DLCV hw4

電機三
b04901081
鄭元嘉

Problem 1 VAE

1. Model

(a) Implementation

		Output Size	Kernel Size	Feat. Map size
Encode				
Layer 1	**Convolution + ReLU**	64	(3, 3)	64
Layer 2	**Convolution + ReLU**	128	(3, 3)	64
Layer 3	**Pooling**	128	-	32
	Reshape	32 * 32 * 128	-	-
Layer 4	**Fully Connected**	1024	-	-
Layer 5	**Tanh**	1024	-	-
Layer 6	**Latent Space**	512	-	-
Decode				
	Reshape	512	-	1
Layer 7	**Transpose Conv.**	64	(32, 32)	32
Layer 8	**Convolution + ReLU**	125	(3, 3)	32
Layer 9	**Convolution + ReLU**	256	(3, 3)	32
Layer 10	**Transpose Conv.**	3	(2, 2)	64
Layer 11	**Tanh**	3	-	64

(b) Parameters

Lambda = 10 ** -5

LR = 0.0001

Optimizer : Adam

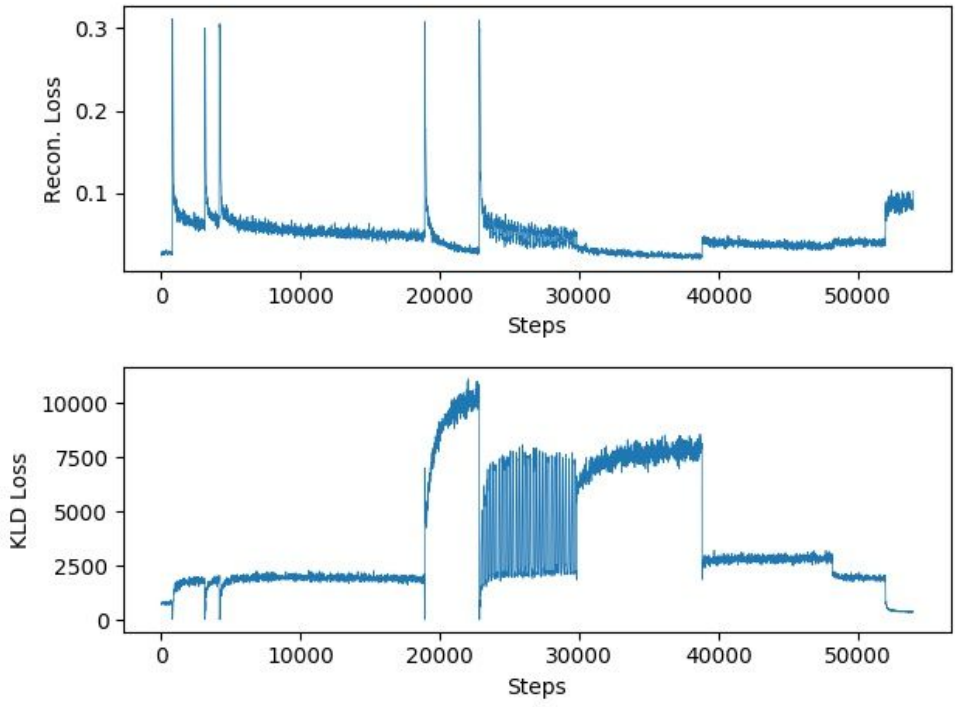
Loss function (Recon. loss) : MSE loss

(c) Preprocess

讀進圖像後，先將圖像pixel value 等比縮到 (0,1)得到x，再進行簡單的normalization至(-1, 1)，即x' = (x-0.5) * 2。

而在decode model最後輸出的image pixel value則通過Tanh activation轉換到(-1, 1)之間並與已經進行過上述normalization的real image運算loss。

2. Learning Curve



此處的KLD Loss是未乘上lambda值的原始KLD Loss。

注意Reconstruction loss有明顯凸起的部份是緣於model中斷training後，繼續training造成的結果，因為Adam optimizer與過去已更新的gradient有關，因此重新resume training會造成起步的gradient較為不穩而造成loss凸起，然而並不會造成整體training上的問題。

另外注意在step為18000、22000、30000、38000、48000與52000處，KLD都有明顯的改變（reconstruction loss亦有對應的改變），原因是我在training過程中試圖改變lambda值來察看training會有如何對應的quality。

