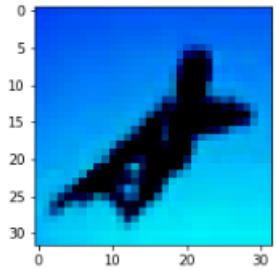
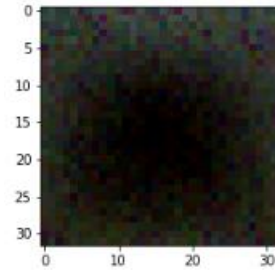
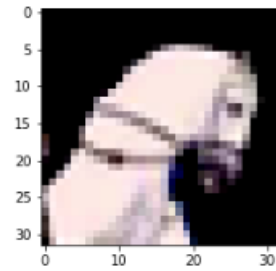
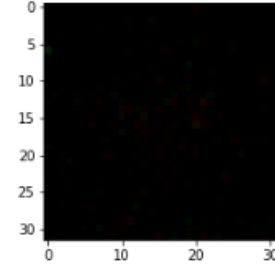
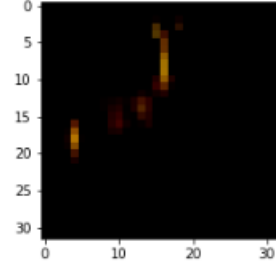
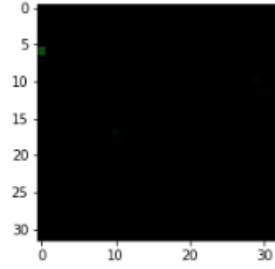
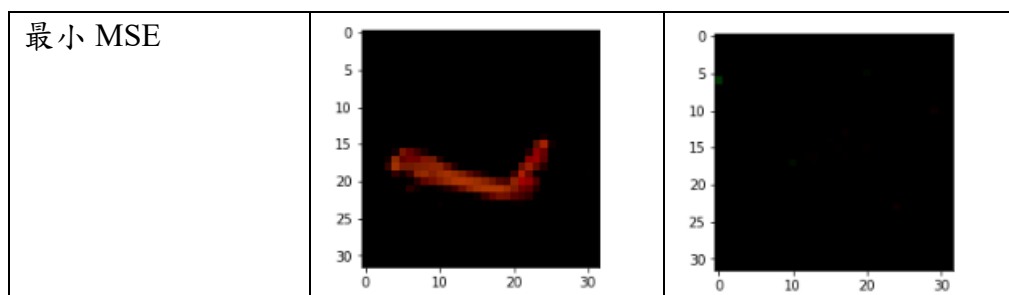


