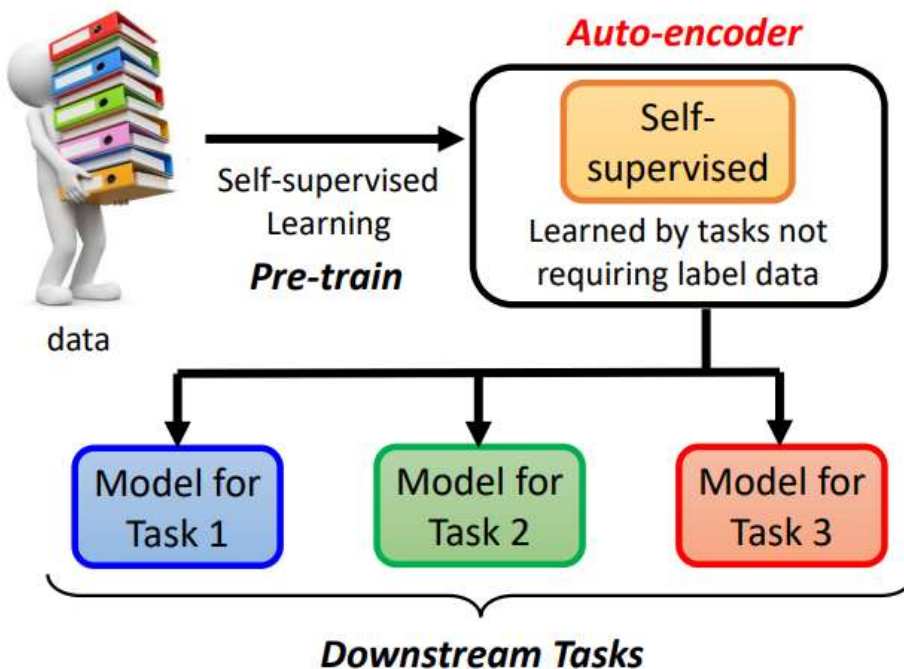


自編碼器 (Auto-encoder)

Create at 2022/06/25

- 自編碼器 (Auto-encoder)
 - Basic Idea of Auto-encoder
 - Feature Disentanglement
 - Discrete Latent Representation
 - More Applications
- 上課資源：
 1. 自編碼器 (Auto-encoder) (上) – 基本概念 (<https://www.youtube.com/watch?v=3oHlf8-J3Nc>).
 2. 自編碼器 (Auto-encoder) (下) – 領結變聲器與更多應用 (<https://www.youtube.com/watch?v=JZvEzb5PV3U>).
- 延伸資源：
 1. Unsupervised Learning - Linear Methods (https://www.youtube.com/watch?v=iwh5o_M4BNU).
 2. Unsupervised Learning - Neighbor Embedding (<https://www.youtube.com/watch?v=GBUEjkpoxXc>).

Self-supervised Learning Framework



- Auto-encoder 也算是 self-supervised Learning 的一環
- 有大量沒有 label 的資料，用這些資料可以去訓練一個模型
- 必須發明一些不需要 label 資料的任務，給模型進行學習
 - 填空題
 - 預測下一個 token
- 不用標註資料的模型稱為 self-supervised learning 或 pre-training
- 做微微的調整就可以用在其他下游的任務裡面

Basic Idea of Auto-encoder

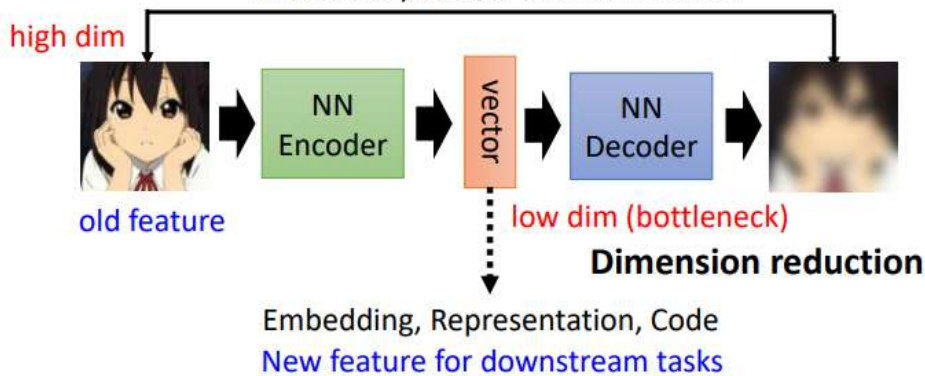
Auto-encoder

Unlabeled
Images



Sounds familiar? We have seen the same idea in Cycle GAN. 😊

As close as possible (reconstruction)



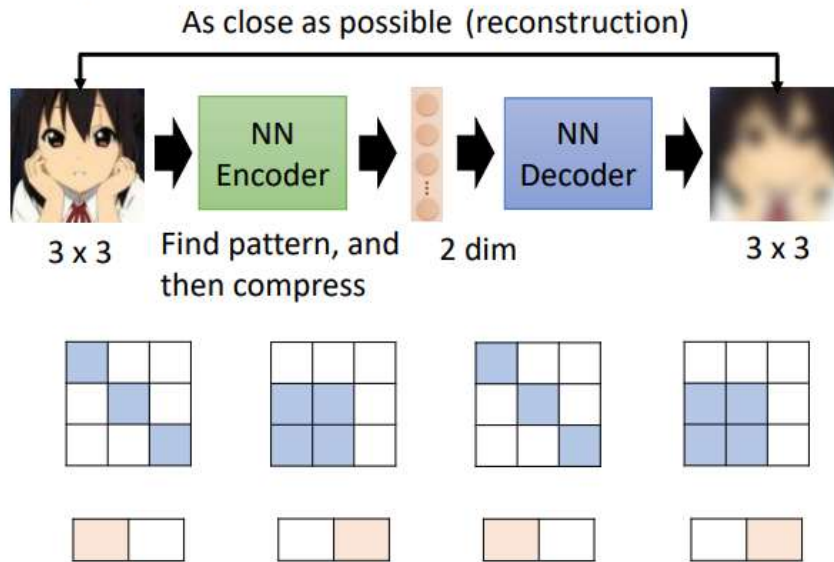
- 換成用影像當作例子
- 在 Auto-encoder 裡面有兩個很大的 network
 - Encoder
 - Decoder
- Encoder 讀進一張圖片，把這張圖片變成一個向量
- 向量會變成 Decoder 的輸入，Decoder 會產生一張圖片
- 所以 Decoder network 的架構可能會像是 GAN 的 Generator
- 訓練目標：
 - 希望 Encoder 的輸入跟 Decoder 的輸出越接近越好 (Reconstruction)
- 概念跟 Cycle GAN 一模一樣
- Auto-encoder 訓練過程完全不需要任何標註資料，只需要收集到大量的圖片，就可以做這個訓練，所以是一個 unsupervised learning 的方法
- 如何使用 Auto-encoder 的技術，怎麼把 train 完的 auto-encoder 用在 downstream 的任務呢？
 - 原來的圖片看作是一個很長的向量，通過 Encoder 輸出另外一個低維度向量 (Embedding, Representation, Code)
 - 再拿低維度的向量去做想做的事情

