Adversarially Learned One-Class Classifier for Novelty Detection

統研碩一 7107018013 郭又嘉

Novelty detection

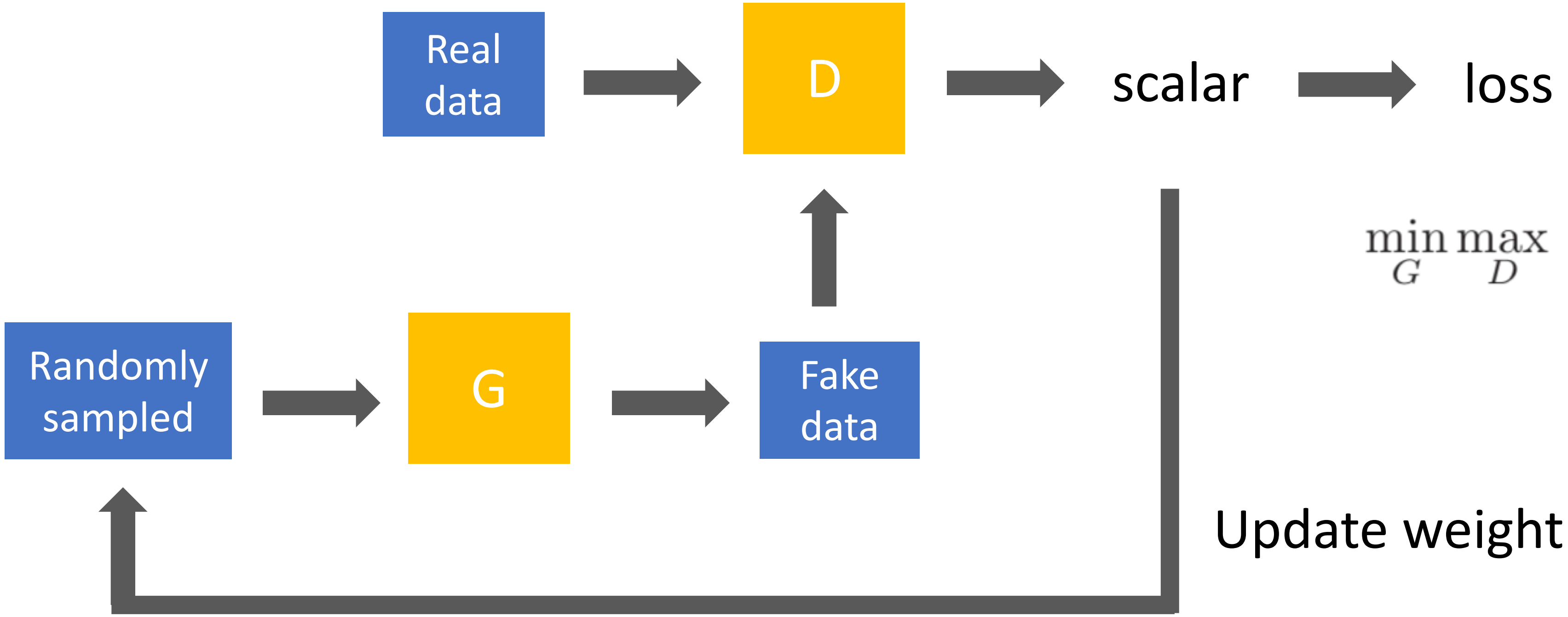
- 困難：
  ① The novelty class is often **absent during training**, poorly sampled or not well defined.
  ② Due to the unavailability of data from the novelty class, training an **end-to-end** deep network is a cumbersome task.
- 解決：
  - end-to-end architecture for one-class classification

輸入是原始數據
輸出是最後結果

①

對僅包含目標的數據訓練
而在所有數據中辨識該目標

Generative Adversarial Networks



$$\min_G \max_D \left(\mathbb{E}_{X\sim p_t}[\log(D(X))] + \mathbb{E}_{Z\sim p_z}[\log(1-D(G(Z)))]\right).$$

