Homework 9 - Unsupervised Learning

學號：r08922051　系級：資工碩一　姓名：吳海韜

1. 請至少使用兩種方法 (autoencoder 架構、optimizer、data preprocessing、後續降維方法、clustering 算法等等) 來改進 baseline code 的 accuracy。

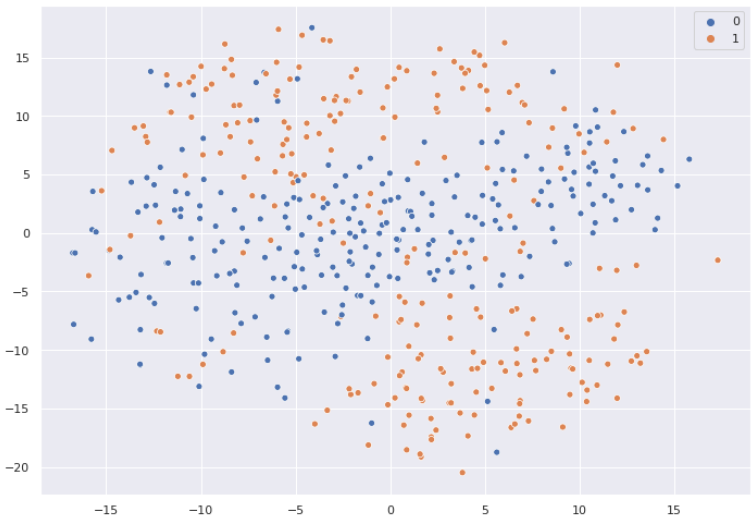
a. 分別記錄改進前、後的 test accuracy 為多少。

Model	Reconstruction Error	Kaggle Public Score
baseline	0.020446502	0.77058
improved	0.020175993	0.86658

其中 baseline model 是採用助教提供 sample code 的架構微調 learning rate (0.0003, 20 epochs) 得到的結果，兩者後續降維做法完全相同先做 kernel PCA、再 t-SNE 投影到二維做 clustering。Preprocessing 的部分是將像素值從 [0,255] 縮放成 [-1,1]，再計算 MSE 得到上表的 reconstruction error。

b. 分別使用改進前、後的方法，將 val data 的降維結果 (embedding) 與他們對應的 label 畫出來。

下圖是改進前的 embedding：



下圖是改進後的 embedding，可以看出代表兩種不同 datasets 的點被更清楚地分開：

c. 盡量詳細說明你做了哪些改進。

(改進1) Autoencoder 架構的部分效仿 resnet 的疊法，每一個 residual block 包含 (Conv3x3 -> Normalization -> Activation function -> Conv3x3 -> Normalization) 這五個步驟。其中 encoder downsample 用有參數的方式取平均，也就是 CNN kernel_size=2, stride=2。後半 decoder upsample 參考了 SRGAN 論文 (Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network) 提出的 pixel shuffle 來取代 ConvTranspose2d。實作部分是利用 pytorch 提供的 `nn.PixelShuffle` 層。假設 upsample 兩倍，那麼它會先通過 CNN 將通道數放大四倍、再分散當作該 pixel 四個方向（左上、右上、左下、右下） upsample 的結果，以下示意圖節自該論文：

