



Embedded Vision Intelligent Laboratory

# 嵌入式智慧影像分析與實境界面

## Fall 2021

Instructor : Yen-Lin Chen(陳彥霖), Ph.D.

Professor

Dept. Computer Science and Information Engineering

National Taipei University of Technology

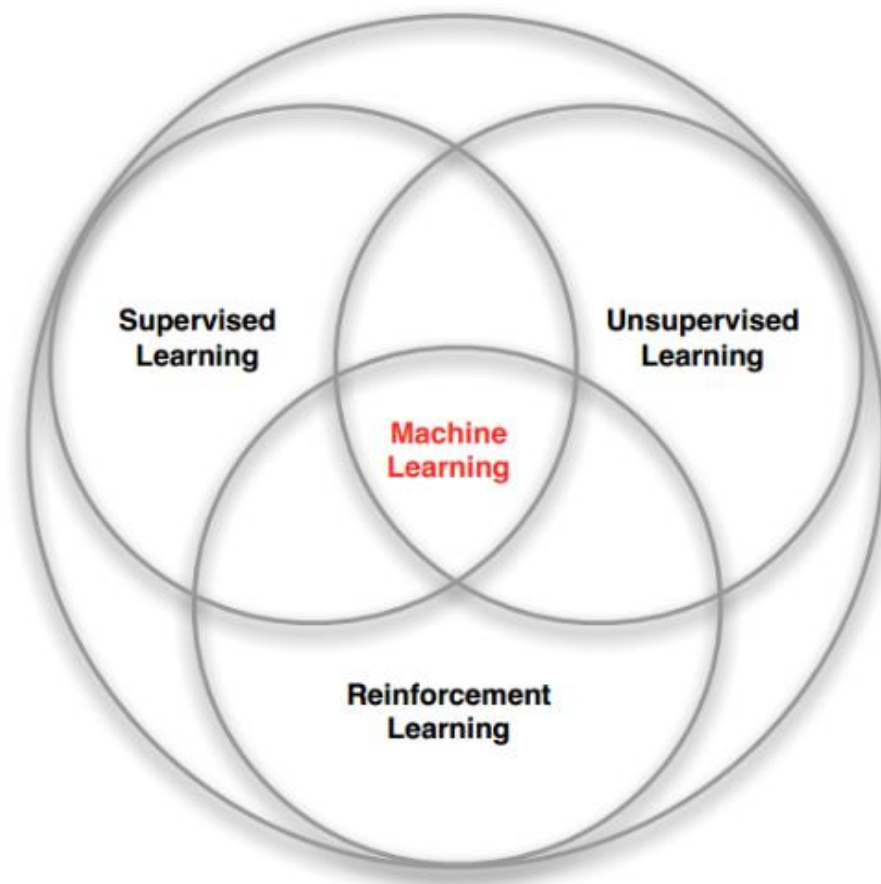
# Lecture 10

非監督式學習與神經網路介紹

# Unsupervised Learning (非監督式學習)



# Machine Learning 三大元素





# Unsupervised Learning

- **非監督學習**（英語：unsupervised learning）是機器學習的一種方法，沒有給定事先標記過的訓練範例，自動對輸入的資料進行分類或分群。
- 非監督學習的主要運用包含：集群分析（cluster analysis）、關聯規則（association rule）、維度縮減（dimensionality reduce）。
- AutoEncoder 是多層神經網絡的一種非監督式學習算法。

