

# 921 U2620 HW8

TOTAL POINTS

**4 / 4**

## QUESTION 1

1 Make a brief introduction about variational autoencoder (VAE). List one advantage comparing with vanilla autoencoder and one problem of VAE. 2

/ 2

- 1 pts Introduction: major mistake(s) / unclarity

- 0.5 pts Introduction: minor mistake(s) /

unclarity

- 0.5 pts Advantage: mistake(s) / unclear

- 0.5 pts Problem: mistake(s) / unclear

✓ - 0 pts Good!

✓ + 0.5 pts Describe the differences based on your reconstructed images (if any).

+ 0 pts Major mistakes / No answer

## QUESTION 2

2 Train a fully connected autoencoder and adjust at least two different element of the latent representation. Show your model architecture, plot out the original image, the reconstructed images for each adjustment and describe the differences. 2 / 2

✓ + 0.5 pts Show your model architecture.

✓ + 0.5 pts Show your two adjustments on the latent representation.

✓ + 0.5 pts Show the original image and your two reconstructed images based on your adjustments.

# 1.

簡單來說 VAE 是從 AutoEncoder 改進過來的，Encoder 產生出 2 組 latent representation，並且從 normal distribution 中 sample 出 noise，把其中一組 latent representation 做 exponential 後和 noise 相成之後再與另一組 latent representation 相加後再餵到 decoder 做還原。

原始的 AutoEncoder 的 latent representation 和 reconstruction image 比較像是一對一的 mapping，若是給的 latent representation 沒有 mapping 到對應的結果，則 reconstruction 會是四不像，而 VAE 則是使 latent representation 加了一些 noise(latent representation 附近) 後都能 mapping 到最後相同的 reconstruction。在不同圖片的 latent representation 的附近範圍會有重疊的部分，但 reconstruction 只會有一種結果，因此可能可以得到介於不同圖片的特徵，使 reconstruction 的結果更為真實。

VAE 的一個缺點就是他沒有學習怎麼樣去產生出新圖片，而是著重在如何產生出跟 database 越相近的圖片，所以以圖片來說，若是相差了一個 pixel 的 reconstruction 圖片，無論這個 pixel 是在哪裡，model 都會認為他們是一樣差的，但對於人類來說，某些 reconstruct 的圖可能即使差了一個 pixel 我們也還是會認為他是真的圖而不是假的圖。

# 2.

Model 架構:

Layer (type)	Output Shape	Param #
Linear-1	[-1, 4096]	50,335,744
BatchNorm1d-2	[-1, 4096]	8,192
ReLU-3	[-1, 4096]	0
Linear-4	[-1, 2048]	8,390,656
BatchNorm1d-5	[-1, 2048]	4,096
ReLU-6	[-1, 2048]	0
Linear-7	[-1, 4096]	8,392,704
BatchNorm1d-8	[-1, 4096]	8,192
ReLU-9	[-1, 4096]	0
Linear-10	[-1, 12288]	50,343,936
Tanh-11	[-1, 12288]	0
Total params: 117,483,520		
Trainable params: 117,483,520		
Non-trainable params: 0		
Input size (MB): 0.05		
Forward/backward pass size (MB): 0.42		
Params size (MB): 448.16		
Estimated Total Size (MB): 448.63		

1 Make a brief introduction about variational autoencoder (VAE). List one advantage comparing with vanilla autoencoder and one problem of VAE. 2 / 2

- 1 pts Introduction: major mistake(s) / unclarity
- 0.5 pts Introduction: minor mistake(s) / unclarity
- 0.5 pts Advantage: mistake(s) / unclear
- 0.5 pts Problem: mistake(s) / unclear

✓ - 0 pts *Good!*

# 1.

簡單來說 VAE 是從 AutoEncoder 改進過來的，Encoder 產生出 2 組 latent representation，並且從 normal distribution 中 sample 出 noise，把其中一組 latent representation 做 exponential 後和 noise 相成之後再與另一組 latent representation 相加後再餵到 decoder 做還原。

原始的 AutoEncoder 的 latent representation 和 reconstruction image 比較像是一對一的 mapping，若是給的 latent representation 沒有 mapping 到對應的結果，則 reconstruction 會是四不像，而 VAE 則是使 latent representation 加了一些 noise(latent representation 附近) 後都能 mapping 到最後相同的 reconstruction。在不同圖片的 latent representation 的附近範圍會有重疊的部分，但 reconstruction 只會有一種結果，因此可能可以得到介於不同圖片的特徵，使 reconstruction 的結果更為真實。