以下附註step所改變的lambda值：

step	0	18000	22000	30000	38000	48000	52000
lambda	10 ** -5	10 ** -6	10 ** -5 10 ** -6 交替	5 ** -6	9 ** -6	10 ** -5	10 ** -4

由上表可以明顯驗證當lambda值大時KLD loss會變小，反之亦然，而recon. loss則反向變動，但recon. loss變動的scale卻非常小，原因是output的range本來就因Sigmoid activation限在（0, 1）區間。而調大lambda時確實可以明顯看出recon. image更為清晰，驗證model的數學理論。

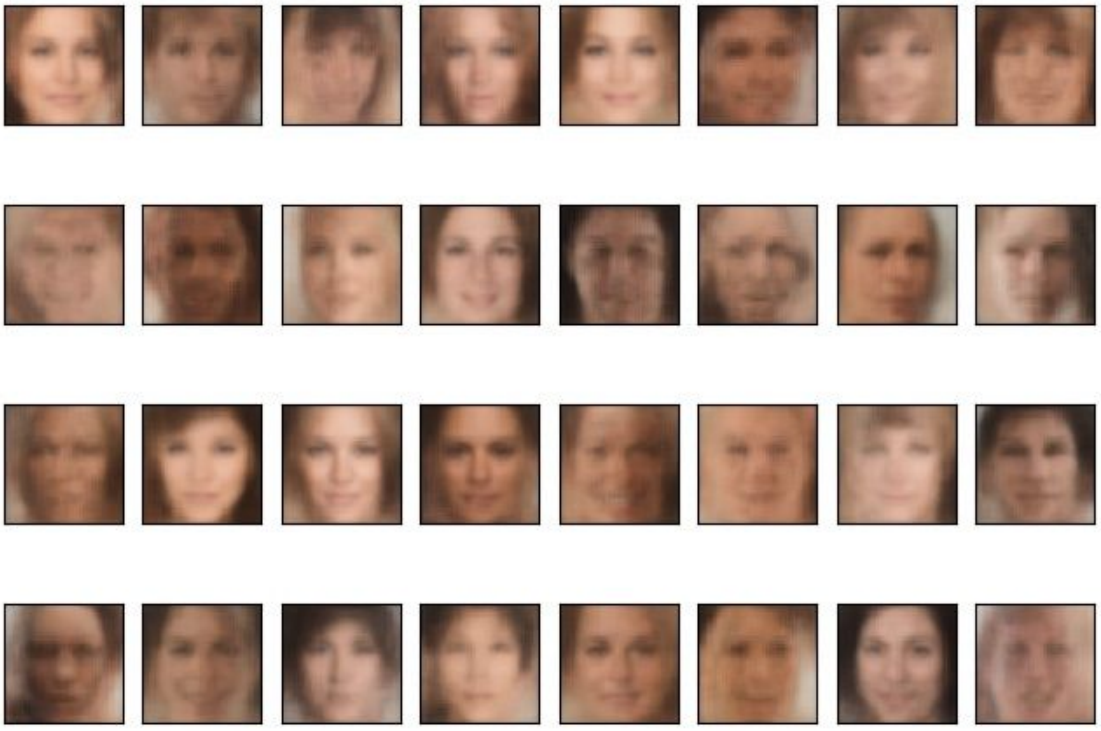
3. Reconstruction image of test images

test images : from 40000.png ~ 40010.png

Reconstruction loss of MSE

Averaged : 0.06824938921024311

4. Random Generation

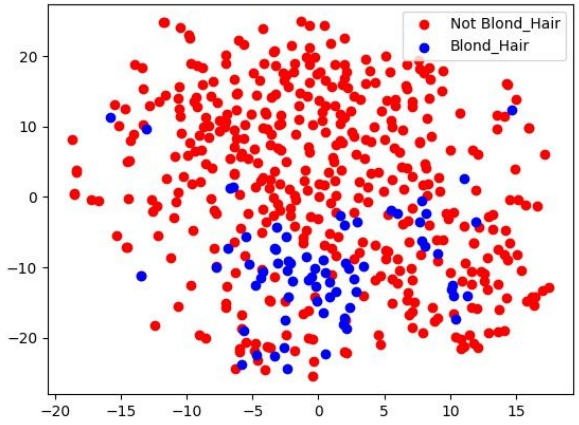


5. Visualization with t-SNE

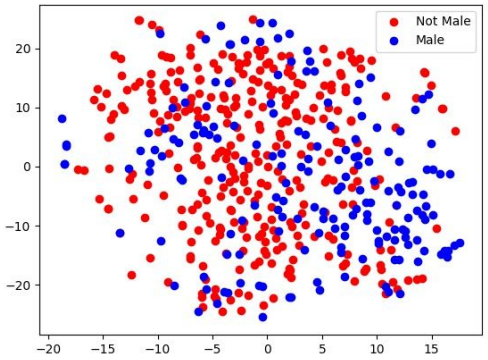
將latent space的數值透過t-SNE轉換，並取出轉換後最前的兩個axis作圖，其中選取500張test images做分析。

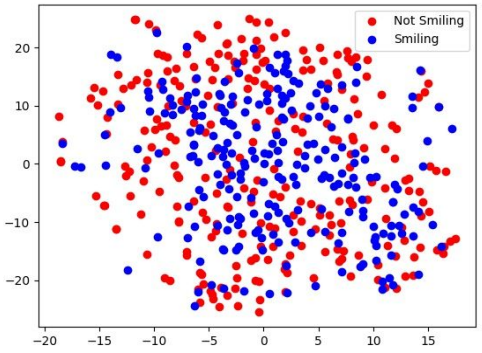
以下挑選「是否為金髮」作為兩項lable標色。

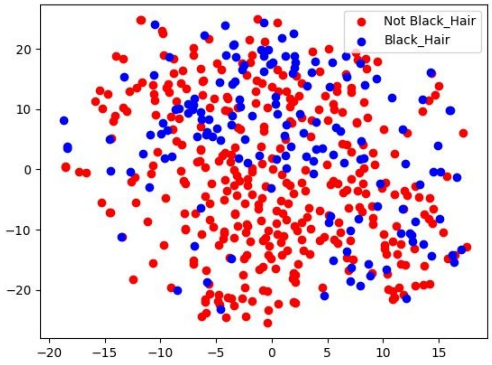
可以觀察到金髮的label在兩個維度中幾乎分佈在第三、四象限，且非金髮的label雖然4個象限皆有分佈，但明顯在第三象限分佈較少，可觀察到在此兩維度中金髮與非金髮的數值分佈是具有一定分離性的。



為了輔佐分離性的判別，以下附上其他特徵做圖，包含「是否為男性」、「是否有笑容」以及「是否為黑髮」。都可看出其在此二維度的分離性都沒有「是否為金髮」來得高（其中是否為黑髮的分離性也相當高）。