# Encoder-Decoder 框架



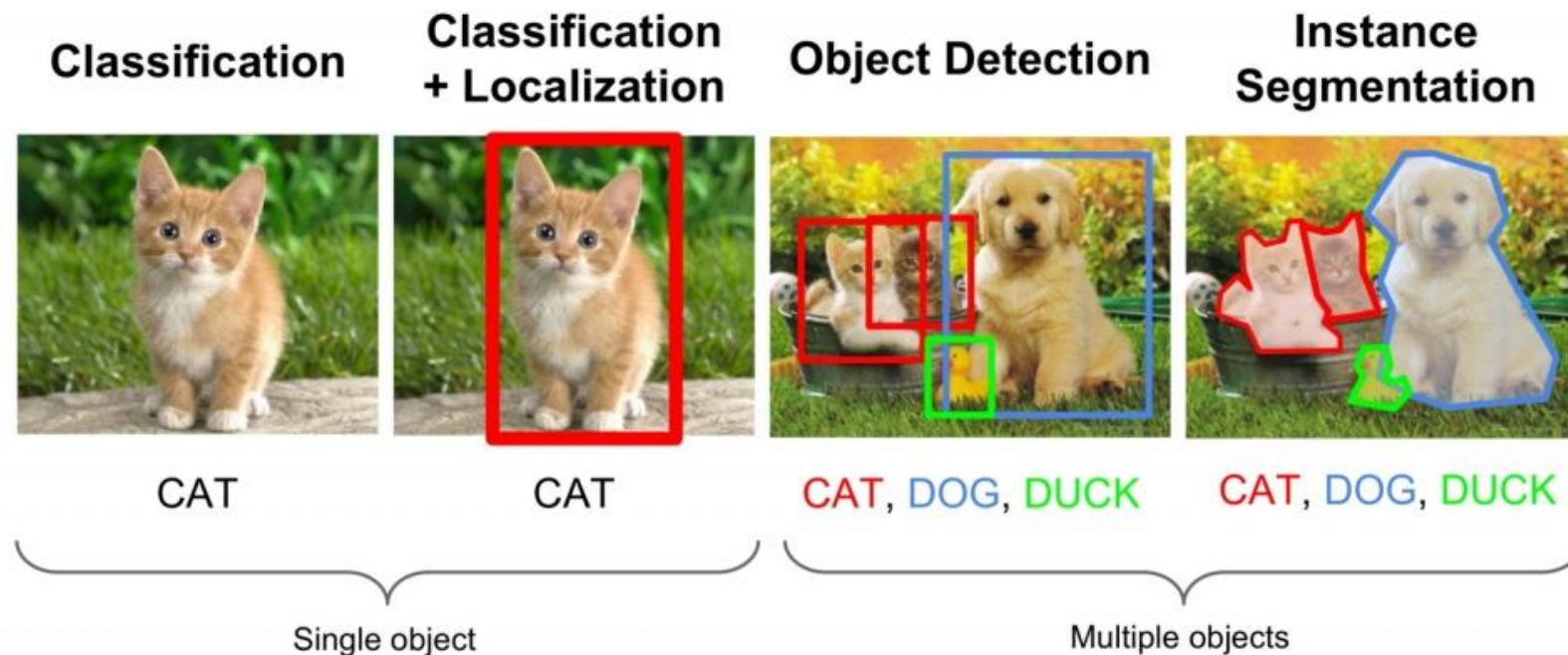
# Encoder-Decoder框架

- Encoder-Decoder（編碼-解碼）是深度學習中非常常見的一個模型框架，比如非監督演算法的auto-encoding就是用Encoder-Decoder的結構設計並訓練的；比如這兩年比較熱門的image caption的應用，就是CNN-RNN的Encoder-Decoder框架；再比如神經網路機器翻譯NMT模型，往往就是LSTM-LSTM的Encoder-Decoder框架。因此，準確的說，**Encoder-Decoder並不是一個具體的模型，而是一類框架。**
- Encoder和Decoder部分可以是任意的文字，語音，影像，視訊資料，模型可以採用CNN，RNN，BiRNN、LSTM、GRU等等。所以基於Encoder-Decoder，我們可以設計出各種各樣的應用演算法。



# 影像Encoder-Decoder框架

- 在電腦視覺的領域中有幾個有名的問題：
  - 影像辨識 (Image recognition)
  - 物件辨識 (Object detection)
  - 語意分割 (Semantic segmentation)







# 影像Encoder-Decoder框架

- 影像辨識是給一張影像，希望模型可以辨識出當中的東西是什麼。輸入模型的會是影像向量，輸出的會是類別向量。
- 物件辨識給的同樣是一張影像，除了需要辨識出當中的物件以外，還要給出這個物體所在的位置，輸出的除了類別向量以外，還有座標。



Input

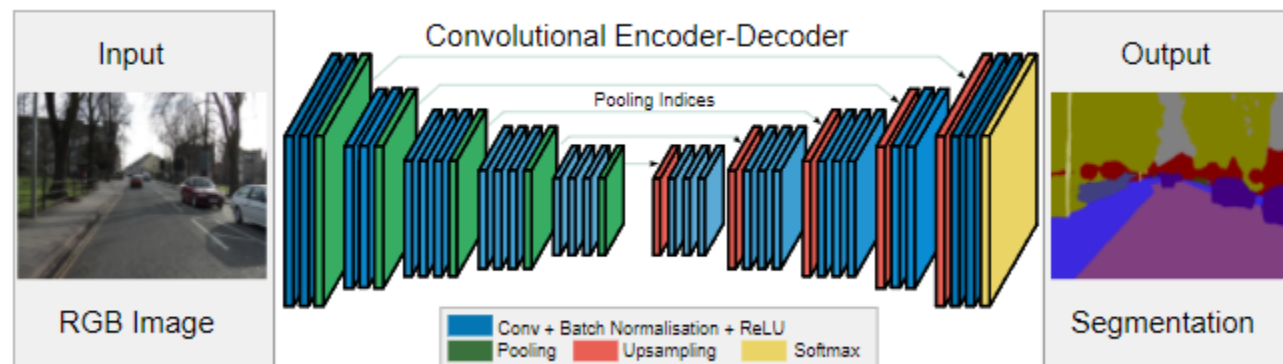


Segmentation [9]



