

# Hw3 report - Anime Faces Generator

R05943135 杯屎賴-

R05943135 江承恩 R05943011 沈恩禾  
R05921016 傅均笙 R05942072 吳昭霆

## 環境

CPU	GPU	Memory	OS	Libraries
Intel(R) Xeon(R) CPU E5-2630 v2 @ 2.60GHz	GTX 980	HYNIX HMT42GR7BFR4C- RD MEMORY 16GB * 8	Ubuntu 15.04 Mint 17.1 Rebecca	Tensorflow 1.0 Skip Thought Theano scipy skimage

## Model 描述

我們使用 **Skip Thought vector**<sup>3</sup> 處理過後之文字描述與 noise 作為輸入丟入 DCGAN<sup>1</sup> 之中。並用 GAN-CLS<sup>2</sup> 之 objective function 作為 loss 訓練。

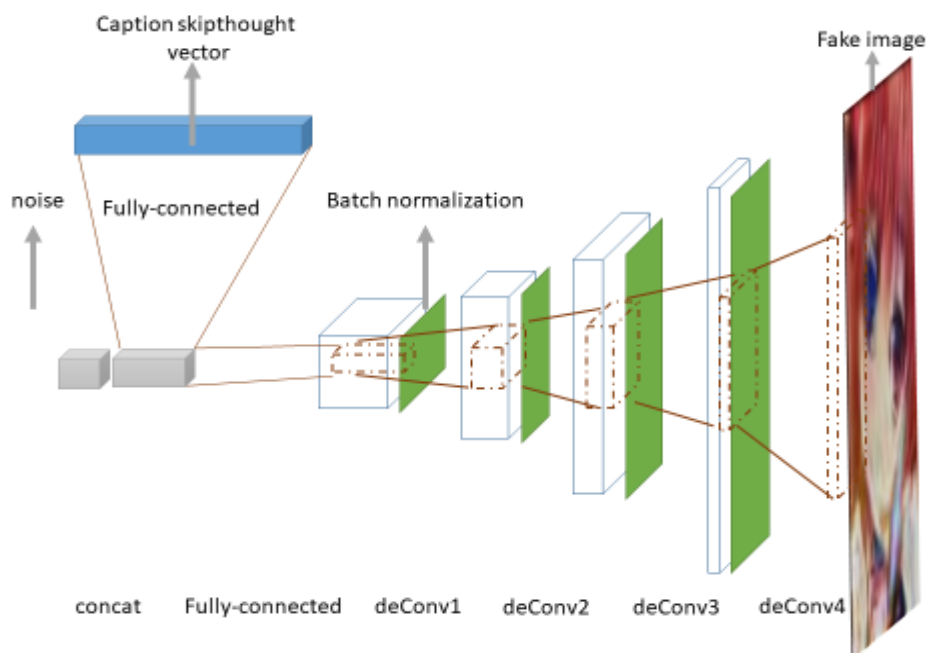


Fig. Generator

- SkipThought 向量先經過 fully-connected 擷取出 256 維 embedding。
- Embedding 與 100 維 Noise 相接，經過 fully-connected 輸出  $4*4*512$  維 activation。
- 三組  $5*5$ , stride  $2*2$  之 transpose convolution，長寬每層增倍，channel 數每層折半。
- 最後通過 transpose convolution 輸出  $64*64$  長寬之三色圖片，channel 數為 3。