Why Auto-encoder?



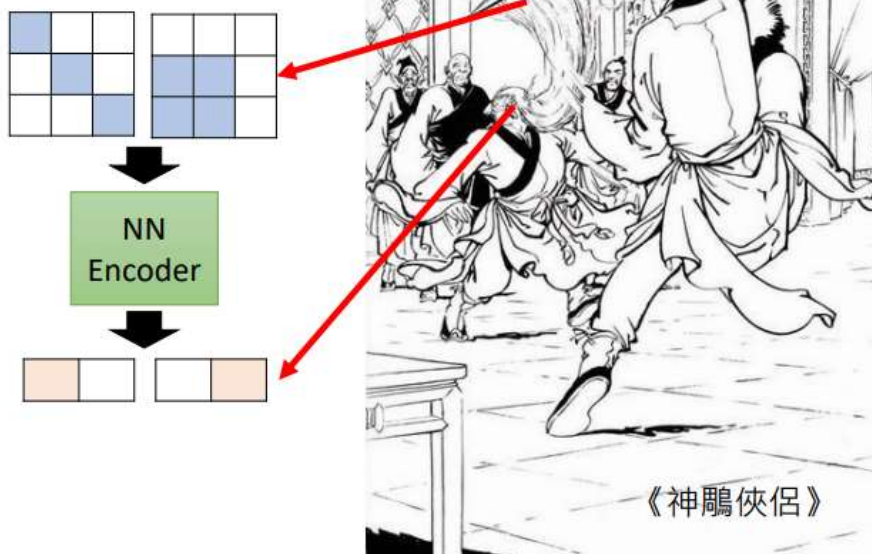
- Auto-encoder 好在哪裡？

Why Auto-encoder?



- Auto-encoder 要做的是把一張圖片壓所再還原回來

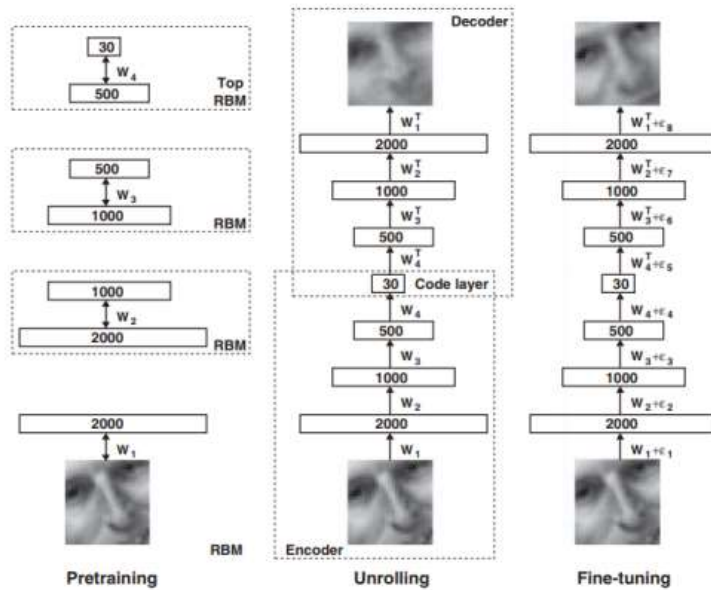
Why Auto-encoder?



- Encoder 做的事情就是化繁為簡
- 把複雜的東西用比較簡單的方法來表示

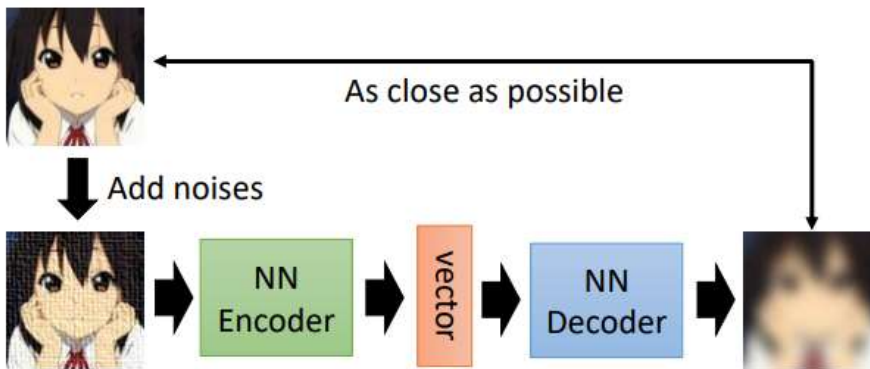
Auto-encoder is not a new idea

Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." *Science* 313.5786 (2006): 504-507