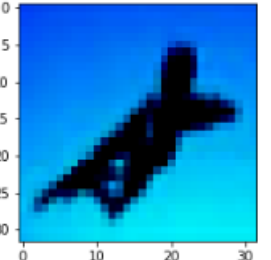
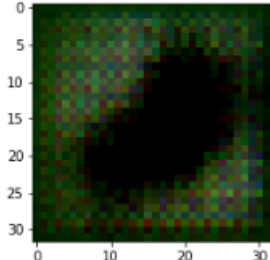
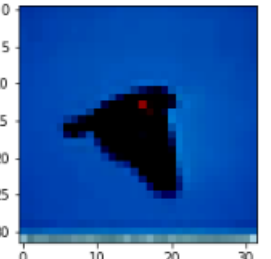
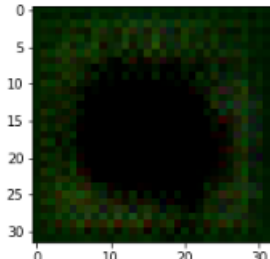
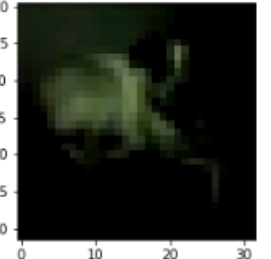
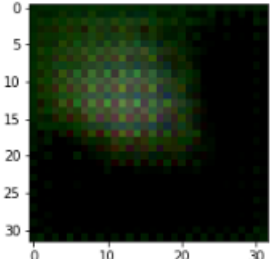
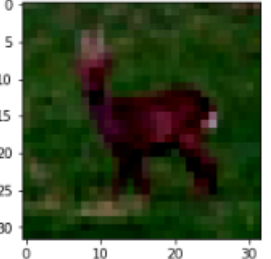
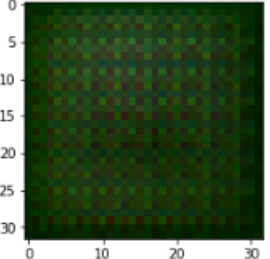
1. (2%) 任取一個 baseline model (sample code 裡定義的 fcn, cnn, vae) 與你在 kaggle leaderboard 上表現最好的單純 autoencoder 架構的 model (如果表現最好的 model 就是 sample code 裡定義的 model 的話就再任選一個, e.g. 如果 cnn 最好那就再選 fcn), 對各自重建的 testing data 的 image 中選出與原圖 mse 最大的兩張加上最小的兩張並畫出來。(假設有五張圖, 每張圖經由 autoencoder A 重建的圖片與原圖的 MSE 分別為 [25.4, 33.6, 15, 39, 54.8], 則 MSE 最大的兩張是圖 4、5 而最小的是圖 1、3)。須同時附上原圖與經 autoencoder 重建的圖片。(圖片總數: (原圖+重建)*(兩顆 model)*(mse 最大兩張+mse 最小兩張) = 16 張)

我的 best model 是將 sample code 中的 fcn epoch 調成 6, lr = 1e-5 然後就在 kaggle 上得到了 0.66641 的分數。

fcn(best model)	原圖	重建
最大 MSE		
次大 MSE		
次小 MSE		


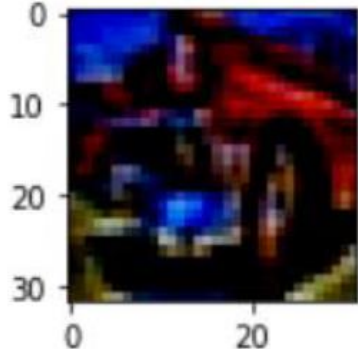

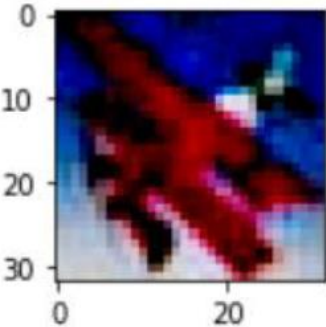

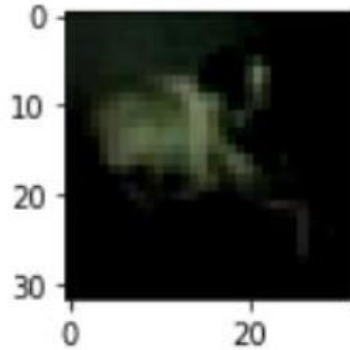

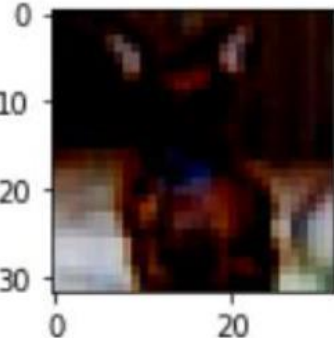


而下面的 model 是將 sample code 的 cnn 跑 6 個 epoch $lr = 1e-5$ 所訓練出來的 model

cnn	原圖	重建
最大 MSE		
次大 MSE		
次小 MSE		
最小 MSE		

有趣的是以上兩個 model 其實都沒辦法將圖片成功的 reconstruct 但是卻在 kaggle 上可以得到滿高的分數。

以下的 model 是我將 cnn model epoch 調成 1000 所訓練而成的，從圖片部分就可以看出來 **reconstruct** 的滿成功的，但是在 kaggle 上的表現卻不是很好。(這部分只是存粹分享一個有趣的發現)