6. Discussion

以下討論中用的model架構請參考本大題第一小題model：

(1) 在Layer5添加tanh對latent space做約束：

在輸出latent space前先將value通過tanh activation做數值約束，可以使訓練更加快速的收斂，loss降低的速度很快，同時也能在random generation時對丟入的random latent space vector作較為狹窄的範圍限制，可避免有過大的數值而造成image distortion。不過就model來說，這樣的作法很有可能會造成back propagation的gradient對layer 5以前的layer產生的效果有效，約束的training的強度，並可能造成產生的圖像模糊與相似。

(2) Lambda驗證：

VAE的理論已經告訴我們，調整lambda值是在reconstruction和generation quality兩者間的trade-off。實際調整lambda值確實能夠驗證，lambda大時reconstruction效果差但generate效果佳（如第2小題所述）。

(3) Batch size：

VAE在訓練時以batch size=8, 16, 32, 64, 128去訓練，最後經過20epochs後，產出的reconstruction image和random generate幾乎相似，並沒有明顯的差異。

(4) Preprocess：

data有進行preprocess步驟的訓練結果效果較佳，但相對整體圖像對比度比沒有做preprocess的訓練結果來的低，呈現柔和感。

(5) Batch Normalization：

在此model架構下，對convolution layers加上B.N.，結果發現更加的模糊與柔和，推測是latent space已經被tanh現在小範圍內，又加上B.N.，對數據做抑制的行為，而在forward的過程中降低數值的變異，故最後出來的圖片會更為平均與模糊。

(6) Drop-out：

VAE的decode network若較大（參數調大），加上drop-out確實可以避免overfitting的情況，不過效果仍有限，加上最後我選擇的decode network參數小，因此完全沒有使用drop out layers。