- Auto-encoder 不是一個新的想法

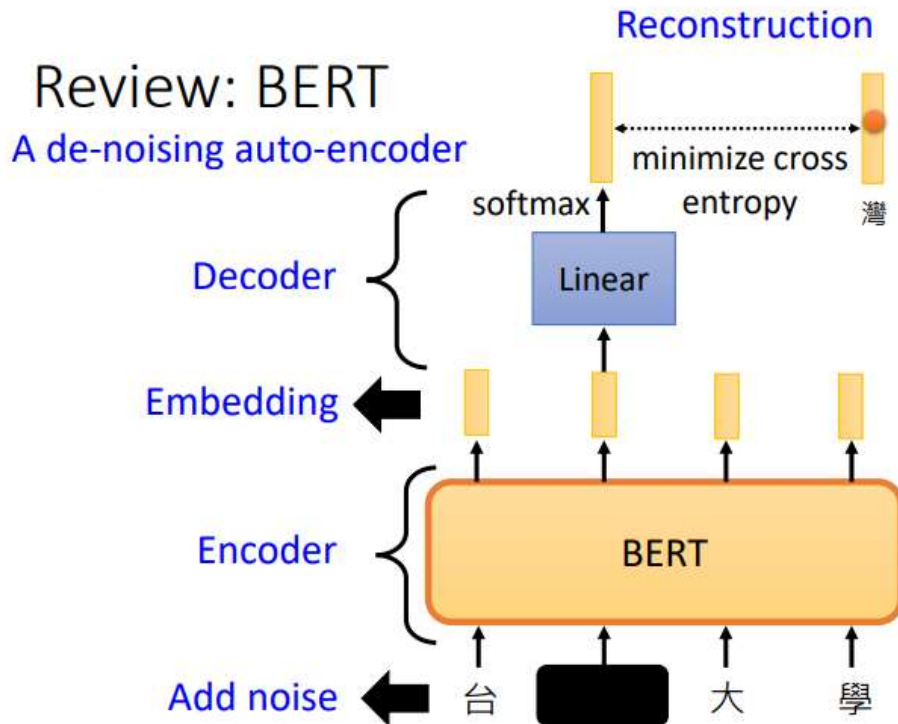
De-noising Auto-encoder



The idea sounds familiar? 😊

Vincent, Pascal, et al. "Extracting and composing robust features with denoising autoencoders." *ICML*, 2008.

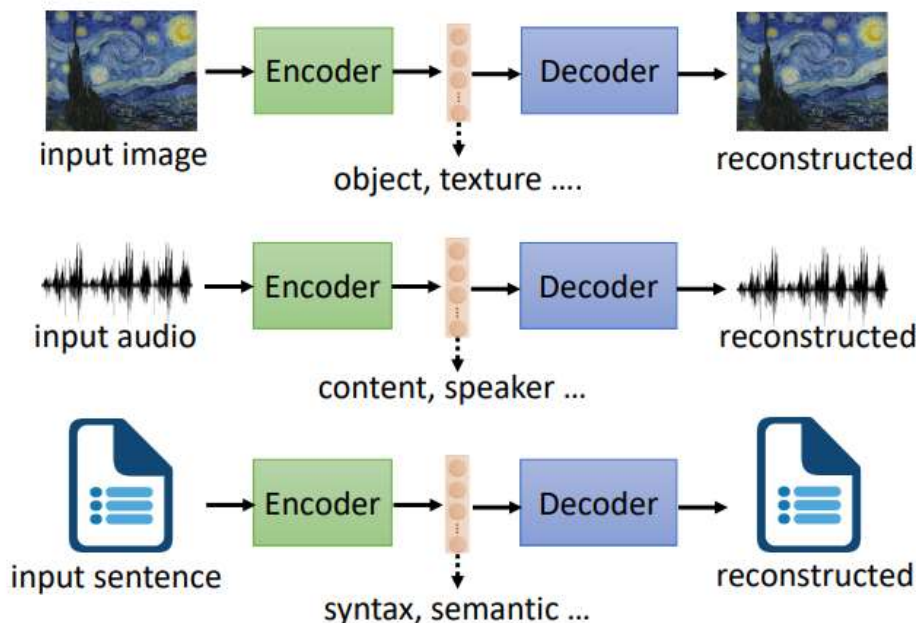
- Auto-encoder 一個常見的類型叫做 **De-noising Auto-encoder**
- 把原來要輸進 encoder 的圖片加上一些雜訊，再通過 encoder 跟 decoder，試圖還原原來的圖片
- 要還原的不是 encoder 的輸入，要還原的是加入雜訊之前的圖片
- 任務是：
 - 還要學會把雜訊去掉



- 可以把 BERT 看做是一個 De-Noising 的 Auto-encoder
- 輸入加上的 masking 其實就是 noise
- BERT 的模型就是 encoder，它的輸出就是 embedding
- 接著有一個 linear 的模型就是 decoder
- decoder 要做的就是還原原來的句子
- 所以 BERT 其實就是 De-Noising 的 Auto-Encoder

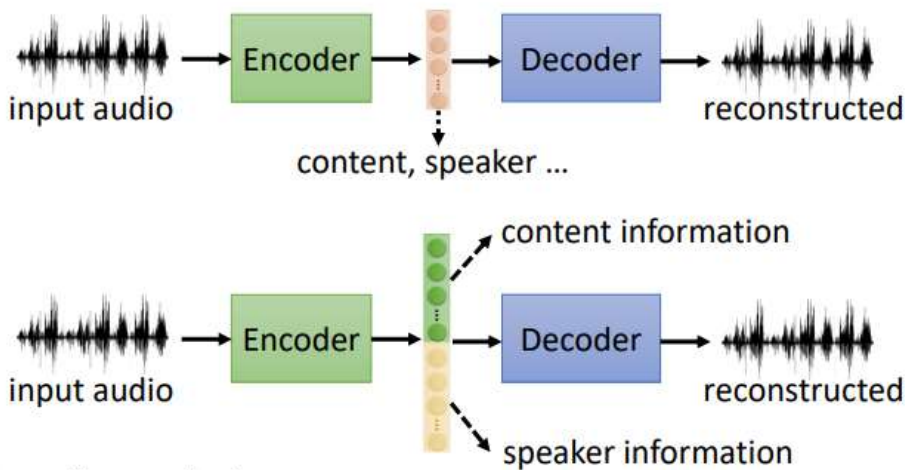
Feature Disentanglement

**Representation includes
information of different aspects**



- Disentanglement : 把一堆本來糾纏在一起的東西把它解開
- Auto-encoder 輸出的 vector 全部的資訊全部糾纏在一個向量裡面，我們不知道這些向量的哪些維代表哪些資訊