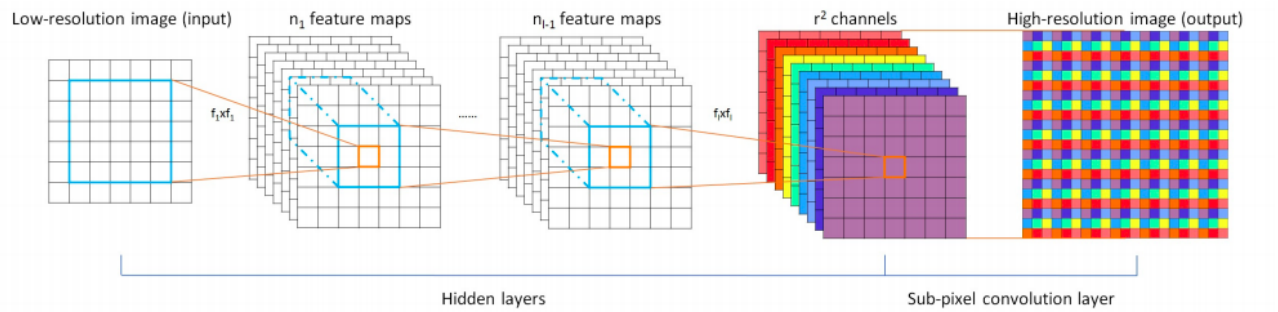


Figure 1. The proposed efficient sub-pixel convolutional neural network (ESPCN), with two convolution layers for feature maps extraction, and a sub-pixel convolution layer that aggregates the feature maps from LR space and builds the SR image in a single step.

(改進2) 訓練階段中將每張圖片 x 隨機旋轉 $d \in \{0, 90, 180, 270\}$ 度成 $\tilde{x}_d$ 再輸入，原本的 MSE Loss 要求 decoder 還原回輸入圖片 $\tilde{x}_d$，測試階段算 MSE 的時候不做旋轉只算 $\tilde{x}_0$ 的部分。另外在 latent code 加入一層 linear classifier 要求模型判斷輸入的圖片被旋轉成四個方向中的哪一種。令 L_{MSE} 為原本的 reconstruction loss，$L_{rotation}$ 為分類四個方向的 cross-entropy loss，給定超參數 β 和模型參數 θ，整個 objective function 如下：（最後上傳的模型 $\beta = 10$）

$$\arg_\theta \min L(\theta) = L_{MSE} + \beta \cdot L_{rotation}$$

這種 self-supervised 的想法曾經被用在 GAN 中作為 discriminator 的一種 regularization，希望能更好地提取真實圖片的特徵，示意圖如下（節自論文 Self-Supervised GAN to Counter Forgetting）：

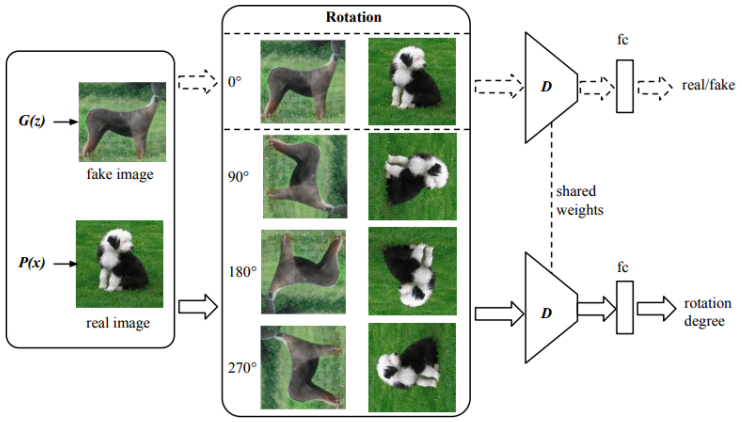
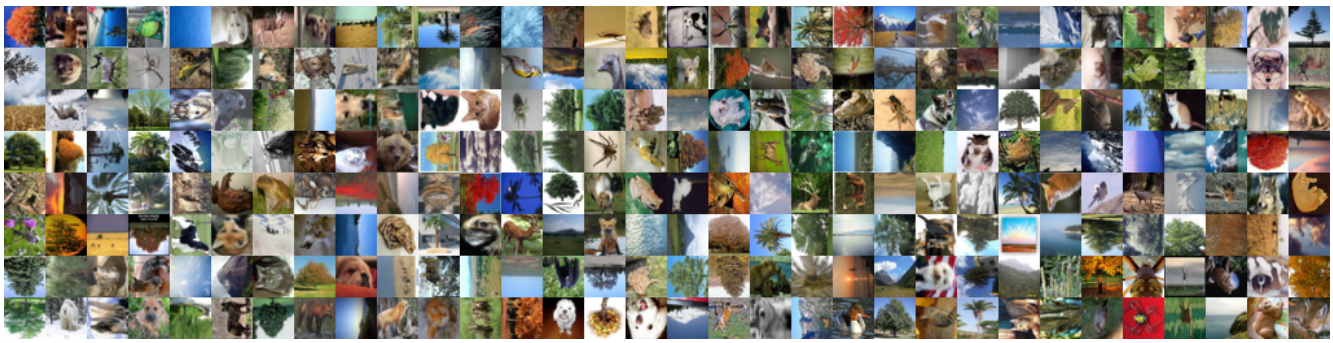