cnn(epoch = 1000)	原圖	重建
最大 MSE		
次大 MSE		
次小 MSE		
最小 MSE		

2. (1%) 嘗試把 sample code 中的 K-means 與 PCA 分別做在 autoencoder 的 encoder output 上，並回報兩者的 auc score 以及本來 model 的 auc。autoencoder 不限。不論分數與本來的 model 相比有上升還是下降，請同學簡述原因。

原本的 auc : 0.66641

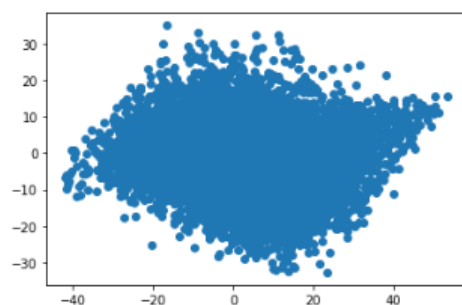
K-means 的 auc : 0.48778

PCA 的 auc : 0.50491

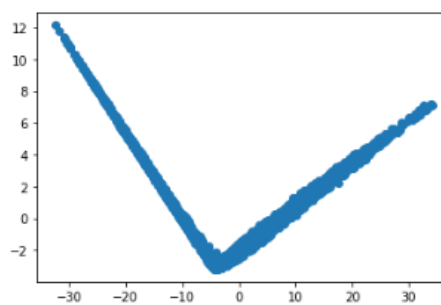
我們可從結果發現加上 K-means 跟 PCA 的 auc 都明顯下降很多，我想可能的原因是因為我的 best model 只有跑 6 個 epoch，雖然可以在 kaggle 上面拿到好的成績，但是這樣的轉換效果應該不會很好，在加上其它分群就會有問題以至於 auc 就下降很多。

3. (1%) 如 hw9，使用 PCA 或 T-sne 將 testing data 投影在 2 維平面上，並將 testing data 經第 1 題的兩顆 model 的 encoder 降維後的 output 投影在 2 維平面上，觀察經 encoder 降維後是否分成兩群的情況更明顯。（因未給定 testing label，所以點不須著色）

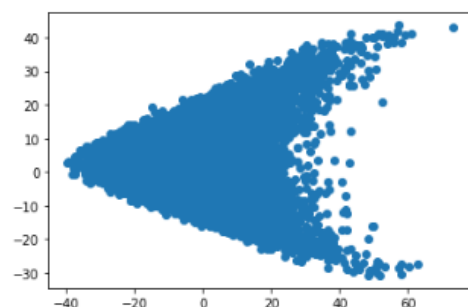
直接 PCA 降維：



經過 fcnn 降維後再透過 PCA 投影：



經過 cnn 降維後再用 PCA 投影：



經過 model 降維之後的投影圖形，fcnn 有更為集中的現象，cnn 則沒有明顯變化，兩者都有特定形狀的排列方式但是都沒有明顯的分群現象。

4. (2%) 說明為何使用 auc score 來衡量而非 binary classification 常用的 f1 score。如果使用 f1 score 會有什麼不便之處？

F1-score 是一個綜合考慮 precision 和 recall 的 metric：

$$2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$$

如果現在有兩個模型，一個 precision 特別高，recall 特別低，另一個 recall 特別高，precision 特別低的時候，這兩組的 f1-score 可能是差不多的，而這個並不是我們所想得到的答案。而 auc score 則不會發生上面的那種狀況。所以在不知道正反(anomaly/normal)資料的比例的情況下，使用 auc score 比起使用 f1 score 會比較好。