

MLDS HW3 Report

環境

- Debian Linux, kernel version 3.16.39
- CPU: Intel® Core™ i7-2600K CPU @ 3.40GHz 8 core
- GPU: GeForce GTX 980, 6GB memory
- tensorflow r1.0
- Keras 2.0.4

模型

DCGAN

- Generator
 1. 用「linear+Irelu」把文字訊息投影到較小的維度
 2. 把「text embedding」和「noise」接在一起
 3. 「concatenate」好之後，進一層「linear layer」
 4. 先「reshape」成 (batch_size, img_col / 16, img_row / 16, filters)，然後加上「batch normalization」和「relu」
 5. 過三層「deconv2d」，把圖片解析度提高、「channel」數量壓低，每層都使用「batch normalization」和「relu」
 6. 進最後一次的「deconv2d」，並套上「sigmoid」當激活函數，得到「generator」最終的輸出圖片
- Discriminator
 1. 和「generator」一樣，先用「linear+Irelu」把文字訊息投影到較小的維度
 2. 對輸入的圖片，先過一層「conv2d」，並使用「Irelu」激活
 3. 過三層「conv2d」，把圖片解析度壓低、「channel」數量提高，每層都使用「batch normalization」和「Irelu」
 4. 把「text embedding」和卷積層最後的輸出接再一起
 5. 再進一次「conv2d」，通樣使用「batch normalization」和「Irelu」
 6. 壓平，接著做一層的「logistic regression」，得到代表圖片真實度的「scalar」
- Loss
 - 三種「Cross Entropy」的總和
 - 對「Discriminator」來說
 - 級正確敘述、正確圖片，應該要是1
 - 級錯誤敘述、正確圖片，則應該要是0

- 紿正確敘述、「Generator」自己生的假圖片，也應該要是0
- 對「Generator」來說
 - 紿正確敘述、「Generator」自己生的假圖片，希望它也能是1

WGAN

- 「Generator」和「Discriminator」並沒有太大的差別，除了「Discriminator」的輸出不在套上「sigmoid」，而是直接輸出數值
- Loss
 - 對「Discriminator」來說
 - 紿正確敘述、正確圖片，輸出數值應該要越大越好
 - 紿錯誤敘述、正確圖片，輸出數值應該要越小越好
 - 紿正確敘述、「Generator」自己生的假圖片，輸出數值也是越小越好
 - 對「Generator」來說
 - 紿正確敘述、「Generator」自己生的假圖片，希望它的輸出數值也能變大
- Update
 - 可以使用「weight clipping」(`tf.clip_by_value`)
 - 或者透過「gradient penalty」，即「regularization term」來限制

```

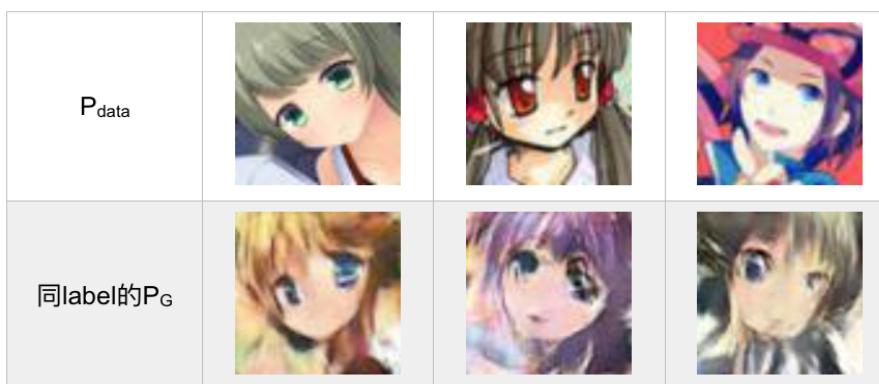
1 epsilon = tf.random_uniform([], 0.0, 1.0)
2 x_hat = epsilon * t_real_image + (1 - epsilon) * fake_image
3 _, d_hat = self.discriminator(x_hat, t_real_caption, reuse=True)
4 ddx = tf.gradients(d_hat, x_hat)[0]
5 ddx = tf.sqrt(tf.reduce_sum(tf.square(ddx), axis=1))
6 ddx = tf.reduce_mean(tf.square(ddx - 1.0) * 10)
7 d_loss = d_loss + ddx

```

如何進步

Text Representation

- 原本使用「skipthoughts」套件，把文字訊息先轉成4800維的向量
- 觀察「Generator」畫的圖，發現顏色和文字幾乎都搭配不起來，如下圖



- 後來改成針對眼睛、頭髮顏色，直接把文字輸入轉成27維的「one-hot representation」，顏色的正確率才提高很多

P_{data}			
同label的 P_G			

DCGAN collapse

- 一開始使用DCGAN做training，在中間的epoch的時候拿出來看，雖然頭髮顏色有對上，但是臉型都是同一張臉，很明顯，發生collapse mode，因此加入wgan(weight clipping, gradient penalty)做嘗試，參考圖在實驗設計與觀察。

實驗設計與觀察

DCGAN

- update次數比例(discriminator:generator = 1:1)
- 可以看到頭髮雖然偏彩色但是還是看得出caption的顏色，眼睛顏色大致對得起來，但五官並不明顯，臉型常常扭曲

green hair blue eyes					
white hair green eyes					
orange hair green eyes					
orange hair red eyes					

WGAN

- weight clipping [update次數比例(discriminator:generator = 5:1)]

- 五官依舊扭曲，且頭髮顏色近乎一樣，兩眼顏色偶爾不對稱，可能參數調不夠好

green hair blue eyes					
white hair green eyes					
orange hair green eyes					
orange hair red eyes					

- gradient penalty [update次數比例(discriminator:generator = 5:1)]
- 很明顯爆炸，猜測原因可能penalty太嚴重

P_{data}			
同label的 P_G			

分工

- 李承軒：Model, Report
- 簡瑋德：Model, Report
- 劉岳承：Experiment, Report
- 黃兆緯：Data preprocessing, Report