# 影像Encoder-Decoder框架

- 對於影像辨識來說，一般架構上會是有 convolution layer 為主的 feature extractor，接著會是以 fully-connected layer 為主的 classifier。在不同階段有不同的目的，在輸入影像之後要先對影像進一步抽取特徵，有了足夠的特徵之後才進行分類。
- 像素的類別預測這件事從另一個角度切入，會很像是一種生成的過程。也就是，我們在前面要將影像的特徵萃取出來，是一種將資訊壓縮的過程，在後半我們希望將壓縮的資訊還原到某種程度，我們需要產生器（generator）。
- 在這邊 encoder 就是 feature extractor，decoder 就是一種 generator。

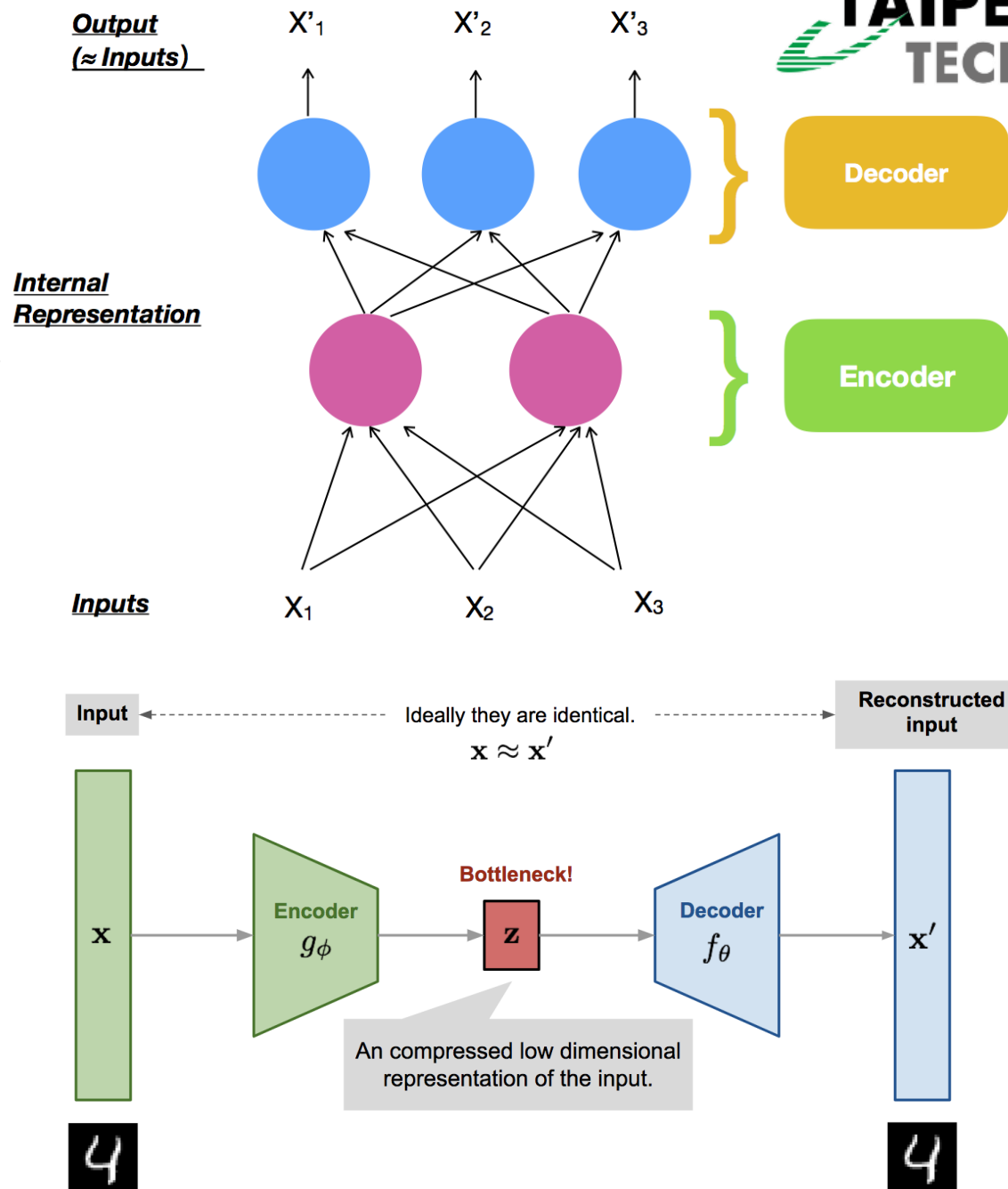


# AutoEncoder



# AutoEncoder構想

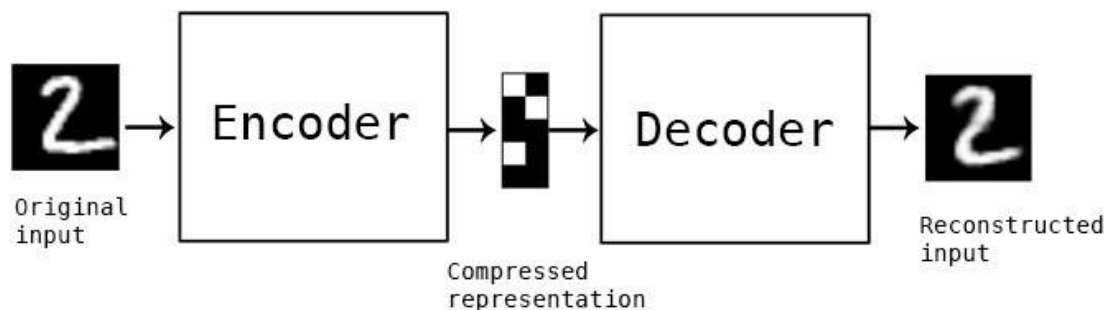
- 在處理異常偵測時。可以像PCA一樣，可以從input擷取重要特徵，代表全體。當新的測試資料進來，和這樣的代表特徵比對，就可以判斷是不是異常。
- 因此設計了一個非監督式學習的神經網路，其中中間的**Internal Representation**(又稱為**Bottleneck**)可以看做是對輸入的資料做壓縮(維度限制)或是加入雜訊到輸入資料





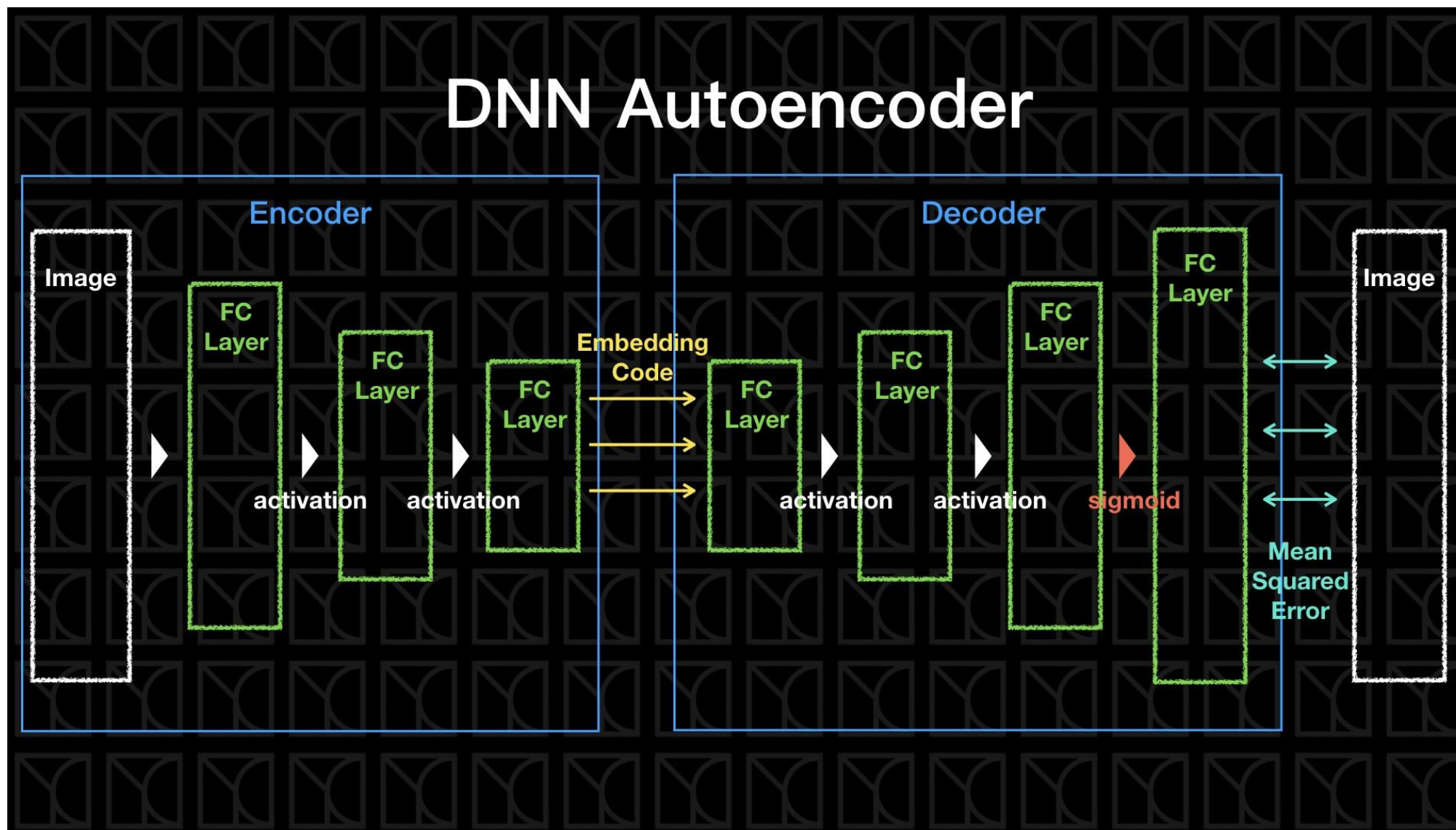
# AutoEncoder(AE)