Feature Disentangle

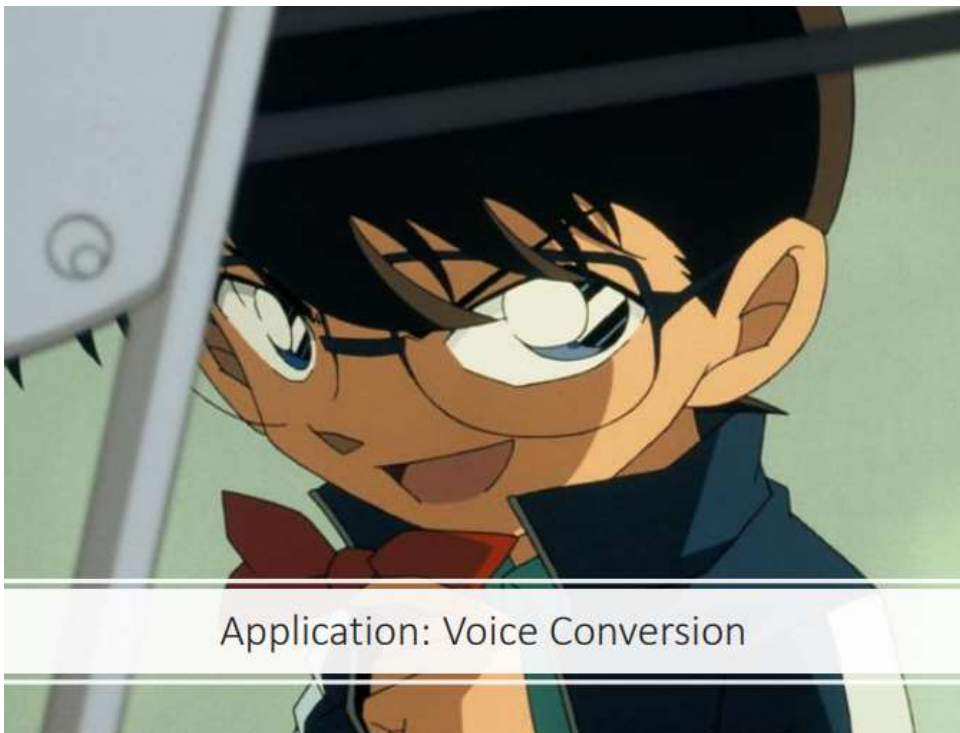


<https://arxiv.org/abs/1904.05742>

<https://arxiv.org/abs/1804.02812>

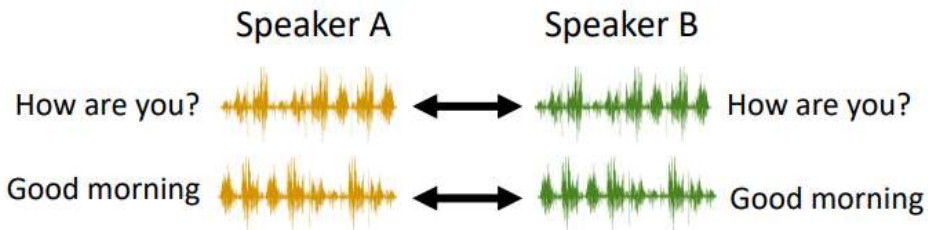
<https://arxiv.org/abs/1905.05879>

- Feature Disentangle 想做到的是：
 - 想辦法在 train auto-encoder 的時候，同時知道 embedding (representation) 的哪些維度代表哪些資訊



- Feature Disentangle 的應用：
 - Voice Conversion (語者轉換)

In the past



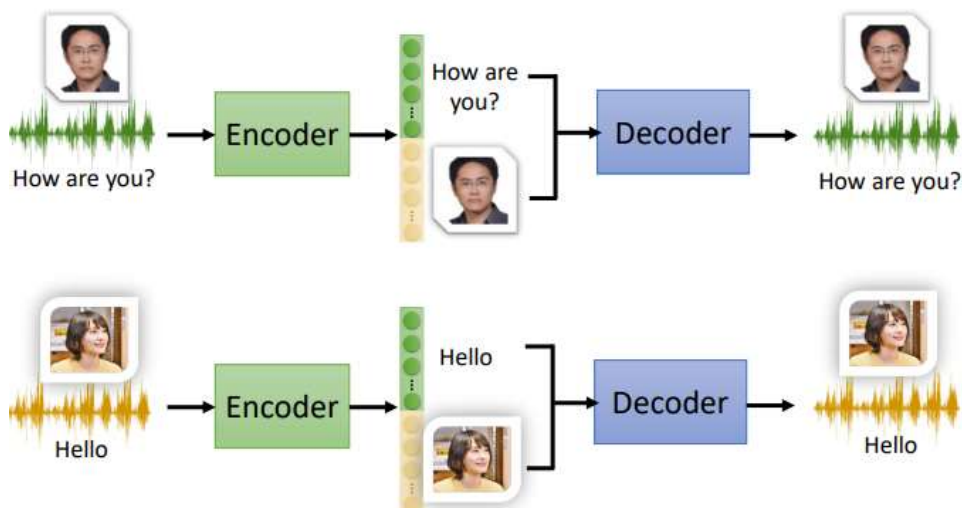
Today



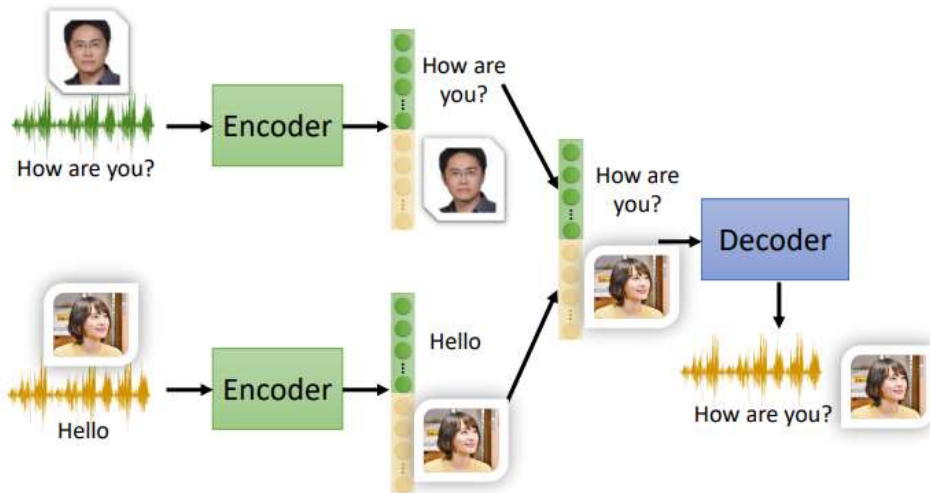
Speakers A and B are talking about completely different things.