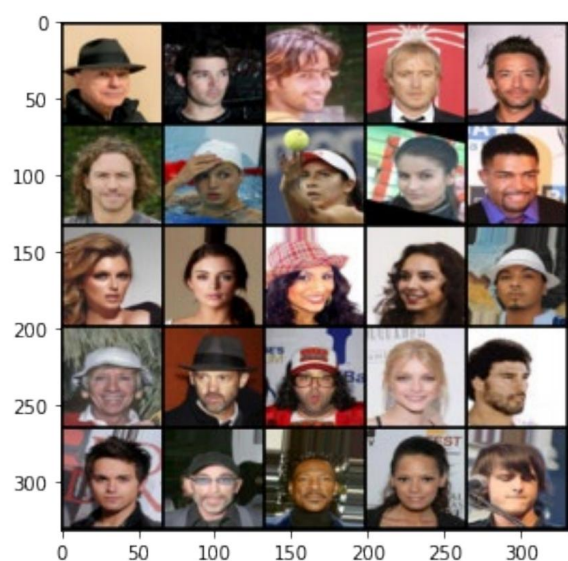
VAE 的一個缺點就是他沒有學習怎麼樣去產生出新圖片，而是著重在如何產生出跟 database 越相近的圖片，所以以圖片來說，若是相差了一個 pixel 的 reconstruction 圖片，無論這個 pixel 是在哪裡，model 都會認為他們是一樣差的，但對於人類來說，某些 reconstruct 的圖可能即使差了一個 pixel 我們也還是會認為他是真的圖而不是假的圖。

# 2.

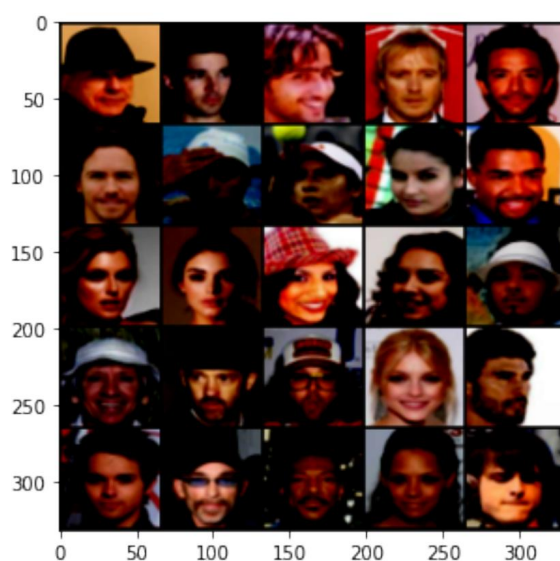
Model 架構:

Layer (type)	Output Shape	Param #
Linear-1	[-1, 4096]	50,335,744
BatchNorm1d-2	[-1, 4096]	8,192
ReLU-3	[-1, 4096]	0
Linear-4	[-1, 2048]	8,390,656
BatchNorm1d-5	[-1, 2048]	4,096
ReLU-6	[-1, 2048]	0
Linear-7	[-1, 4096]	8,392,704
BatchNorm1d-8	[-1, 4096]	8,192
ReLU-9	[-1, 4096]	0
Linear-10	[-1, 12288]	50,343,936
Tanh-11	[-1, 12288]	0
Total params: 117,483,520		
Trainable params: 117,483,520		
Non-trainable params: 0		
Input size (MB): 0.05		
Forward/backward pass size (MB): 0.42		
Params size (MB): 448.16		
Estimated Total Size (MB): 448.63		

