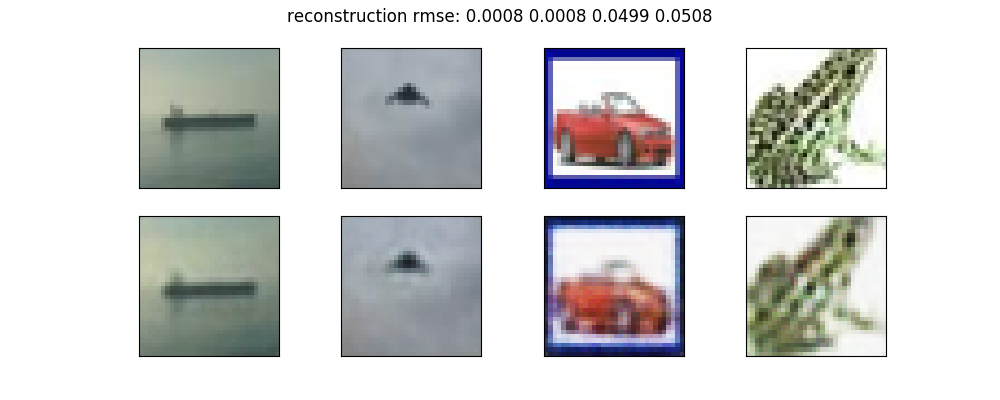
學號：r07921001 系級：電機所二 姓名：李尚倫

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）

首先第一個model用的是cnn-based autoencoder，重建的結果如下圖：

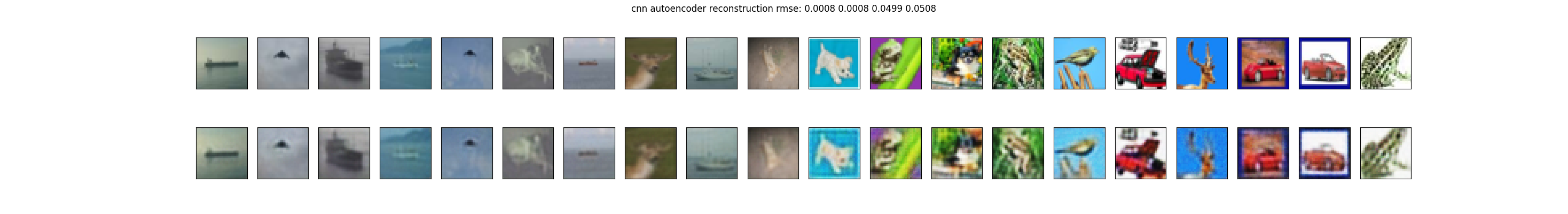


第二個model用的是fcn-based autoencoder（best），重建的結果如下圖：

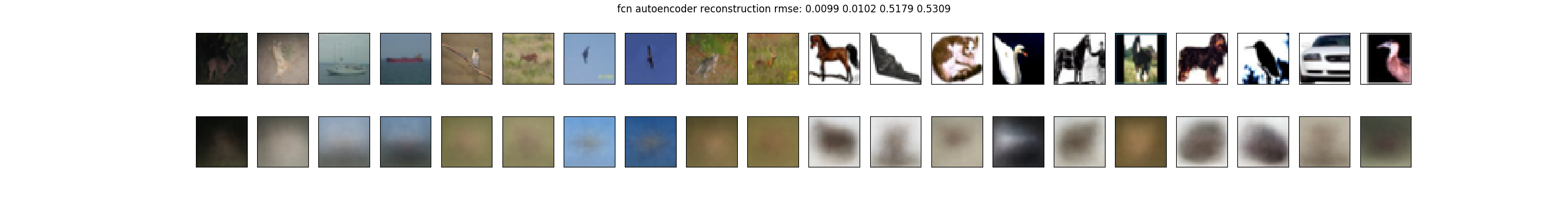


由上面的結果可以得知，cnn-based的autoencoder在圖片上的重建效果還是較fcn-based來的好很多，但是這次的作業考慮的點並非和重建效果完全相關，如果model能夠完美的重建出任何圖片，那anomaly的reconstruction error也會很低，就喪失了鑑別度。所以這個autoencoder如果要拿來做anomaly detection應該要對正確的類別的reconstruction error很低（在左側），沒看過的類別的reconstruction error很高（在右側），而為了看出更多的趨勢，下圖列出最高和最低error的十筆：

cnn-based autoencoder：



fcn-based autoencoder



並觀察的一些train dataset裡的圖片，原本有的圖片種類有貓 狗 鳥 飛機 車 青蛙 鹿 馬 船，因此還是看不太出來anomaly是指什麼，因為此作業並沒有給任何label，所以後來就直接丟kaggle，結果如下一題，用reconstruction error來當score算AUC的結果正確率僅不到六成。