- 現在 training 不需要兩個聲音都說一句話

Application: Voice Conversion

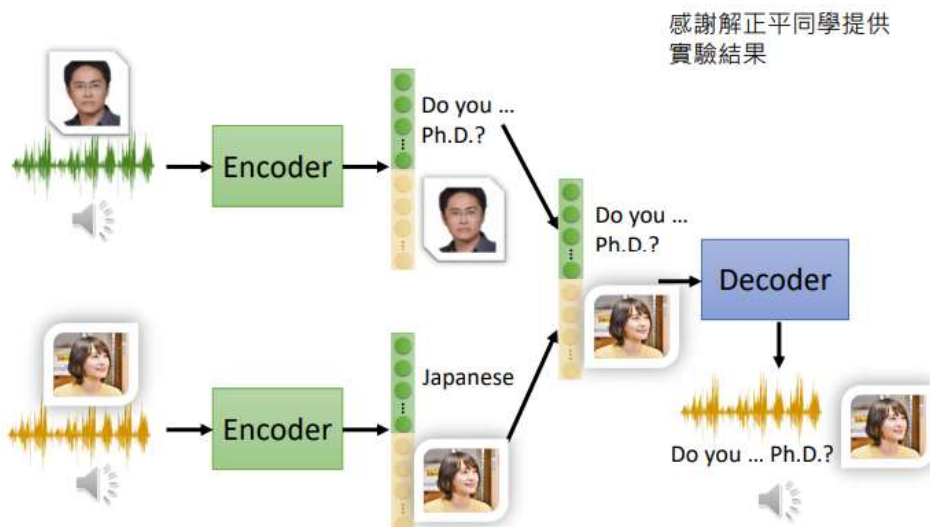


Application: Voice Conversion



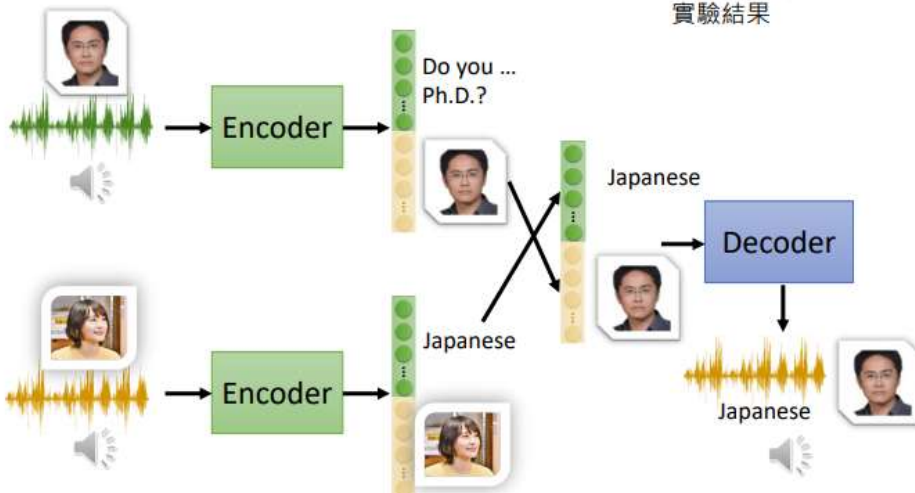
- 現在知道 embedding 的哪些片段代表哪些資訊
- 把需要的資訊取出來

Application: Voice Conversion



Application: Voice Conversion

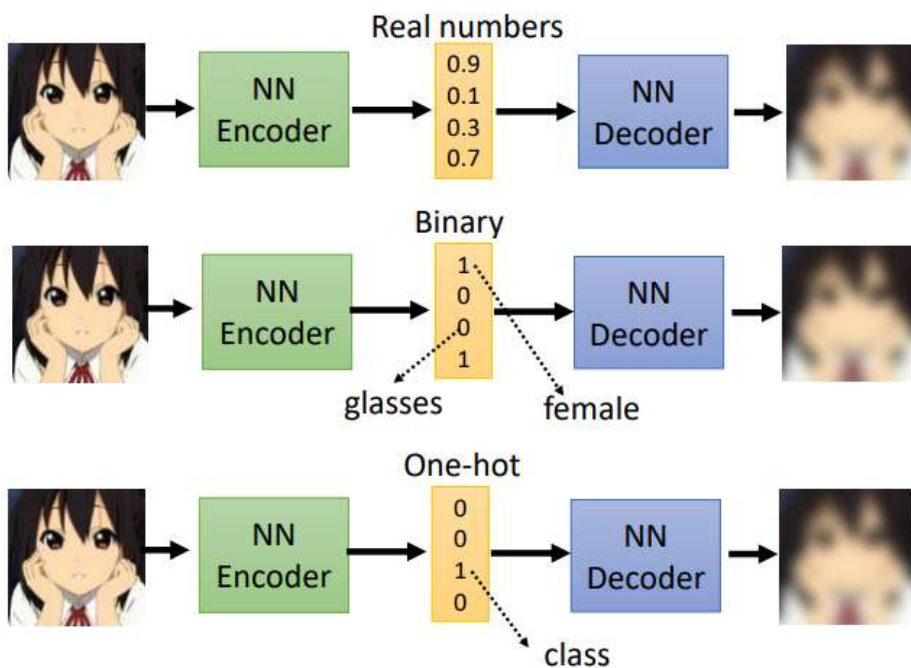
感謝解正平同學提供
實驗結果



- 把需要的資訊取出來之後拼起來
- 丟進 Decoder

Discrete Latent Representation

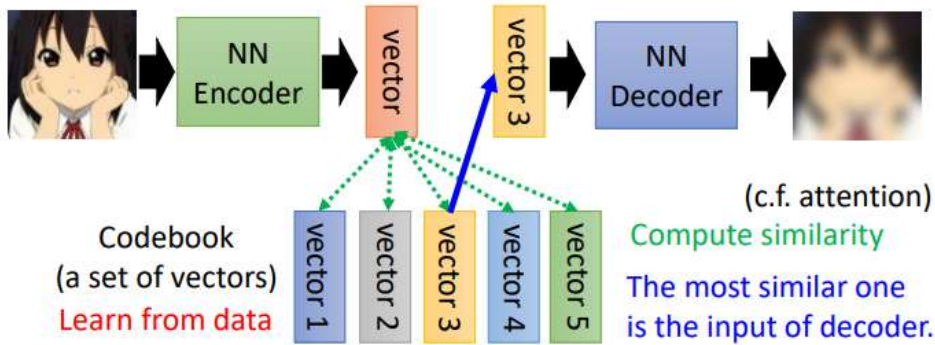
Discrete Representation



Discrete Representation

<https://arxiv.org/abs/1711.00937>

• Vector Quantized Variational Auto-encoder (VQVAE)



