

MLDS: HW4

January 7, 2018

Professor Hung-Yi Lee

電機四 B03901018 楊程皓

Model Description

我的架構基本上以 dcgan 為主，除了 Optimizer 使用 RMSPropOptimizer 外其他都跟論文差不多。另外也有實作 wgan，但因產出的圖不比 dcgan 好，最後就以 dcgan 為繳交的 model。

我的 generator 吃進由 skip thoughts 產生的 dimension 為 2400 的 vector 與 dimension 250 random noise，中間架構基本上跟論文方式一致，先做 dimension reduction，將 skip thoughts 產出的 dim: 2400 過一個 fully connected layer，將 dimension 縮減為 256，concatenate noise 後為 dim: 506，再接 fully connected 後轉為 $4*4*512$ 。最後經過四輪的 deconvolution 分別轉為 $8*8*256$, $16*16*128$, $32*32*64$, $64*64*3$ 作為最後的 image output。最後一層的 activation function 是 tanh。

我的 discriminator 吃進一個 $64*64*3$ image 及 skip thoughts 產出的 dim: 2400 vector。對 image 做在 generator 中相反的事情，經過四輪 convolution 分別轉為 $32*32*64$, $16*16*128$, $8*8*256$, $4*4*512$ ，對 skip thoughts vector 則做一樣的 dimension reduction 轉為 dim: 256，接上 image convolution 後的 $4*4$ 後過一個 convolution 轉為 $4*4$ 後最後做一個 fully connected layer 轉為 dim: 1 的 output，代表 discriminator 認為這個 image 是不是真的且對上正確的 text caption。最後一層的 activation function 是 sigmoid。

Generator 的 objective function (loss) 為 cross entropy of output of discriminator(generated images, right captions) and ones (output 經過 sigmoid activation function)，希望讓 discriminator 誤認 generated images 為真實 images。

Discriminator 的 objective function (loss) 有三項: cross entropy of output of discriminator(real images, right captions) and ones, cross entropy of output of discriminator(generated images, right captions) and zeros, cross entropy of output of discriminator(real images, wrong captions)，希望能辨認對的 image match 對的 caption，其餘都辨認為錯誤的配對。

Performance Improvement

我有修改 noise dimension (100, 120, 250)，但影響不大，並且因我最後用的 dcgan 並不會太被 noise dimension 所影響（產出的五張圖差異非常少），但我發現 epochs 之間的產圖差異不小，詢問過弘毅教授可以使用 ensemble 的方式，將五個不同的 epochs 各產五張圖作為最終結果。

另外論文上的 dimension reduction 是從一千多維降為 128 維，由於我們的 skip thoughts 產出的是 2400 維，故我最後使用的 dimension reduction 做的就是降為 256 維再與 noise 合併。

在 tags 的選擇方面，我首先把沒有包括 hair, eyes 的所有 instance 都刪掉，並在有多個 hair or 多個 eyes 的 instance 只保留其中分數（數量）最高的那個 feature，故每個 instance 最多有一個 hair & 一個 eyes 的顏色描述，以避免 skip thoughts 產生不太合理的 vectors。

另外我也有實作的 wgan , 設定都跟 slides 一樣 : no sigmoid on discriminator, clip weight -0.01 - 0.01, RMSProp, discriminator update : generator update = 5:1 。但產生的圖比 dcgan 模糊很多, 雖然他會被 noise 影響而對同個 caption 產生出不同的圖。

Experiment Settings and Observations

dcgan model 架構如上, discriminator update : generator update = 1:2, Optimizer learning rate = 0.0001, batch size = 64, image size down size 96 \rightarrow 64 (因最後產的圖也是要 64*64*3), epoch: 1800, wrong caption 挑選方式為其他 caption random 出一個與當前 right caption 不一樣的。

在實驗中觀察到 dcgan 看來不太容易被 noise 影響產出圖的特性, 只有一些 pixel 深淺值會被稍微改變, 然而 wgan 並沒有這個問題, 產出的圖還是有許多不一樣的地方。另外 dcgan 的 epoch 之間變化其實很大, 讓我有點懷疑他其實是在不同的 training instances 之間震盪。

下面的圖們即為 dcgan 最後產出的圖。雖然還略顯模糊, 不過已看得出來屬於動漫人物。



下面的則是我的 wgan 產出的圖, 清晰度比之 dcgan 仍有一段差距。