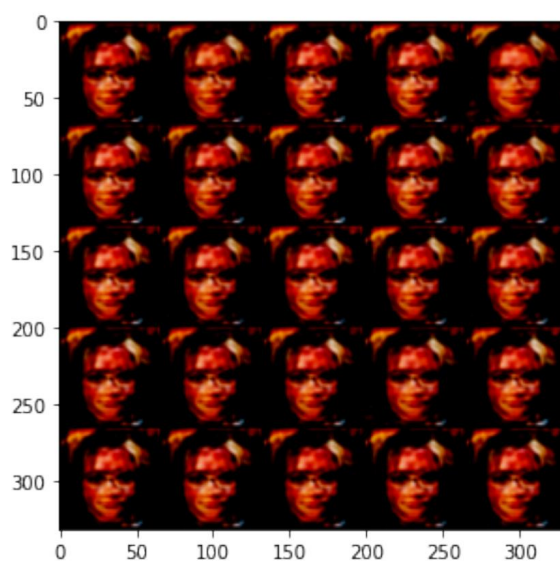
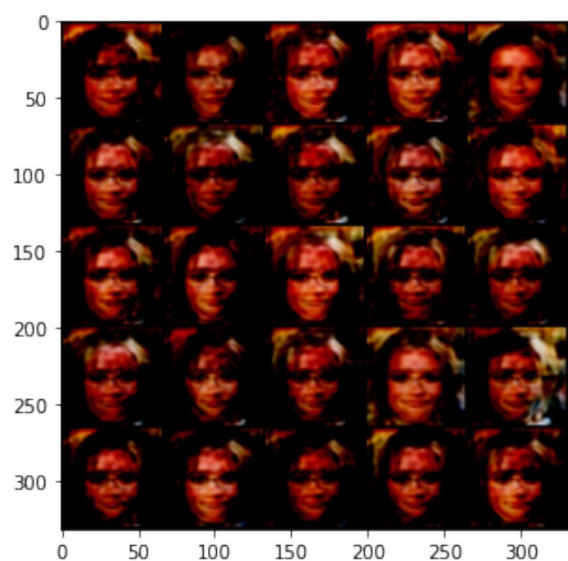
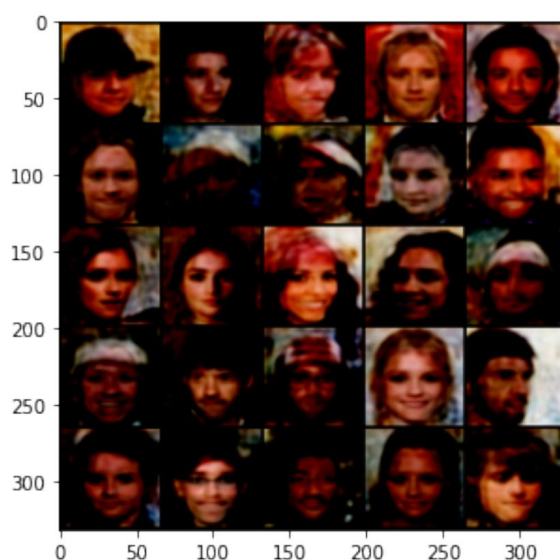
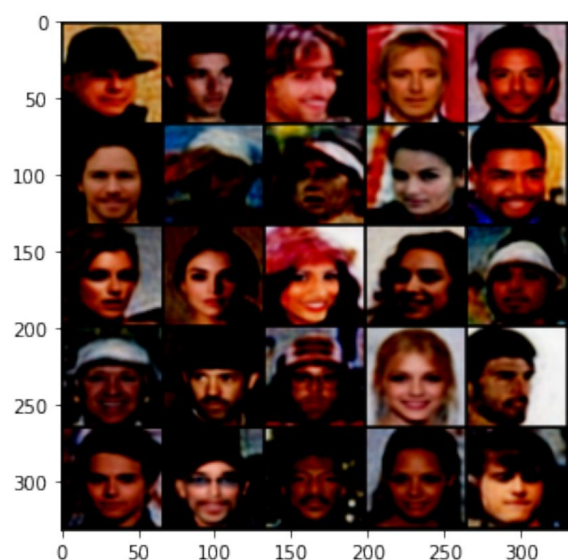
原始圖:



經過 Transform 的圖:



還原圖:



左上: 不做任何改動的 reconstruction

右上:  $\text{latent pixel value} * 0.8 + 0.2 * (\text{random noise} \rightarrow \text{每個 latent 都用相同的 noise})$

左下:  $\text{latent pixel value} * 0.5 + 0.5 * (\text{random noise} \rightarrow \text{每個 latent 都用相同的 noise})$

右下:  $\text{latent pixel value} * 0.2 + 0.8 * (\text{random noise} \rightarrow \text{每個 latent 都用相同的 noise})$

觀察:

首先這個 model 在 kaggle 的 public score 拿到了 0.83170 的成績。

我覺得添加 random noise 的比例會影響還原程度，像是我上面的是 20%、50%、80% random noise，可以從 20% 的 random noise 圖片開始有一些雜訊了，加到 50% 的 noise 可以開始看出有一些其他臉的輪廓和某些特徵浮現，像是臉的右上角有白白的一搓很像頭髮的東西，加到 80% 基本上所有的圖都長得一樣了，還原圖取決於我比例的調配會影響某些特徵的浮現。

2 Train a fully connected autoencoder and adjust at least two different element of the latent representation. Show your model architecture, plot out the original image, the reconstructed images for each adjustment and describe the differences. 2 / 2

✓ + 0.5 pts *Show your model architecture.*

✓ + 0.5 pts *Show your two adjustments on the latent representation.*

✓ + 0.5 pts *Show the original image and your two reconstructed images based on your adjustments.*

✓ + 0.5 pts *Describe the differences based on your reconstructed images (if any).*

+ 0 pts Major mistakes / No answer