Problem 2 GAN

1. Model

(a) Implementation

		Output Channels	Kernel Size	Feat. Map size
Gen. Net				
Layer 1	**Transpose Conv.**	64	(32, 32)	32
Layer 2	**Convolution + ReLU**	128	(3, 3)	32
Layer 3	**Convolution + ReLU**	256	(3, 3)	32
Layer 4	**Transpose Conv.**	3	(2, 2)	64
Layer 5	**Sigmoid**	3	-	-
Dis. Net				
Layer 1	**Convolution + ReLU**	64	(3, 3)	1
Layer 2	**Convolution + ReLU**	128	(3, 3)	32
Layer 3	**Pooling**	128	(3, 3)	32
Layer 4	**F. C.**	32 * 32 * 128	-	-
Layer 5	**F. C.**	1024	-	-
Layer 6-1 Real/Fake	**F. C.**	1	-	-
Layer 6-1 Attribute	**F. C.**	1	-	-
Layer 7	**Sigmoid**	1	-	-

(b) Parameters

LR = 0.0001

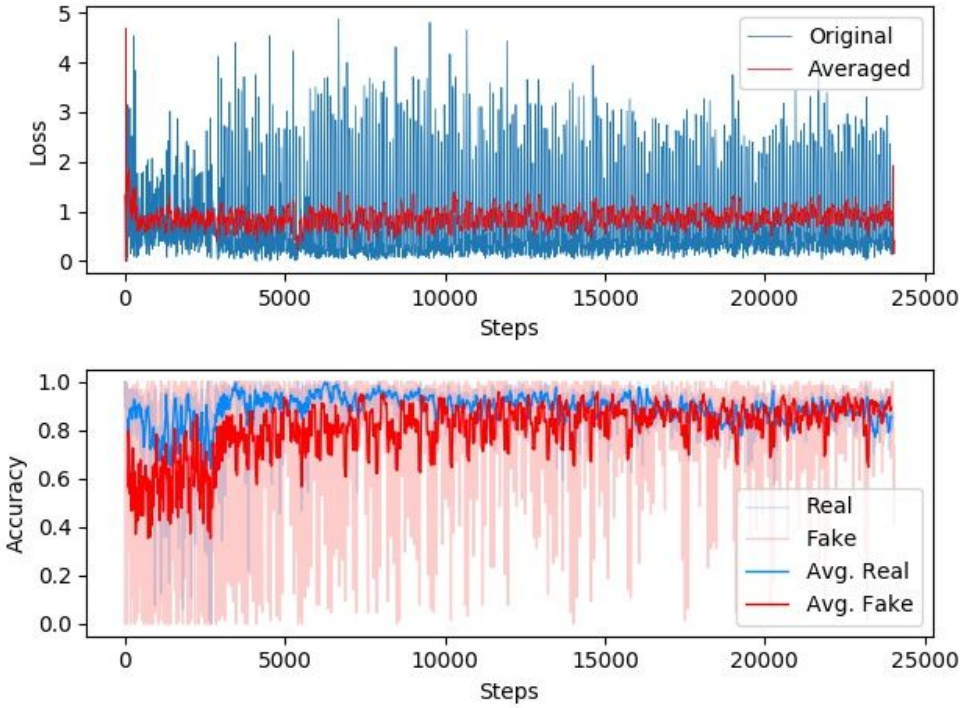
Optimizer : Adam

Loss function (for both) : BCE loss

Steps for one model :