- AutoEncoder 是多層神經網絡的一種**非監督式學習算法**，稱為自動編碼器，它可以幫助資料分類、視覺化、儲存。
- 其架構中可細分為 **Encoder**（編碼器）和 **Decoder**（解碼器）兩部分，它們分別做壓縮與解壓縮的動作，讓輸出值和輸入值表示相同意義
- 透過重建輸入的神經網路訓練過程，隱藏層的向量具有降維的作用。特點是編碼器會建立一個隱藏層（或多個隱藏層）包含了輸入資訊的低維向量。然後有一個解碼器，會通過隱藏層的低維向量重建輸入資料。通過神經網路的訓練最後AE會在隱藏層中得到一個代表輸入資料的低維向量。





# AutoEncoder(AE)



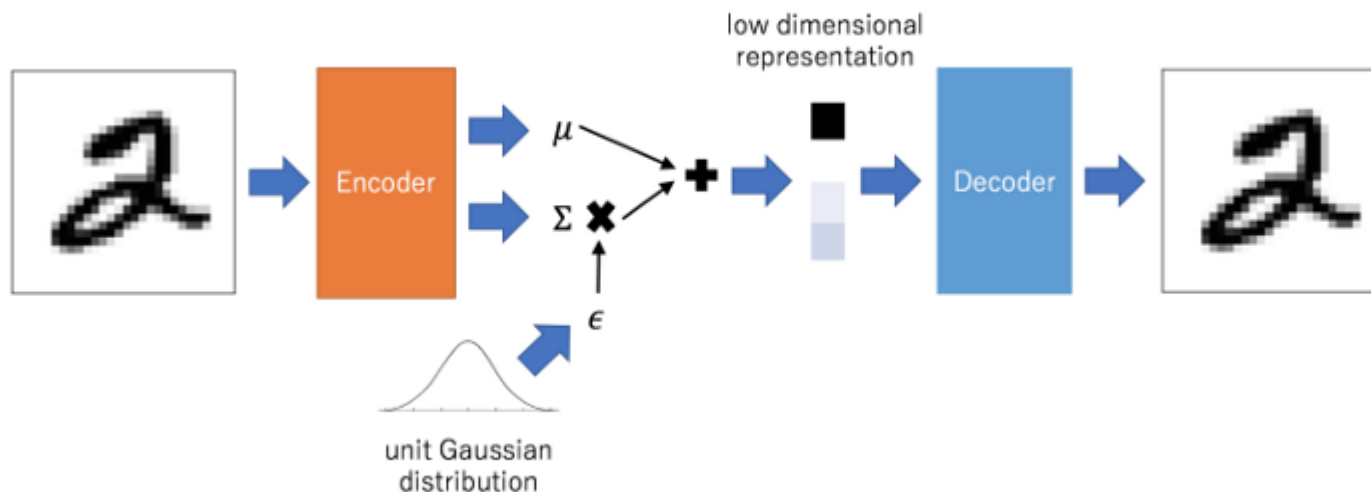
# Variational AutoEncoder(VAE)





# Variational AutoEncoder(VAE)

- 可以看出與 AutoEncoder 不同之處在於 VAE 在編碼過程增加了一些限制，迫使生成的向量遵從高斯分佈。由於高斯分佈可以通過其 *mean* 和 *standard deviation* 進行參數化，因此 VAE 理論上是可以讓你控制要生成的圖片

