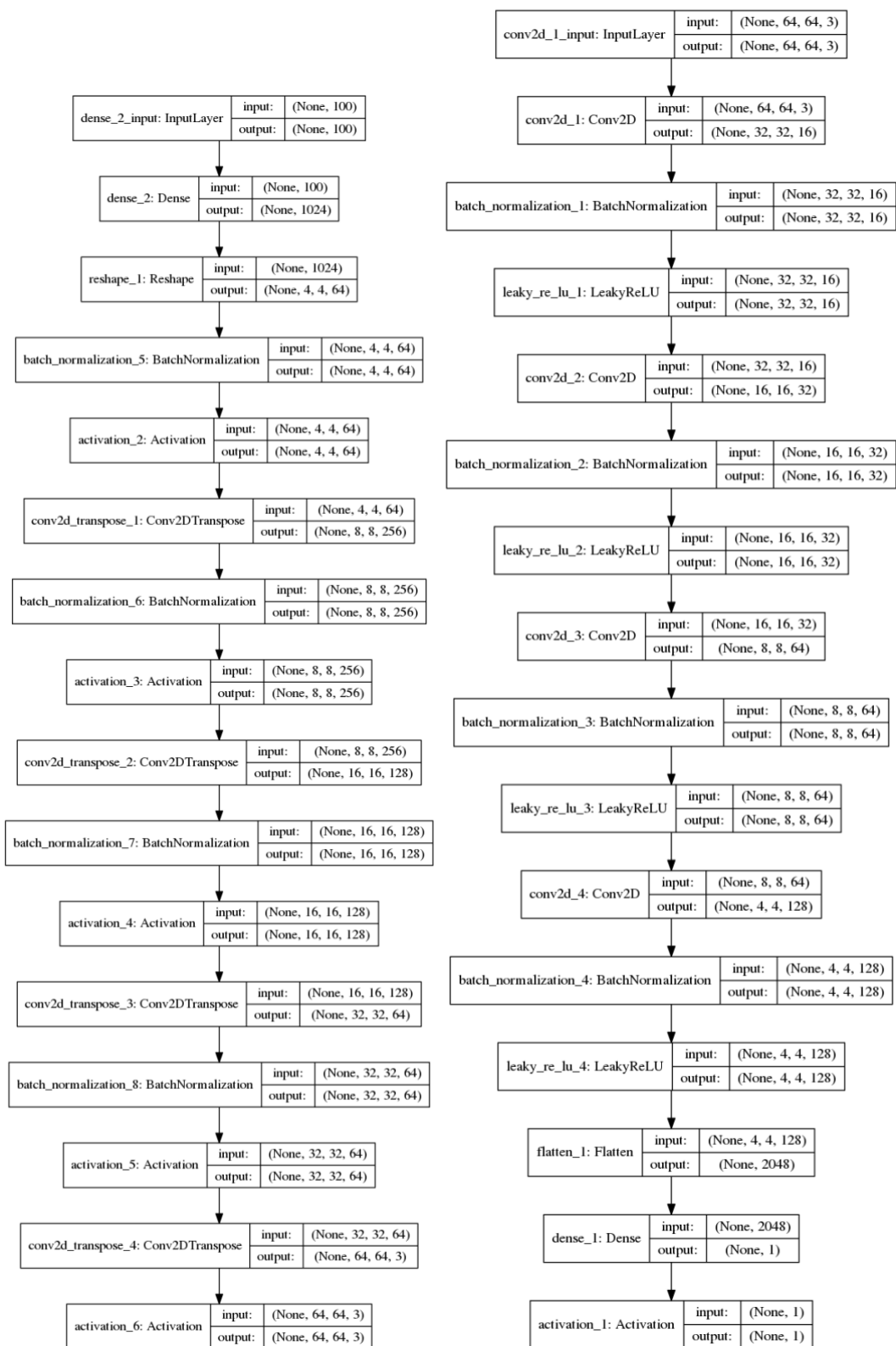
我發現 VAE 雖然看似加入 Gaussian distribution 的概念進入 auto-encoder，但是可能也正是因为這個隨機的因素導致在 training 的過程並不像 auto-encoder 這麼單純，需要同時考量 reconstruction error 以及與 Gaussian distribution 的相似程度，因此在調整參數的過程中還是花了蠻多時間的。而且如上面的範例所示，效果其實還並不是很好 (上面的圖還是從幾百張裡面挑出比較好的 seed 的結果了...)