[init] D. net : G. net = 1 : 1
[final] D. net : G. net = 3 : 1

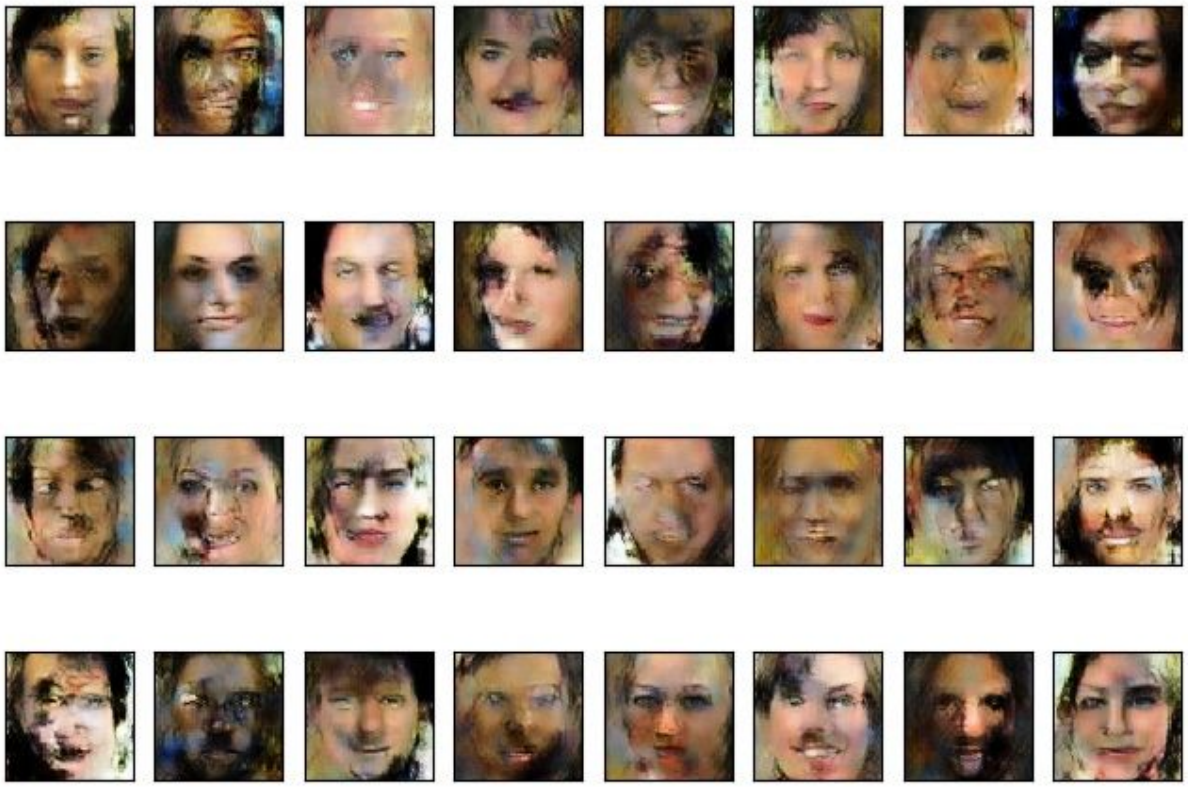
2. Learning Curve



第一張圖是Loss對steps作圖，其中loss算入training Discriminator和training Generator過程中的loss，因此看藍色的線（原始紀錄點）很明顯有來回高低高低振盪的情形，高點是對應訓練Discriminator的時間點，低點是對應訓練Generator的時間點。整體而言，由藍色的紀錄點可看出後期（step>15000）時，最高點比前期的低，還有降低的趨勢，而最低點則有微微攀升的趨勢，可得知Discriminator的loss和Generator的loss彼此靠攏，往一個平衡點靠近，代表是正確收斂的過程。此外紅色線是取10的藍色原始紀錄點做平均所得到平均趨勢，整體而言並沒有太大浮動，可代表這個訓練預期的平衡點（但不代表會到達該平衡點）。

第二張圖是針對training Discriminator時紀錄的accuracy，其中藍色紀錄real images所對應的accuracy，紅色紀錄fake images所對應的accuracy，淺色代表真實紀錄點，深色代表平均趨勢（以十個真實數據做平均）。可明顯看出紅色針對fake images的accuracy是平均上升的，代表我們產生的fake images愈來愈真實，而藍色的線大致上平穩，唯有些微下降，代表Discriminator為了區辨fake image，而有點無法完全掌握real images的特徵。

3. Random Generation



4. Discussion

以下討論中用的model架構請參考本大題第一小題model：

(1) Mode collapse：

原先的generation model為在Layer 1前再插入一層F.C. layer，但發生generated image quality確實變好，卻幾乎長的一樣的mode collapse。將該F.C. layer拔去後mode collapse的情況就消失了。可能是在F.C. layer中產生dominat的network而造成後續的forward全部bias。

(2) No early stopping：

在訓練GAN時，除了用肉眼觀看每次產出的image quality外幾乎沒有其他辦法能夠判斷何時該停止training。尤其到訓練中後期(15~30epoch)，image quality產生像是振盪的情形，一下好一下子稍差（但不會整個爛掉）。查詢資料後發現這是非常正常的情況，因為可能是遇到loss的saddle point，而造成類似振盪的情況，此時除了重新加大lr讓它離開local minimum外，就是要重train或是要改model參數。