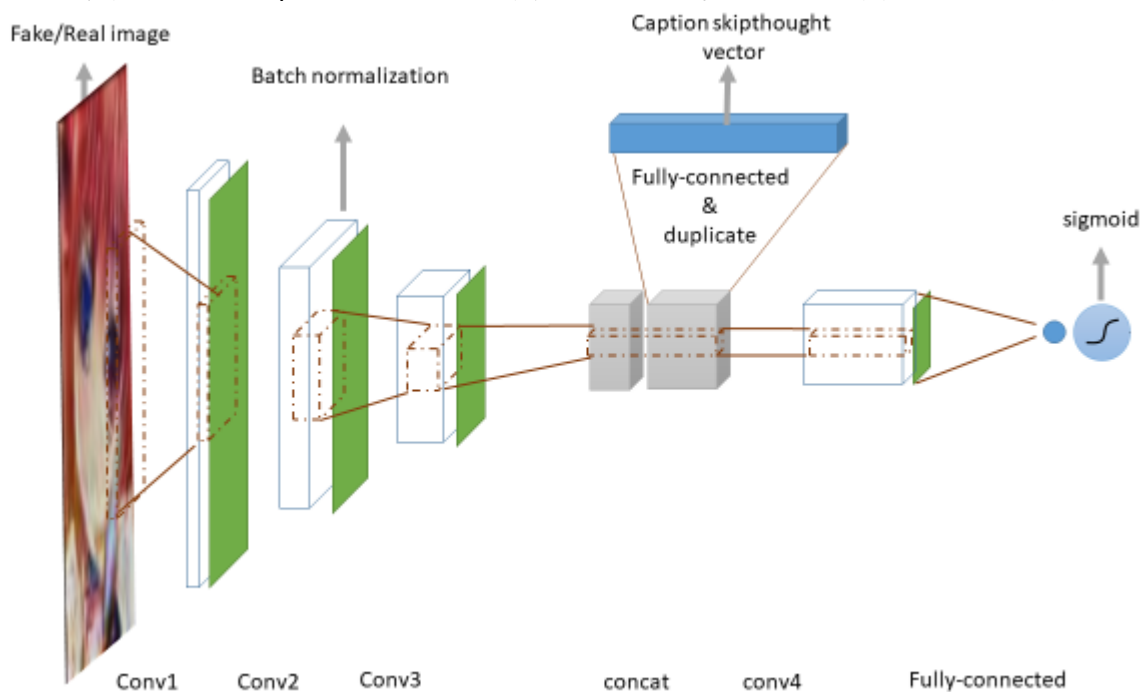


Fig. Discriminator

- Discriminator 為一 deep convolution binary classifier。
  - 通過三組  $5*5$ , stride  $2*2$  之 convolution，長寬每層折半，channel 數由 128 起每層加倍。
  - Caption 之 skipthought 向量先經過 fully-connected 擷取出 256 維 embedding。
  - 將 caption embedding 延展為  $4*4*256$  維。
  - 將 embedding 與 convolution 之  $4*4*512$  activation 相接。
  - 通過  $1*1$  之 convolution。
  - 通過 fully-connected layer，輸出一維 logit，通過 sigmoid。
- Batch normalization 使用 decay 0.9。

## GAN-CLS training

Conditional GAN 必須額外處理 caption，除了產生合理的圖片以外，圖片必須符合 caption。因此訓練時 Discriminator 須考慮三種情況的 cross entropy，令其 optimizer 這三組 Loss 之總和，三個 Loss weighting 相同。Generator 則以 discriminator 輸出之 cross entropy 為 Loss。

- Discriminator :
  - real image loss :輸入真實圖片與文字描述為 positive example(label 為 1)。
  - wrong image loss:輸入與文字敘述不合之真實圖片為 negative example(label 為 0)。
  - fake image loss :輸入 Generator 產生之圖片與文字敘述作為 negative example。

*minimize*

$$-\log(D(x: \text{real image}, t: \text{real text})) - \log(1 - D(x: G(x: \text{real text}), t: \text{real text})) \\ - \log(1 - D(x: \text{real image}, t: \text{irrelevant text}))$$

- Generator loss : 將 generator 產生之假圖做為 positive example。

*minimize*

$$-\log(D(x: G(x: \text{real text}), t: \text{real text}))$$

## Training Data 處理

Training caption 僅擷取 tag 中出現 hair, eyes 者，並實驗額外擷取 blush, ponytail, braid。因為資料僅切割人臉一小部分，並且有 tag 錯誤之樣本(原始圖片有多個人物，切割出錯誤 tag)，因此額外 tag 之實驗結果不理想。兩者各有樣本 8k 與 18k 組資料。

- Caption 皆先通過 Skip Thought 產生 4800 維向量作為 input。
- Noise 使用 uniform distribution [-1,1]。
- image 使用 skimage.transform.resize()調整解析度至 64\*64 (無作用，原圖為 64\*64)。
- 1/2 機率將 training image 左右翻轉。

## Training 參數

使用 Adam optimizer。

- batch size:64
- learning rate:0.0002
- beta1:0.5
- epoch: 300, 350, 400, 600





## 實驗

A:在處理 wrong image 時有出錯，有機會 wrong image 與 caption 是對應的，這組實驗同時使用 braid, blush, ponytail 三組額外 tag。

B:僅擷取 hair, eyes 兩種 tag，若 tag 顏色不同的話，才可以作為 wrong image 使用。觀察相近 epoch 數產生圖片結果。

- A : noise 毫無效果，僅能產生簡單 caption 到 image mapping 的結果。



### ■ 300 epoch

		blue hair blue eyes blue hair green eyes	
		blush blue hair blue eyes & blush blue hair red eyes	 ponytail green hair blue eyes

### ■ 400 epoch

	blue hair blue eyes blue hair green eyes		blue hair red eyes & green hair blue eyes
	blush blue hair blue eyes & blush blue hair red eyes		short hair blue eyes

### ■ 500 epoch

	Blue hair blue eyes blue hair green eyes blush blue hair blue eyes		blue hair red eyes & green hair blue eyes
		Short hair blue eyes Long hair blue eyes	

- B:使用 epoch 596 ~ 600 五個 model 分別產生一張圖片。

	blue hair blue eyes
	blue hair green eyes
	blue hair red eyes
	green hair blue eyes

## 分析

- Noise 在實驗 A 無法產生顯著效果，若調整 noise 範圍，僅改變色澤深淺。
- Epoch 過高可觀察到 D Loss 下降到 0.001 尺度，但 generator 仍在 5~10 範圍中震盪。表示 Discriminator 過強，此時產生之圖非常保守，只能產生一種臉型。
- Epoch 過低會產生不合理之臉型。300~400 epoch 在 A 實驗為佳。
- 猜測使用 WGAN 可以改善單純 mapping 的問題，我們的 model 似乎收斂到某個資料點，而非一個 distribution。
- WGAN 實作將 Model 最後一層 output sigmoid 去除，將 weight 限制在-0.01~0.01 之間，使用 WGAN object function，但尚未訓練夠多 epoch 產生合理結果。

## 分工

沈恩禾	江承恩	傅均笙	吳昭霆
實驗設計, GAN-CLS	DCGAN	SkipThought	WGAN 嘗試

## References

- 1 **Tensorflow DCGAN implementation:** <https://github.com/carpedm20/DCGAN-tensorflow>
- 2 *Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee*, **Generative Adversarial Text to Image Synthesis**, arXiv:1605.05396
- 3 **Skip Thought implementation:** <https://github.com/ryankiros/skip-thoughts>