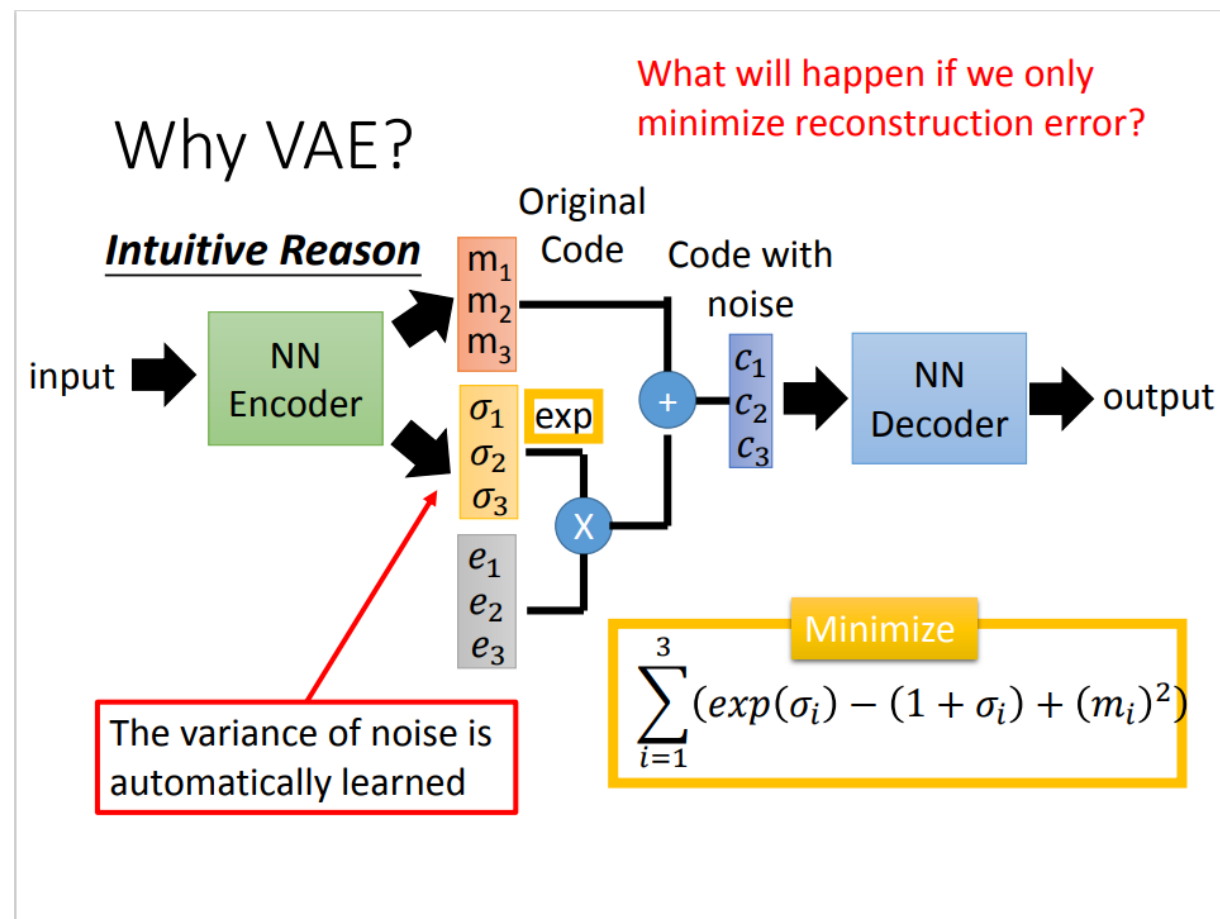






# Variational AutoEncoder(VAE)

- VAE 的內部做法：
  - 先輸出兩個向量:mean和standard deviation
  - 用normal distribution產生第三個向量
  - 把第二個向量做exponential，之後跟第三個向量做相乘後，把它跟第一個向量相加，極為中間層的隱含向量

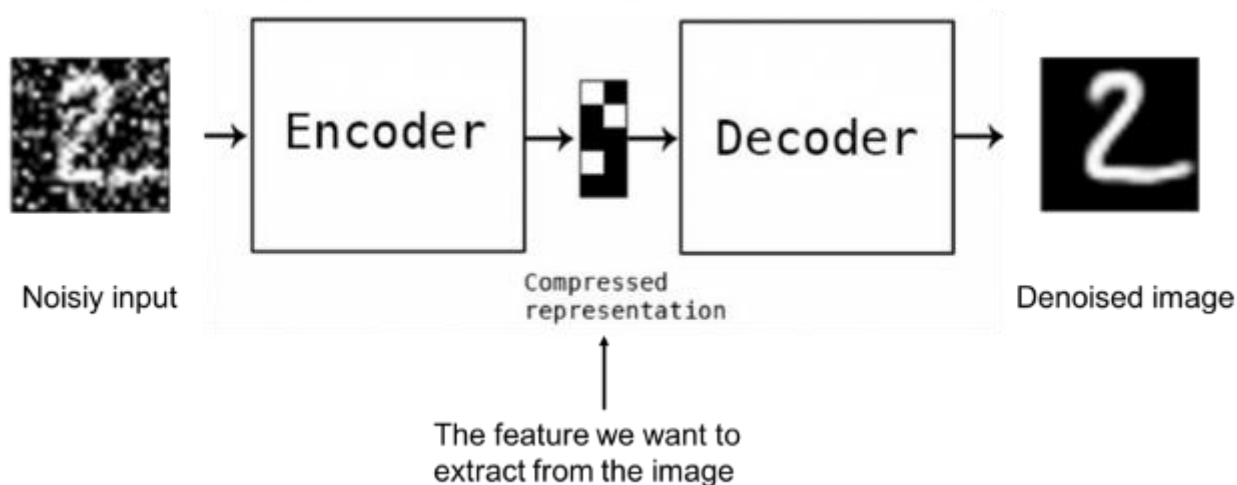


# Denoising AE ( DAE )



# Denoising AE ( DAE )

- Denoising AE 是一種學習對圖片去噪（denoise）的神經網絡
- 可用於從類似圖片中提取特徵到訓練集
- 實際做法是在 **input** 加入隨機 **noise**，然後使它回復到原始沒有 **noise** 的資料，使模型學會 **denoise** 的能力。