(3) Training steps ratio：

發現training GAN中，調配training Gen. Net和 Dis. Net的訓練step比例與大小是非常重要的！

經過多次測試，個別的training step落在1~30steps較佳，且比例為1:1到3:1較優（Dis. Net: Gen. Net）。另外，在訓練初期由於Gen. Net非常弱，所以可以把訓練Gen. Net的steps ratio調較高（1:1），但在中期Dis. Net 判別的accuracy開始振盪時，可以將訓練Dis. Net的steps ratio調高（3:1~4:1），如此可避免train壞。

(4) Drop-out ：

根據這次的training經驗，是否有加上drop-out並不太會影響訓練結果，其他因素與參數與運氣幾乎決定一切。

(5) Batch size：

在training的過程中發現batch size調大會有明顯的穩定效果，產生的圖像穩定度比較高（局部爛掉的情況比較少），應該是batch帶來較為穩定的gradient使training的方向較為穩定。

(6) Transpose Convolution：

該layers的bias應該要儘量拿掉，由於算法的關係，若bias有dominant的情況整個整個network產生非常大的影響，所以儘量不要加bias。

(7) 如何判定model是否work：

每搭好一次model或調好參數進行training時，初期非常難判斷是否有work。以下紀錄幾個自己掌握到的pattern。

最多3~5的epoch產出的image還使沒有人臉的話幾乎就是不會work。

lr非常重要，以我的model深度為例，若深度在此附近，則lr約在0.0005~0.000001是會work的，太大的話會完全沒有在training。

印出loss，以我們loss算法為例，若Loss出現0或是超過10，幾乎都是不會work的，要調整lr，甚至是調整model與參數。

(8) Transfer from VAE：

從VAE transfe共同的架構參數，不過要避免偷吃步.....（例如其實幾乎沒調VAE太多就停下來）。為了避免偷吃步的判定，所以最後重頭train而不是用transfer的方式。

5. Comparison

GAN VAE

從GAN random generate出的image比起VAErandom generate出的image來得色塊清晰、對比度高，但image周邊處理以其左右對稱度就較差。整體而言，GAN的

images比較「大膽」、「塊狀」與「色彩鮮豔」，而VAE的images較為「保守」、「圓滑」與「柔色」。

色塊清晰、對比度高應該是由於GAN在training過程中並不像VAE中有KLD去約束sample的分佈，再加上VAE的圖像loss是有與原圖計算reconstruction loss的，在設計上已經考慮與原圖的差異，且是以「pixel」的方式去算loss，所以圖像叫不會有色塊差異，但GAN僅僅是交由Discriminator去做判斷是否為真實圖片，因此無法確保Discriminator究竟學習到什麼樣的features做為判斷，而且不是以pixel為loss計算，故產生色塊分佈大致是合理的。

GAN image周邊處理與左右對稱度的問題也是由於GAN的Discriminator僅針對是否為真實圖片做學習，因此很難保證它會學習到左右是否對稱這個feature，雖然在convolution的設計上是有考慮到feature maps上的特徵相對位置的分佈，不過在train過程中仍然很有可能不會再初期學習到這個較為高層次的判別。

Problem 3 ACGAN

1. Model

(a) Implementation

		Output Channels	Kernel Size	Feat. Map size
Gen. Net				
Layer 1	**F.C.**	512	-	-
	Reshape	512	-	1
Layer 2	**Transpose Conv.**	64	(2, 2)	32
Layer 3	**Convolution + ReLU**	128	(3, 3)	32
Layer 4	**TConvolution + ReLU**	512	(3, 3)	32
	Transpose Conv.	3	(2, 2)	64
Layer 5	**Tanh**	3	-	-
Dis. Net				
Layer 1	**Convolution + ReLU**	64	(3, 3)	64
Layer 2	**Convolution + ReLU**	128	(3, 3)	64
Layer 3	**Pooling**	128	(3, 3)	32
Layer 4	**Convolution + ReLU**	256	(3, 3)	32
Layer 4	**Convolution + ReLU**	512	(3, 3)	32
Layer 5	**Pooling**	128	(3, 3)	16
	Reshape	16 * 16 * 512	-	-
Layer 6	**F. C.**	1024	-	-
Layer 7-1 Real/Fake	**F. C.**	1	-	-
Layer 7-2	**F. C.**	1	-	-