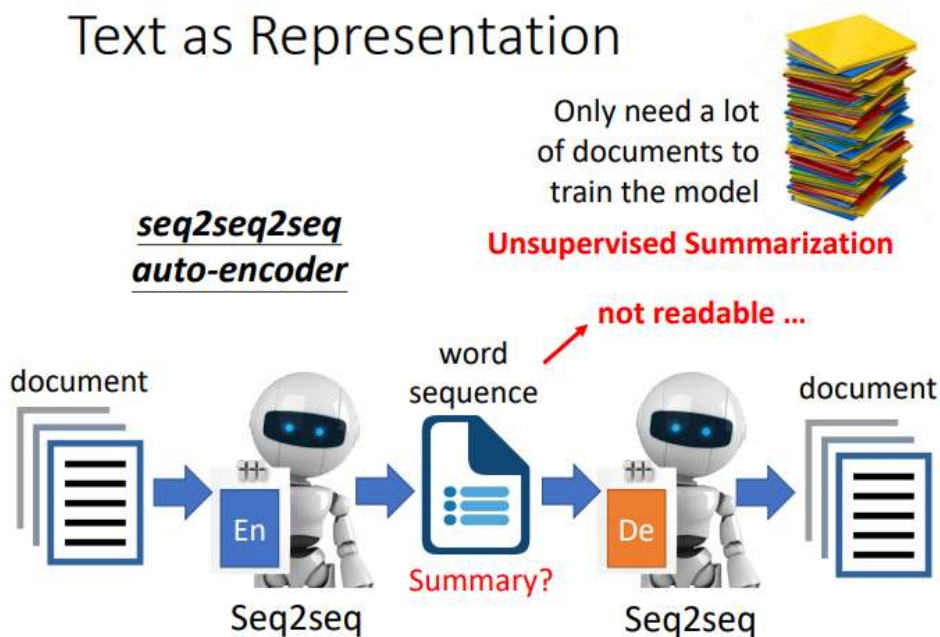
For speech, the codebook represents phonetic information

<https://arxiv.org/pdf/1901.08810.pdf>

- Discrete Latent Representation 最知名的是 VQVAE (Vector Quantized Variational Auto-encoder)
- 運作過程：
 - 輸入圖片到 Encoder 輸出一個向量
 - 有一個 codebook (a set of vectors)
 - 把 encoder 的輸入跟 codebook 都去算相似度，看誰的相似度最大
 - 把相似度最大的 vector 拿出來，丟到 decoder 輸出一張圖片
- codebook 可能可以學到最基本的發音部位 phonetic

<https://arxiv.org/abs/1810.02851>

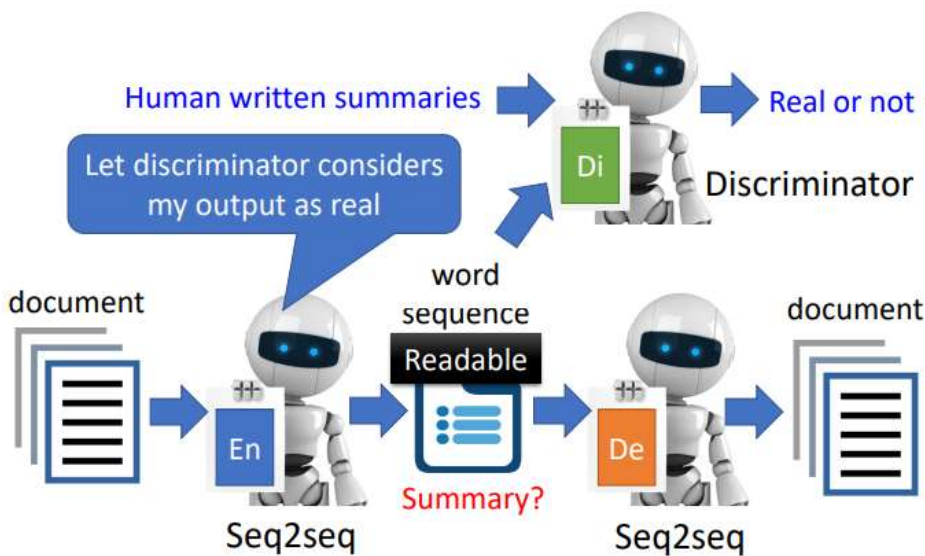
Text as Representation



- Representation 可以是一段文字
 - 把 embedding 變成一串文字，也許這串文字就是文章的摘要
 - 把文章丟到 encoder 裡面輸出一串文字，這串文字可以透過 decoder 還原回原來的文章
 - 但 encoder、decoder 需要是 seq2seq 的 model，ex. Transformer
 - 所以整個合起來是一個 seq2seq2seq auto-encoder
-
- 實際 train 起來會發現是行不通的
 - 因為 encoder 跟 decoder 之間會發明自己的暗號，產生的摘要人類看不懂的

This is cycle GAN 😊

Text as Representation



- 所以用 GAN 的概念加上一個 discriminator，discriminator 看過人寫的句子，所以知道人寫的句子長怎樣
- Encoder 要想辦法去騙過 discriminator (讓 discriminator 覺得是人寫的句子)，還要能讓 decoder 還原成原來的文章
- 其實就是 CycleGAN

感謝 王耀賢 同學提供實驗結果

Text as Representation

- **Document:** 澳大利亞今天與13個國家簽署了反興奮劑雙邊協議,旨在加強體育競賽之外的藥品檢查並共享研究成果
- **Summary:**
 - **Human:** 澳大利亞與13國簽署反興奮劑協議
 - **Unsupervised:** 澳大利亞加強體育競賽之外的藥品檢查
- **Document:** 中華民國奧林匹克委員會今天接到一九九二年冬季奧運會邀請函,由於主席張豐緒目前正在中南美洲進行友好訪問,因此尚未決定是否派隊赴賽
- **Summary:**
 - **Human:** 一九九二年冬季奧運會函邀我參加
 - **Unsupervised:** 奧委會接獲冬季奧運會邀請函