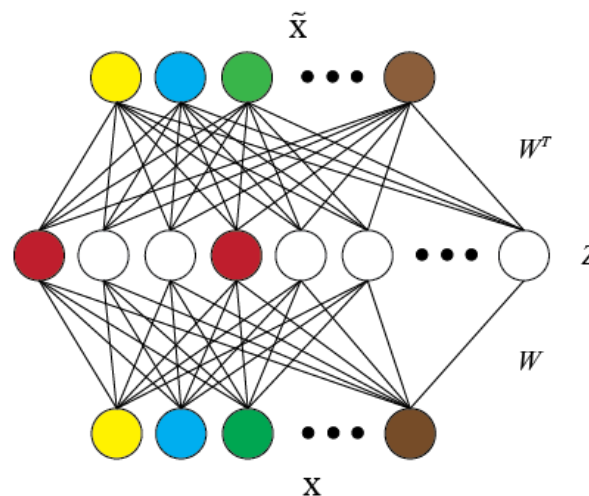


Sparse AE ( SAE )



# Sparse AE (SAE)

- Sparse AE 的作法是在 AutoEncoder 基礎上加上 **Sparsity Regularization** 和 **L2 Regularization**，限制每次得到的 representation 盡量稀疏，迫使自動編碼器將每個輸入表示為少量節點的組合，只有一小部分節點具有非零值，稱為活動節點。
- 在特徵稀疏的過程裡可以過濾掉無用的訊息，每個神經元都會訓練成辨識某些特定輸入的專家，因此 Sparse AE 可以給出比原始資料更好的特徵描述。





# Sparsity Regularization

- 我們想要做的事就是讓 autoencoder 中每個神經元的輸出變小，而實際上的做法則是如下
- 先設定一個值，然後讓平均神經元輸出值 (average output activation value) 越接近它越好，如果偏離這個值，cost 函數就會變大

$$\hat{\rho}_i = \frac{1}{n} \sum_{j=1}^n h(w_i^T x_j + b_i)$$

$\hat{\rho}_i$  : average output activation value of a neuron  $i$

$n$  : total number of training examples

$x_j$  :  $j$ th training example

$w_i^T$  :  $i$ th row of the weight matrix  $W$

$b_i$  :  $i$ th element of the bias vector



# L2 Regularization

- 經過了 Sparsity Regularization 這一項，理想上神經元輸出會接近我們所設定的值，而這裡想要達到的目標就是讓 **weight** 盡量的變小，讓整個模型變得比較簡單，而不是 **weight** 變大，使得 **bias** 要變得很大來修正

$$\Omega_{weights} = \frac{1}{2} \sum_l^L \sum_j^n \sum_i^k (w_{ji}^{(l)})^2$$

$L$  : number of the hidden layers

$n$  : number of observations

$k$  : number of variables in training data

# Applications of AutoEncoder





# Model pretrained weight

- Autoencoder 可以用於 weight 的 pretrain，讓模型找到一個較好的起始值。
- 如果label data 非常少，但卻有一堆的沒 label data 時，可以用這樣的方法去做 weight pre-train，因為 Autoencoder 本身就是一個 unsupervised learning 的方法，用沒 label 過的 data 先得到 weight 的 pre-train，再用有 label 的 data 去 fine-tune 權重，如此一來就可以得到不錯的模型。



# Image segmentation

- 首先將 input data 進行label產生output，接著建構一個 network，輸入原始的圖片(左圖，牙齒 X 光照)後，要產生output (右圖，牙齒構造分類)。
- 在這種 情況下，encoder & decoder 都會是圖形判斷較為強大的 convolution layers，萃取出有意義的特徵，並在 decoder 裡 deconvolution 回去，便可以得到 segmentation 的結果。