2. Generative Adversarial Network

a. Architecture

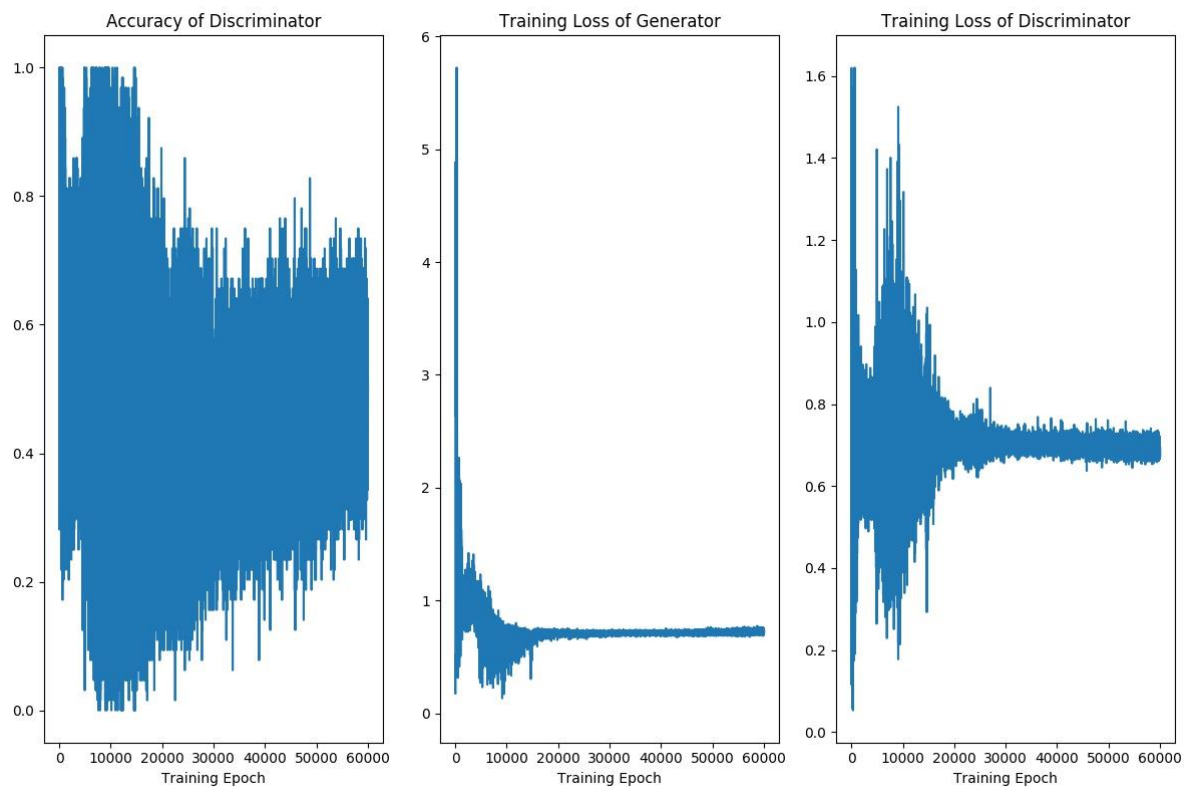
左邊為 generator 右邊為 discriminator，結構上基本上是将 VAE 的 encoder 跟 decoder 做一些修改，譬如說 discriminator 的 activation 改成 leaky ReLU，也將 discriminator 的 feature map 的數量調小，避免陷入 discriminator 太強與 generator 差距過大沒有 gradient 的問題。

另外一個比較特別的地方是雖然我將 input image 減去 127.5 再除以 127.5 讓數值維持在 -1 跟 1 之間，但有別於一般 generator input 的 noise 會從 -1 到 1 之間 sample，我發現限縮在 0 到 1 之間對於 training 的過程會更穩定，成功率也較高。我想可能是因為 generator 所需要學習的 range 較小(?)，因此可能比較可以跟得上 discriminator 的腳步。因為原本從 -1 到 1 之間 sample 的時候，discriminator real fake accuracy 很容易一下就 100% 了。



b. Learning curve

我發現在後期，generator 跟 discriminator 的 loss 都很穩定，這可能也是為什麼將 sample noise 改成 0 到 1 之間之後 training process 會比較穩定，而且產生的圖的品質也會較好的圖示。



c. 32 Random generated images



d. Findings and learnings

雖然說 GAN 就是 discriminator 跟 generator 在那邊左右互搏，聽起來概念很簡單，但真正 training 的時候就會發現兩邊很難平衡，一邊太強或太弱都會導致 gradient 不見，另一邊直接放棄。經過這次作業之後大概可以掌握到一些 training 的技巧，就是看哪邊太強就想辦法砍那邊，譬如說加入 random label flip, noisy label 或是減少 feature map 的數量。不過最後還沒能掌握這些技巧的加入時機，導致加入之後 image 品質都不是很好。而在陰錯陽差發現將 sample 的 range 限縮在 0 跟 1 之間似乎也不錯的效果。