- Example (成功的例子)

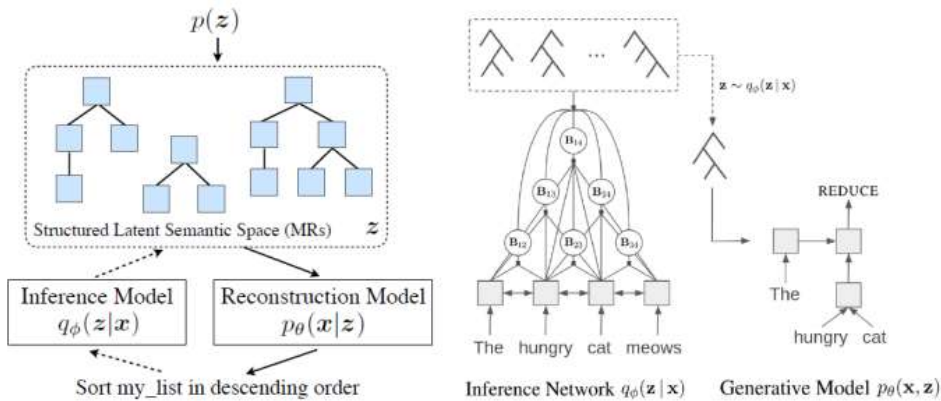
感謝 王耀賢 同學提供實驗結果

Text as Representation

- **Document:** 據此間媒體27日報道,印度尼西亞蘇門答臘島的兩個省近日來連降暴雨,洪水泛濫導致塌方,到26日為止至少已有60人喪生,100多人失蹤
- **Summary:**
 - **Human:** 印尼水災造成60人死亡
 - **Unsupervised:** 印尼門洪水泛濫導致塌雨
- **Document:** 安徽省合肥市最近為領導幹部下基層做了新規定:一律輕車簡從,不準搞迎來送往、不準搞層層陪同
- **Summary:**
 - **Human:** 合肥規定領導幹部下基層活動從簡
 - **Unsupervised:** 合肥領導幹部下基層做搞迎來送往規定:一律簡

- Example (犯錯的例子)

Tree as Embedding



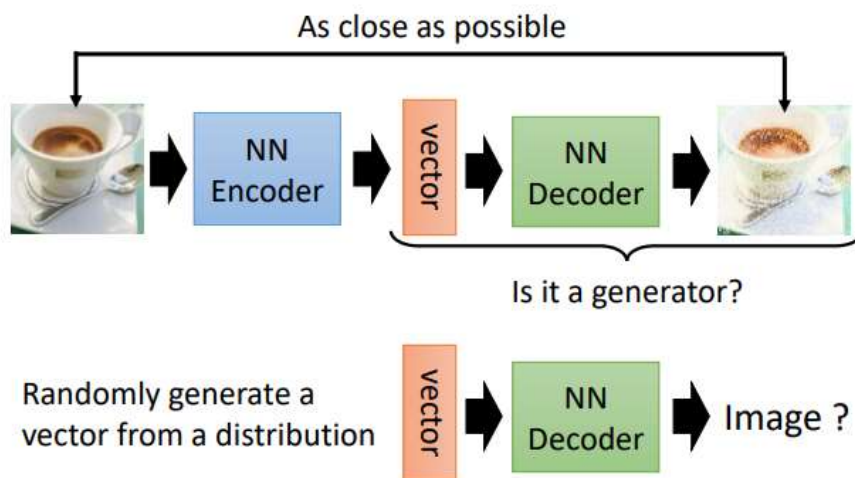
<https://arxiv.org/abs/1806.07832>

<https://arxiv.org/abs/1904.03746>

- 有 tree structure 當作 embedding 的例子

More Applications

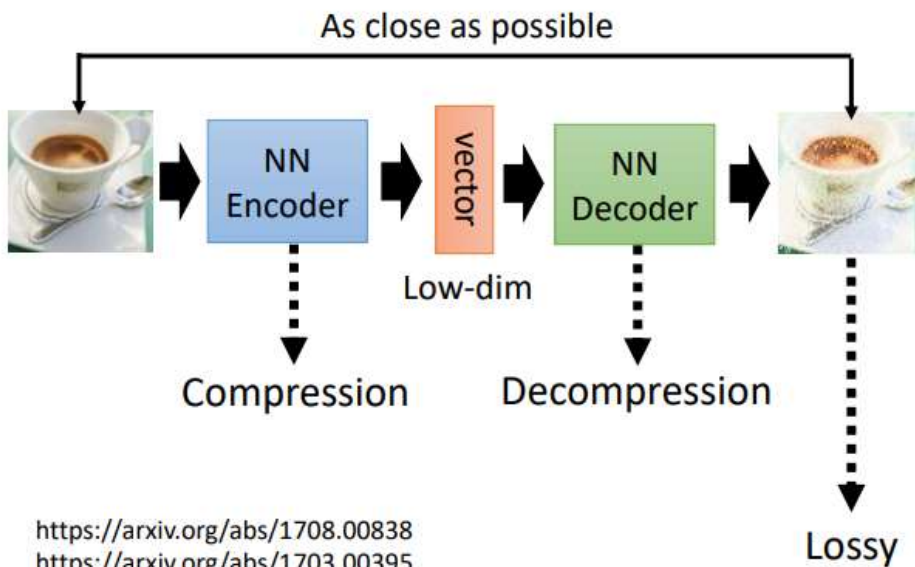
Generator



With some modification, we have **variational auto-encoder (VAE)**.

