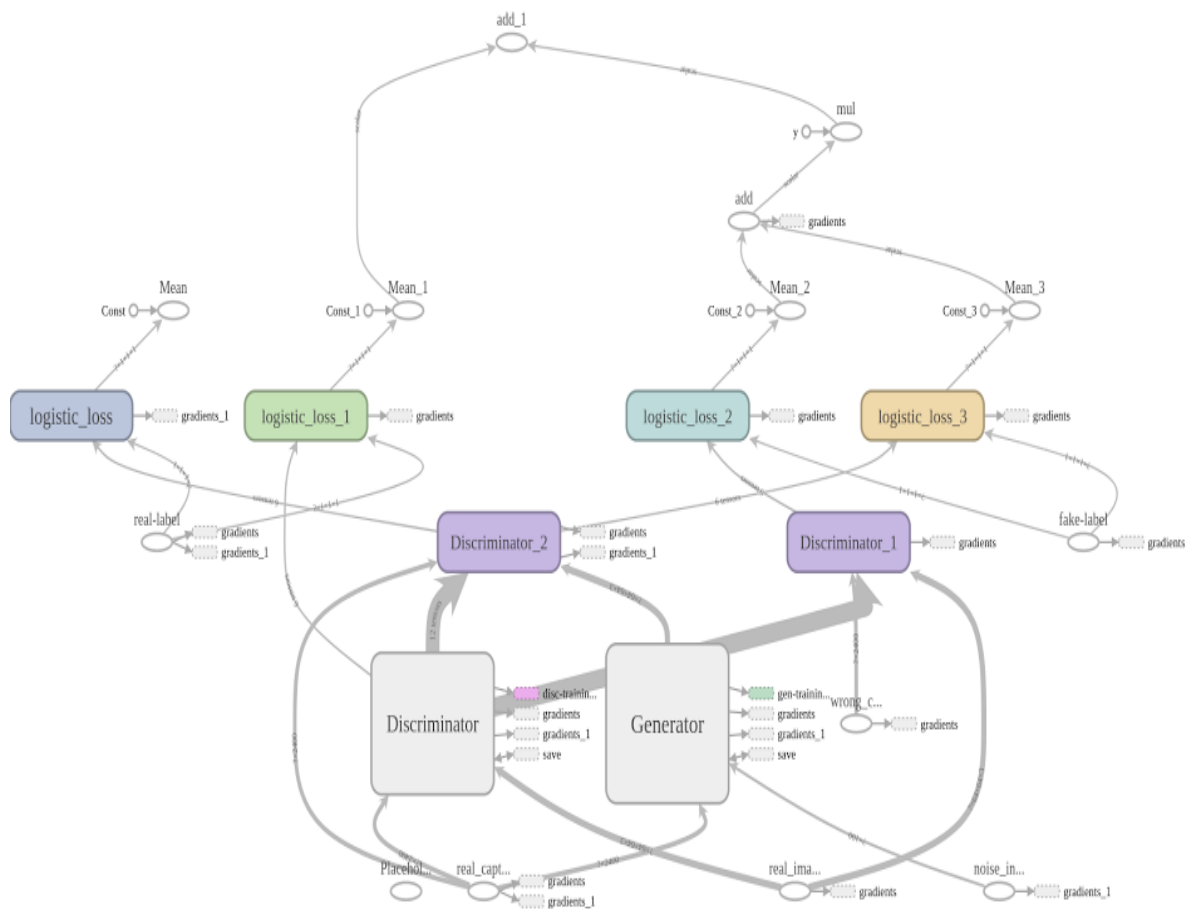


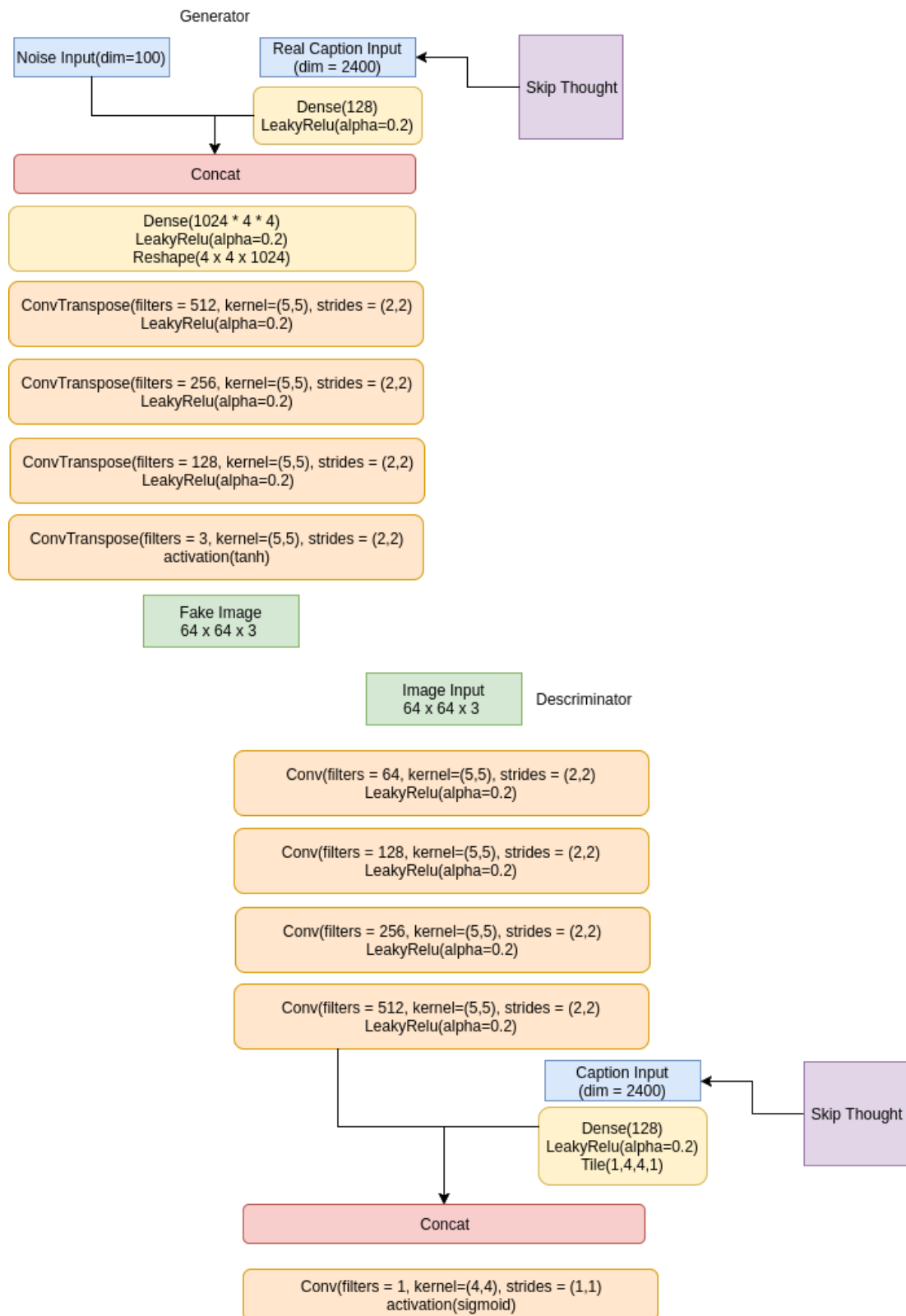
# 作業四報告

## 模型敘述:

我最後使用的模型是普通的GAN-CLS, 與Generative Adversarial Text to Image Synthesis那篇文章的架構很像. 唯一的差別就是不用batch normalization另外就是generator 和 discriminator 都是用leaky relu.



## 模型架構：



目標函數：

$$\begin{aligned}s_r &\leftarrow D(x, h) \text{ \{real image, right text\}} \\s_w &\leftarrow D(x, \hat{h}) \text{ \{real image, wrong text\}} \\s_f &\leftarrow D(\hat{x}, h) \text{ \{fake image, right text\}} \\\mathcal{L}_G &\leftarrow \log(s_f) \\\mathcal{L}_D &\leftarrow \log(s_r) + (\log(1 - s_w) + \log(1 - s_f))/2\end{aligned}$$

模型的改進：



我嘗試用了WGAN-GP，用同一個架構，RMSprop 可是出來的圖片都很模糊  
爲了讓discriminator能夠比較穩定我加了label smoothing，讓real值在每個batch隨機在0.7與1.2,結果效果不錯。  
另外就是把輸入標準化在-1和1之間，還有利用gaussian sampling而不是uniform sampling.

實驗設定與觀察：



實驗的時候發現batch normalization讓生成出來的臉都變得特別的亮，可是線條變得很模糊，只看得到眼睛。發現如果不用batch normalization效果會比較好。

在實驗中也發現用leaky relu discriminator和generator會比較穩定，loss 比較不會突然跳得很高。  
還有觀察到的另一點就是有如果convolution和convTranspose的維度太小，訓練一樣久生成出的照片也會比較模糊。

最後決定普通的GAN-CLS效果最好，以下是實驗的設定和生成出的照片。

Adam Optimizer( learning rate = 0.0002, momentum = 0.5)

Batch Size = 64

Epoch = 100

Noise\_dim = 100

Encode\_dim = 128