Dental X-Ray Image with 7 classes



# Video to Text

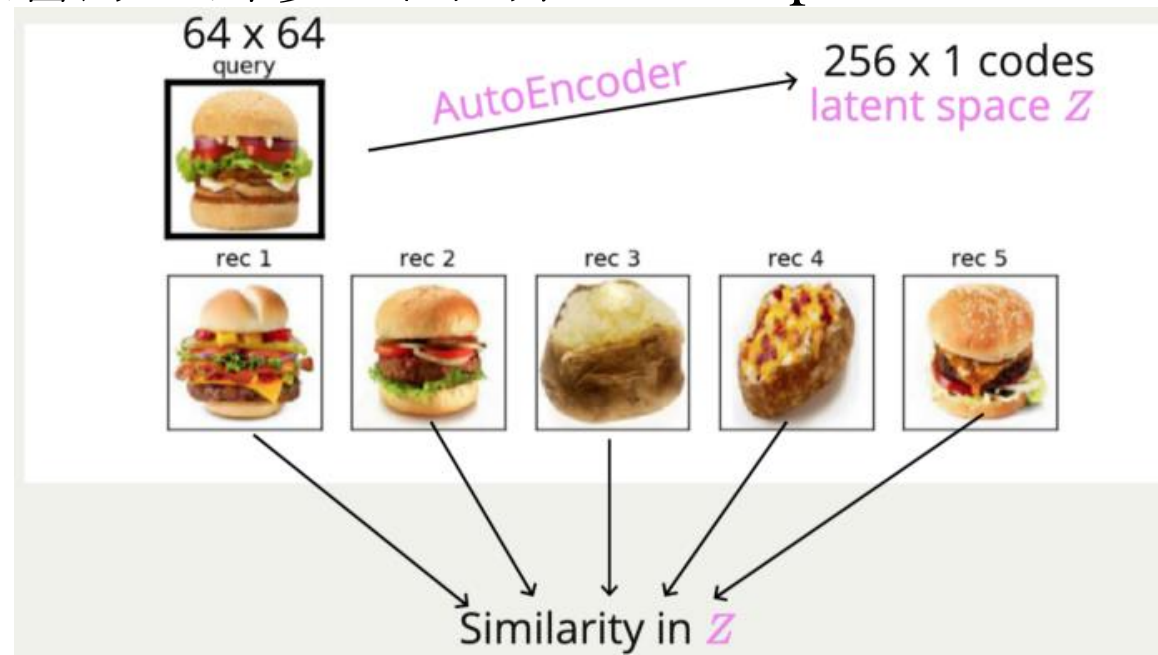
- 關於Image caption 問題，會用到 sequence to sequence 模型。
- Input data 照片，output 則是描述照片的一段文字，在sequence to sequence 模型使用 LSTM + ConvNet 當作 encoder & decoder，可以很好的描述一連串的有順序性的動作。





# Image Retrieval

- Image Retrieval 試著從輸入圖片中找出最相近的圖。
- 以 pixel-wise 的方式去比照，很容易找到完全不相干的東西，因為機器沒有學到特別的特徵，只是將每個 pixel 去計算 loss。
- 若使用 Autoencoder 的方式，先將圖片壓縮到 latent space，再對 image 的 latent space 計算 similarity，出來的結果便會好上許多，因為在 latent space 裡的每一個維度，可能代表了某種特徵。

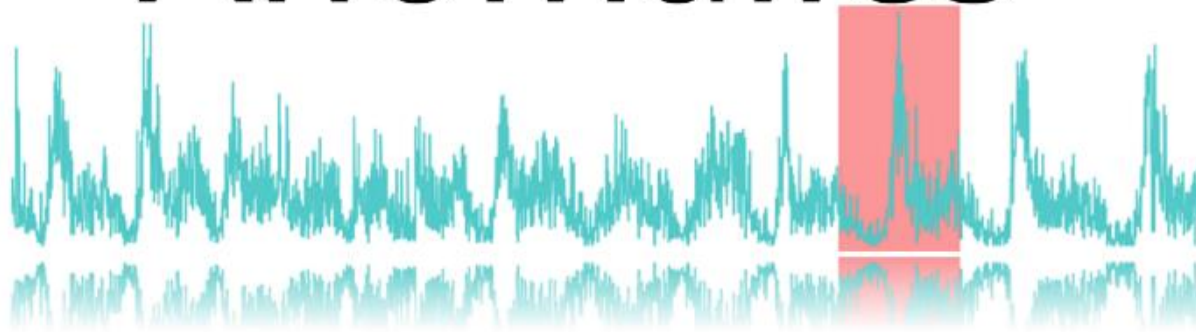




# Anomaly Detection

- 生活中有時會有anomalies 的狀況發生，如機台的訊號異常，溫度突然飆升.....等情況。但在多數情況機器運作正常的資料會遠大於異常的資料。
- 對於這種狀況，最好的方法就是去拿原有的大量正常情況資料拿去訓練出一個好的 Autoencoder，這時若有 anomalies 進來，那麼reconstruct 後的圖形就會壞掉，進而找出anomalies 的狀況。

## Anomalies



# 參考資料



# 參考資料

- [AutoEncoder \(一\)-認識與理解 - NLP-ML筆記 - Medium](#)
- [VAE \(ntu.edu.tw\)](#)
- [22 Convolutional encoder-decoder 架構 - iT 邦幫忙::一起幫忙解決難題，拯救 IT 人的一天 \(ithome.com.tw\)](#)
- [Tensorflow Day17 Sparse Autoencoder - iT 邦幫忙::一起幫忙解決難題，拯救 IT 人的一天 \(ithome.com.tw\)](#)
- [What are Autoencoders?. 簡單介紹 Autoencoder的原理，以及常見的應用。 | by Yu-Ru Tsai | Taiwan AI Academy | Medium](#)