

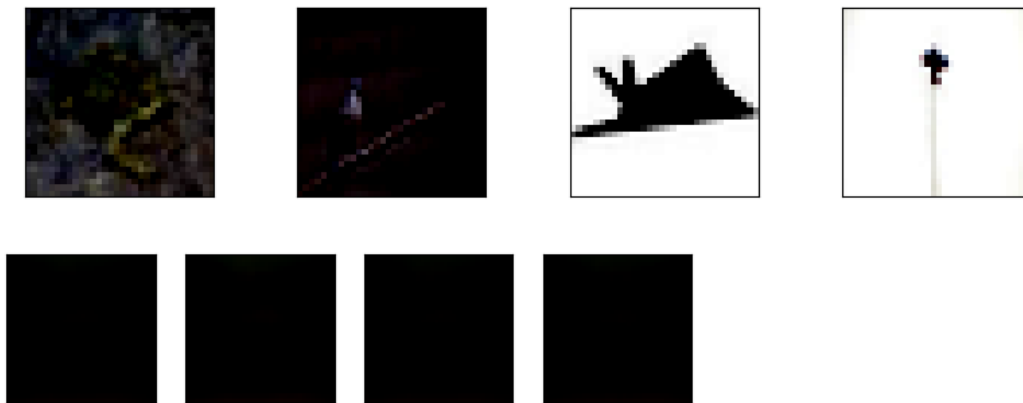
ML HW10 Report

學號： B05902010 系級： 資工四 姓名： 張頌平

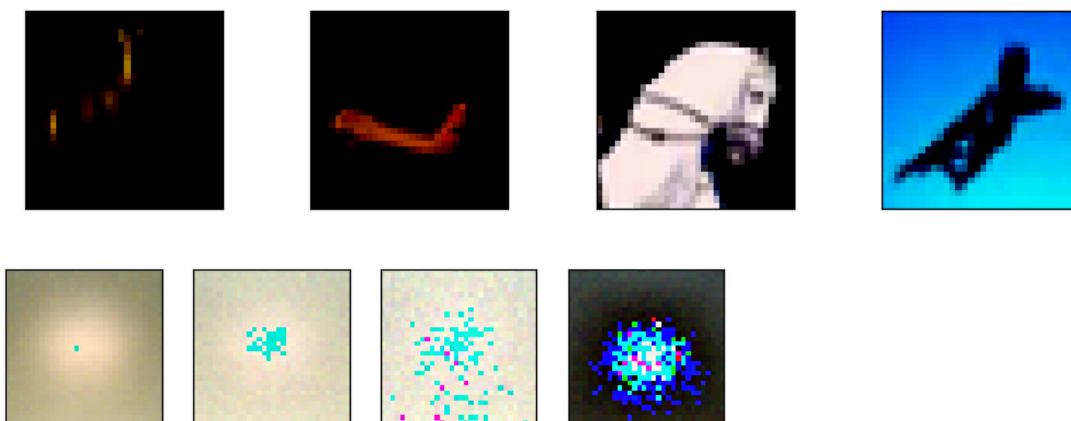
1. (2%) 任取一個baseline model (sample code裡定義的 fcn, cnn, vae) 與你在kaggle leaderboard上表現最好的單純autoencoder架構的model (如果表現最好的model就是sample code裡定義的model的話就再任選一個, e.g. 如果cnn最好那就再選fcn), 對各自重建的testing data的image中選出與原圖mse最大的兩張加上最小的兩張並畫出來。(假設有五張圖, 每張圖經由autoencoder A重建的圖片與原圖的MSE分別為 [25.4, 33.6, 15, 39, 54.8], 則MSE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

Ans:

Baseline model(vae): kaggle test auc: 0.58206



Best model(fcn): kaggle test auc: 0.65222



2. (1%) 嘗試把 sample code中的KMeans 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。

Ans:

原auc score: 0.65222

+KMeans: 0.48306

+PCA: 0.53061

原因:

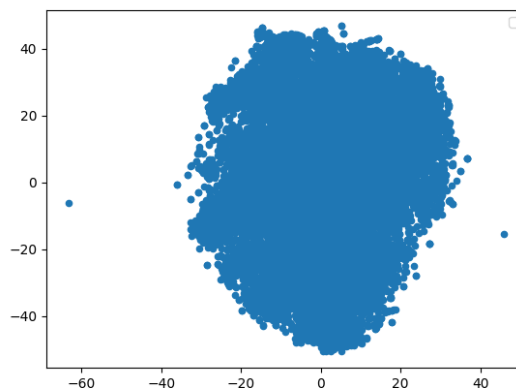
因為使用的autoencoder為fcn，經過encoder完的資訊其實只有3維，因此在只有fit test data的情況下，相比原本是對於整個reconstruct完的image做MSE，只對encoder output做降維並且算MSE結果可能因著整體資訊量太少，因此結果反倒不如原本不做降維處理來得好。

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）

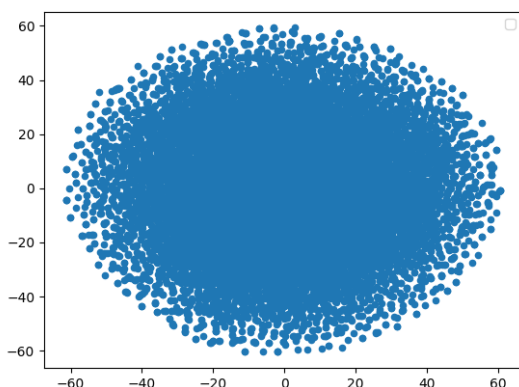
Ans:

可以發現使用我訓練的best model(fcn)降維後分群的效果的確比baseline model(vae)和原本的testing data來得更明顯

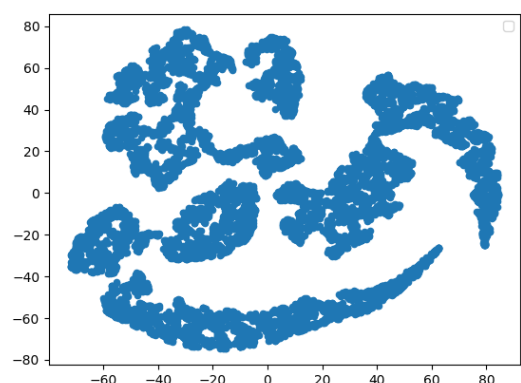
Testing data



Baseline(vae)



Best(fcn)



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？

Ans:

使用auc score來衡量的原因主要在於，我們這次主要的目標是判斷異常輸入，而我們的方法是透過Mean square error的大小來判斷圖片是否能夠被reconstruct，使用F1的話，我們可能比較難去定義到底哪個threshold才能夠判斷出最好的分數，而對於auc score來說，其主要的算法是去計算在各種threshold下，比較True Positive Rate 與 False Positive Rate 間的變化。因此如果二分類的分配差異越顯著(正常和異常差異越大)，AUC的分數就會越高，也就不會遇到像F1 score一樣不知道哪個threshold適合的問題，因此在這一個task，使用auc score來評估偵測結果會更加適當。