Attribute				
Layer 8-1	**Sigmoid**	1	-	-
Layer 8-2	**Sigmoid**	1	-	-

(b) Parameters

LR = 0.0001

Optimizer : Adam

Loss function (for both) : BCE loss

Steps for one model :

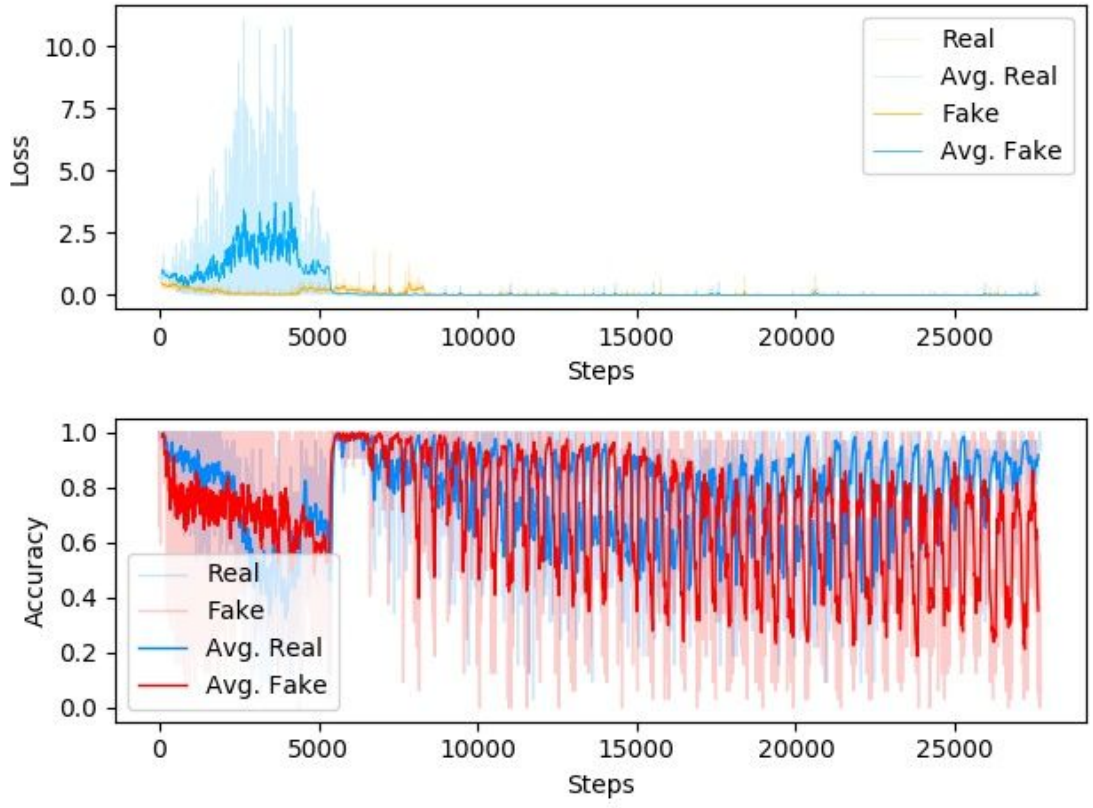
[init] D. net : G. net = 1 : 1

[final] D. net : G. net = 4 : 1

(c) Preprocess

讀進圖像後，先將圖像pixel value 等比縮到 (0,1)得到x，再進行簡單的normalization至(-1, 1)，即x' = (x-0.5) * 2。

2. Learning Curve



第一張圖中紀錄Attribute Classifier部份的loss，其中約在step<5000時的training steps ratio 是D net: G net = 10 : 10，這時的訓練狀況不穩定使得loss往上升

，此時我將training steps ratio調整為D net: G net = 20 : 10，訓練狀況變穩定了，loss也幾乎維持在非常低的水準，約在10*-1~10*-2數量級。

第二張圖紀錄Discriminator判斷圖片是real/fake的準確度。同樣的約在step<5000時的training steps ratio調整不佳，故accuracy接往下掉。調整完後訓練變得穩定且可以注意到，對於real image的accuracy(藍線)始終維持相當高的水平，唯有在中期出現微微下降的情形，因為中期在real/fake images 的判別中調整平衡點，整體呈現向上彎的曲線。另一方面，對於fake image accuracy則是出現微微下降的趨勢，代表我們產生的圖片漸漸能夠fool Discriminator。

3. Random Generation