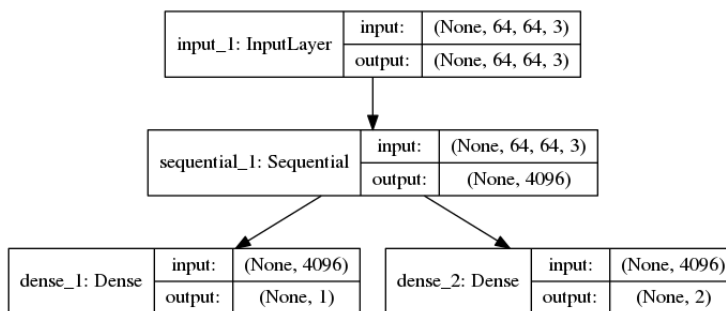
e. Differences between VAE and GAN

可以發現 VAE 所產生的圖片基本上很糊而且顏色都怪怪的也很單一，GAN 相對來說就比較多元，而且甚至有一些背景的感覺。這根上課所學到的理論也差不多，而這也是當初 GAN 被發明出來的原因—解決 VAE 的不足。

3.

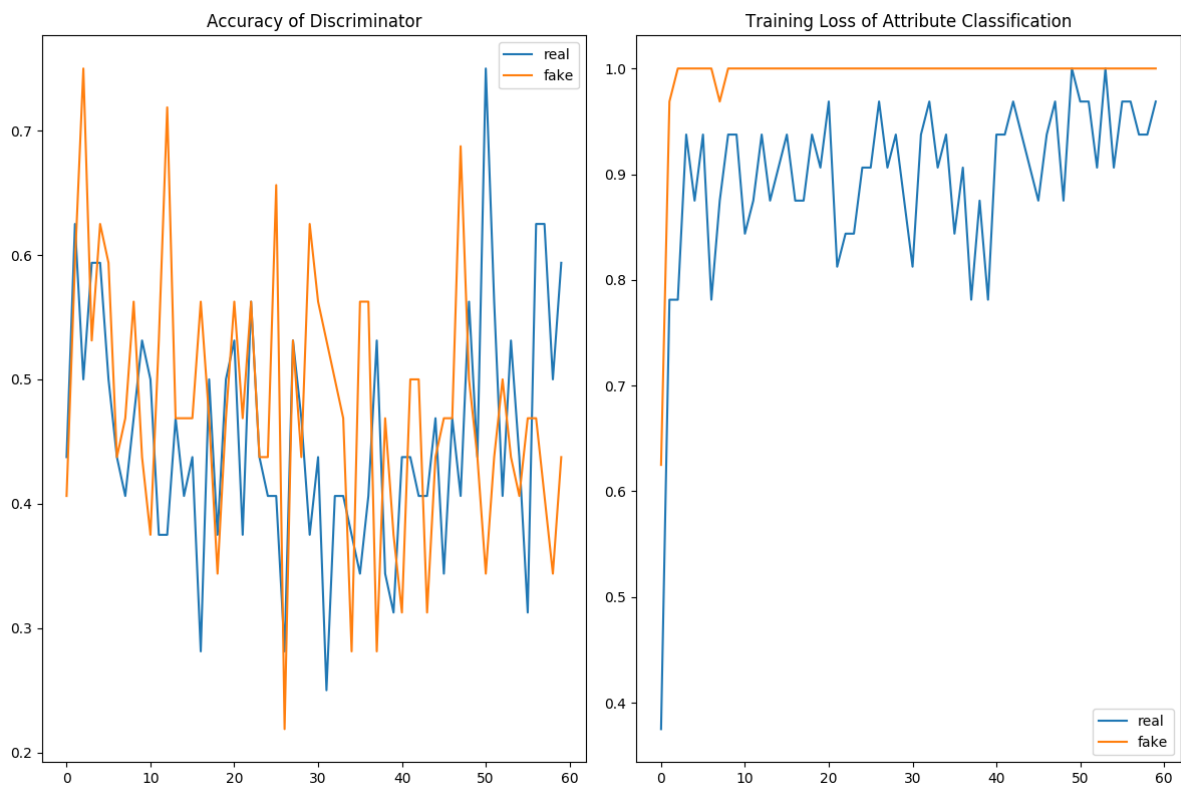
a. Architecture

Generator 的部分跟 DCGAN 是一樣的，只差在 input 的地方多了一個 attribute 的 0 或 1，而 generator 則差在最後面會分別將 latent feature 過兩個 fully connected layer 去判斷真假跟 attribute 是否成立。



b. Learning curves

兩張都是每 1000 個 step 畫一次圖。左邊 discriminator 在判斷是否為真實圖片的 loss 在 real 跟 fake 上是差不多的，但可以看到最後面有一點分開的趨勢，可能再 train 下去會 mode collapse 或是爛掉。而右邊在判斷是否某張照片有 smile 則是在 smile=1 的時候比較不這麼肯定，這可能說明了有些時候笑並不是那麼明顯，可能只是嘴角上揚等等，較不好學。



c. 10 pairs

Not Smiling