1. (1%) 嘗試把 sample code中的KNN 與 PCA 做在 autoencoder 的 encoder output 上，並回報兩者的auc score。

|  |  |  |
| --- | --- | --- |
| Judging method | AUC score | |
| Sample FCN autoencoder | Best CNN autoencoder |
| Reconstruction error (rmse) | 0.59659 | 0.57497 |
| KNN on latent vector (dist to center) | 0.61389 | 0.62203 |
| PCA on latent vector (recon error) | 0.53198 | 0.49579 |
| Fusion (recon & knn) | 0.64277 | 0.63424 |

由上表可知靠reconstruction error做出來的AUC是FCN model比較好，這也驗證了第一題的想法，重建效果好並不等於anomaly detection的效果好，要必須有針對性才行。而針對autoencoder的latent factor做再處理，法ㄧ是KNN，對latent vector做分群，然後再計算各個點距離群中心的距離當作評斷的score，以兩種model來說都是分成三群(k=3)的效果最好。而法二則是PCA，對latent vector做PCA降維後再升回去，一樣看reconstruction error，這個方法在這兩個model上效果不太明顯，不如用原本的autoencoder的output的reconstruction error。法三是把auto encoder output image 的 reconstruction error和KNN k=3 距離中心的distance的score一起根據兩者個mean縮放到同一個scale後的平均，為最佳結果。

1. (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）

|  |  |  |
| --- | --- | --- |
| tSNE projection | | |
| Testing data (32\*32\*3) | Latent vector from cnn\_autoencoder (4\*4\*96) | Latent vector from fcn\_autoencoder (3) |
|  |  |  |

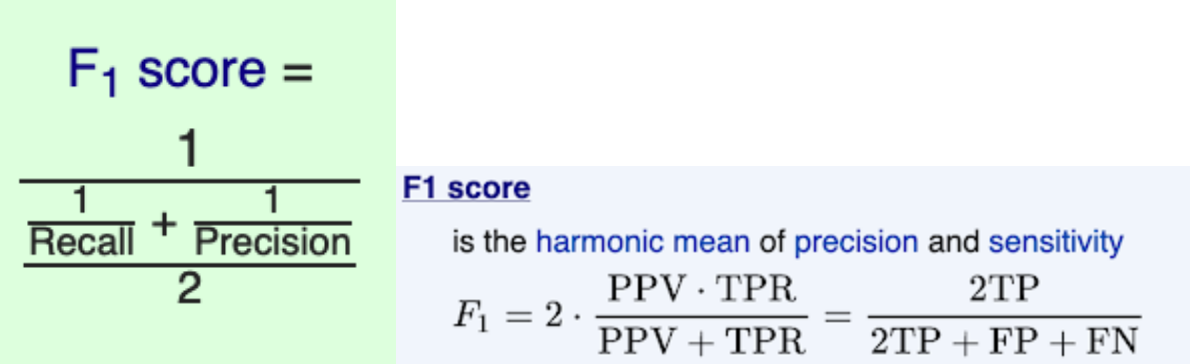
由上圖可以看出結果其實非常不明顯，畢竟最終結果的AUC也只有六成多，代表這兩個model都還是不能很明確的分出兩類。而若使用第二題提到的方式，對latent vector用knn分群，然後再畫一次tSNE投影會如下表：

|  |  |  |
| --- | --- | --- |
| tSNE projection with knn clustering label | | |
|  | Latent vector from fcn\_autoencoder | Latent vector from cnn\_autoencoder |
| k=2 |  |  |
| k=3 |  |  |
| k=4 |  |  |

由上表可知，隨著k值越大分群效果越差，且在k=2的時候亦不能明確的分出兩類，僅能使用第三題的方法，各個latent vector計算自己與cluster中心的距離當作score，距離中心越遠的越不屬於這個cluster，越有可能為Anomaly。

1. (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？

因為我們這次作業所輸出的並不是一個已經分完類的結果(0 or 1)，而是一個連續的score，F1 Score是以Recall 和 precision來計算的，所以勢必得直接先決定一個threshold然後根據以下公式算出值：



但如果用area under ROC curce (AUC score)的話，就可以針對這個output（一個連續的distribution）在不同的threshold底下畫出的以TPR和FPR為兩軸的ROC curve，然後計算曲線下面積，這個做法就不受限於threshold取得如何，能夠直接judge這個model所預測出的分佈是否正確。比起直接決定threshold得到分類結果再算F1 score，用ROC curve算AUC score能更好的判斷輸出不是直接分成兩類而是一個連續score的model的binary classification performance.