- 可以把 Decoder 當作一個 Generator 來使用
- 可以從一個已知的 distribution sample 一個向量丟給 decoder，看能不能輸出一張圖
- 除了 GAN 以外還有另外兩種 generative model
 - VAE (其實就是把 auto-encoder 的 decoder 拿出來，當作 generator 用)

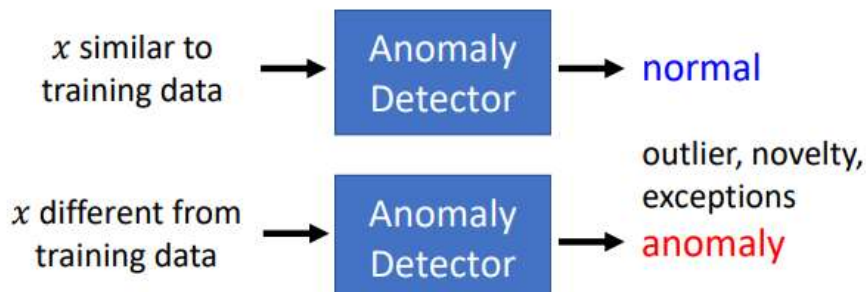
Compression



- auto-encoder 可以拿來做壓縮
- 可以把 encoder 的輸出當作是一個壓縮的結果，因為一張圖片是一個非常高維的向量，而 encoder 的輸出是一個非常低維的向量，可以把這個向量看作是一個壓縮的結果
- encoder 做的就是壓縮 (是 lossy 的壓縮，會失真)
- decoder 做的就是解壓縮

Anomaly Detection

- Given a set of training data $\{x^1, x^2, \dots, x^N\}$
- Detecting input x is *similar* to training data or not.



- 在作業會用 auto-encoder 做 Anomaly Detection (異常檢測)
- 假設有一堆訓練資料 $\{x^1, x^2, \dots, x^n\}$
- 輸入一筆資料，判斷它跟我們之前在訓練資料裡面看過的資料相不相似
- 有一個異常檢測的系統，是透過大量已經看過的資料訓練出來
 - 輸入一筆資料，如果看起來像是訓練資料裡面的 data，輸出 normal
 - 如果看起來不是訓練資料裡面的 data，輸出 anomaly (outlier, novelty, exceptions)

Anomaly Detection

Training Data:



Training Data:



Training Data:



- Example

Binary Classification?

Anomaly Detection **We only have one class.**
Training auto-encoder

• Fraud Detection

- Training data: credit card transactions, x : fraud or not
- Ref: <https://www.kaggle.com/ntnu-testimon/paysim1/home>
- Ref: <https://www.kaggle.com/mlg-ulb/creditcardfraud/home>

• Network Intrusion Detection

- Training data: connection, x : attack or not
- Ref: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>

• Cancer Detection

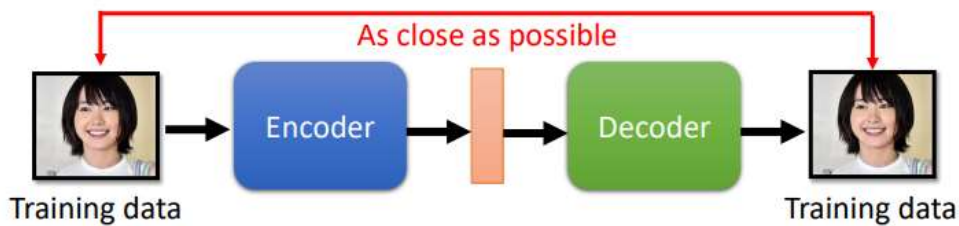
- Training data: normal cells, x : cancer or not?
- Ref: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data/home>

- 常見應用
 - 詐欺偵測
 - 網路的侵入偵測
 - 癌症檢測
- 不是一般的分類問題，叫做 one class 分類問題
 - 不容易收集到異常的資料

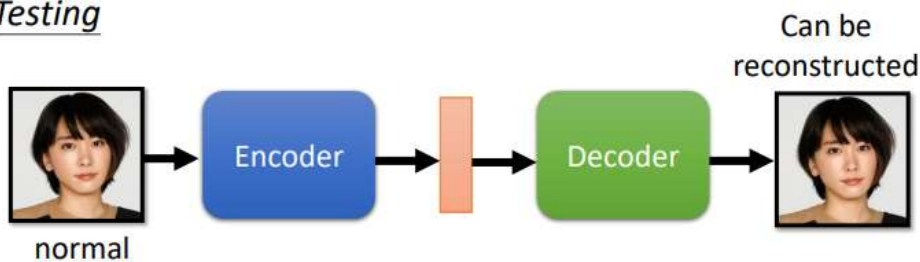
Approach: Auto-encoder

Training

Using **real human faces** to learn an autoencoder



Testing

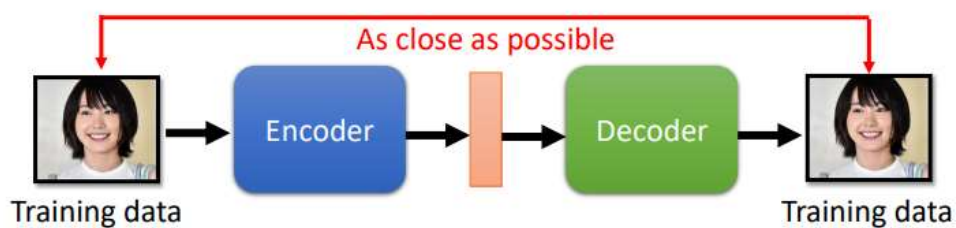


- 因為已經看過，所以可以順利還原回來

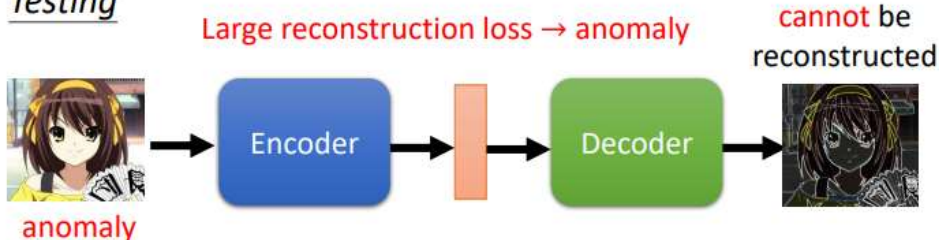
Approach: Auto-encoder

Training

Using **real human faces** to learn an autoencoder



Testing



- 因為沒有看過，所以 decoder 很難還原回來
- 如果計算輸入跟輸出的差異非常大
 - 代表輸入 encoder 的照片，可能是一個異常的狀況 (訓練的時候沒有看過)

tags: 2022 李宏毅_機器學習