Figure 4: Rotation feature learning discriminator. The dotted line arrow uses only the non-rotated images, for real/fake classification task. The solid line arrow uses all the four rotated images, for the rotation classification task.

以下從 training data 中 sample 一些隨機旋轉的圖片：



已知這次作業要區分的 datasets 是有生物的照片或風景照的前提下，模型可能會因為要正確分類出轉置的方向，在 encoder 的部分提取更多有意義的資訊。例如：貓狗這種動物照片的腳應該在圖片下半部、風景照的雲朵出現在圖片上半部、樹的照片樹幹在下樹葉在上……等。從結果而言加入這段 self-supervised loss 對準確率提升是有一些幫助。

2. 使用你 test accuracy 最高的 autoencoder，從 trainX 中，取出 index 1, 2, 3, 6, 7, 9 這 6 張圖片

a. 畫出他們的原圖以及 reconstruct 之後的圖片。

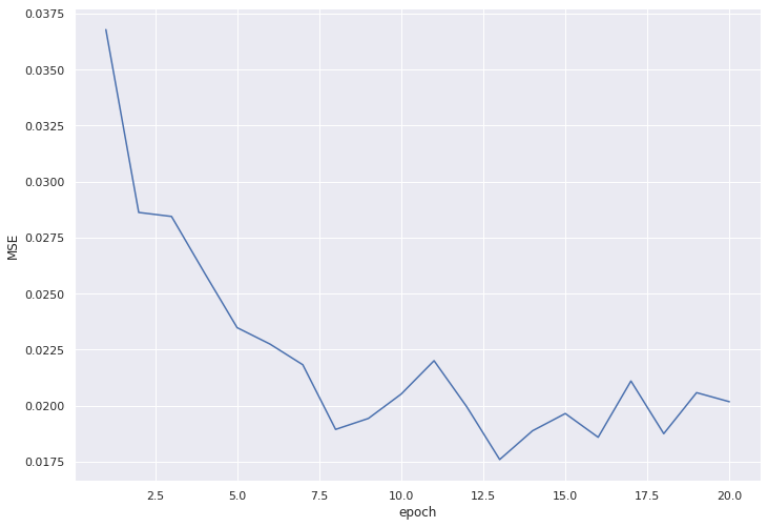
還原結果如下：

我在 Kaggle 上 accuracy 最高的模型在 reconstruction error 的部分和 baseline model 差不多，還原結果（下排）比原圖（上排）略為模糊一些，這部分和我們一般對 autoencoder 的直覺相符。

3. 在 autoencoder 的訓練過程中，至少挑選 10 個 checkpoints

a. 請用 model 的 train reconstruction error (用所有的 trainX 計算 MSE) 和 val accuracy 對那些 checkpoints 作圖

以下是總共 20 epochs 訓練過程 training set 的 MSE 值：



每個 epoch 結束都針對 500 筆 validation set data 做降維分群並計算準確率，為了預防準確率小於 0.5 的情形，以下都選擇記錄 $acc_{rev} = \max(acc, 1 - acc)$：

b. 簡單說明你觀察到的現象。

訓練剛開始的階段，MSE 下降的同時 validation set accuracy 也跟著上升。但我們 autoencoder 所使用的 loss function 並不是專門為了 downstream task 所設計（因為沒有 labels），所以當 autoencoder 訓練到一個程度以後，MSE 下降就不必然會伴隨 accuracy 的改進，此時 encoder 提取出的特徵就可能不是最適合做分群的特徵。