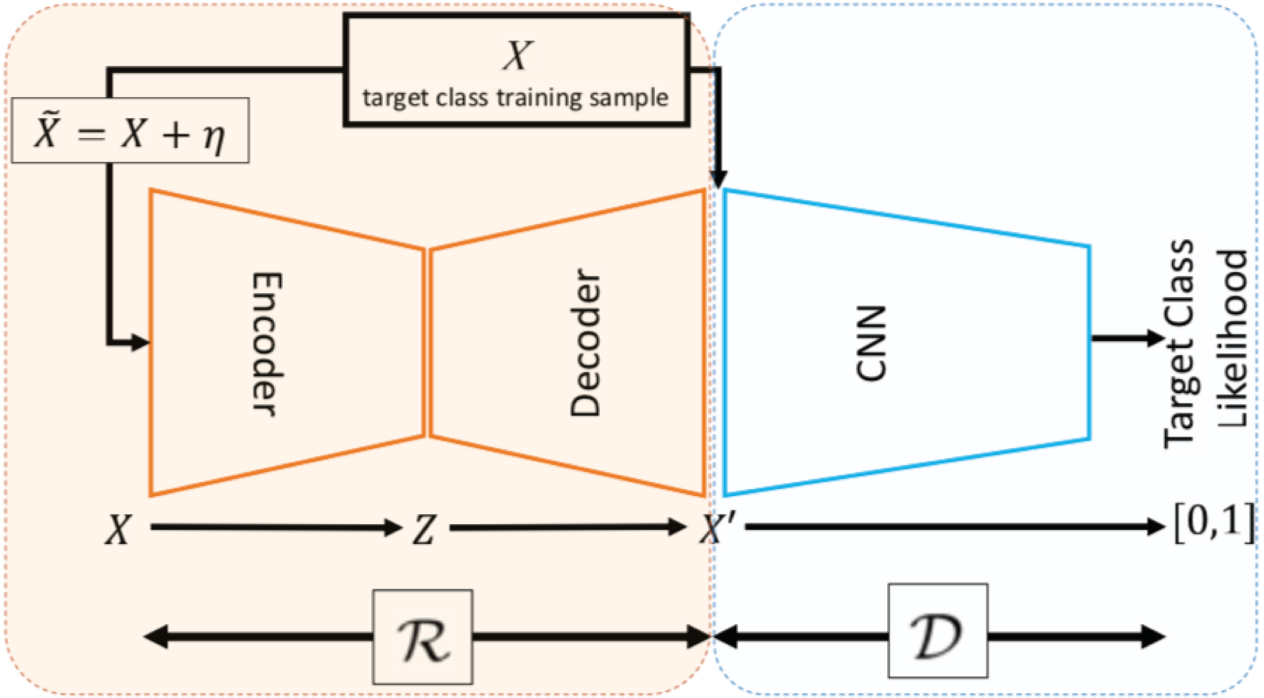
G努力做出逼近真實的假資料
D努力分辨真實與假資料的差別

相互對抗

G產生最真實的資料（GAN目的）
D可精確分別真假（paper重點：異常偵測）

Approach

- Network R：Reconstruction
- Network D：Detection
- X：目標類別的資料



$$\min_{\mathcal{R}} \max_{\mathcal{D}} \Big(\mathbb{E}_{X\sim p_t}[\log(\mathcal{D}(X))] + \mathbb{E}_{\tilde{X}\sim p_t+\mathcal{N}_\sigma}[\log(1-\mathcal{D}(\mathcal{R}(\tilde{X})))]\Big)$$

$$\tilde{X} = (X \sim p_t) + (\eta \sim \mathcal{N}(0, \sigma^2\mathbf{I})) \longrightarrow X' \sim p_t$$

Maximize $\mathcal{D}(\mathcal{R}(X \sim p_t; \theta_r)).$

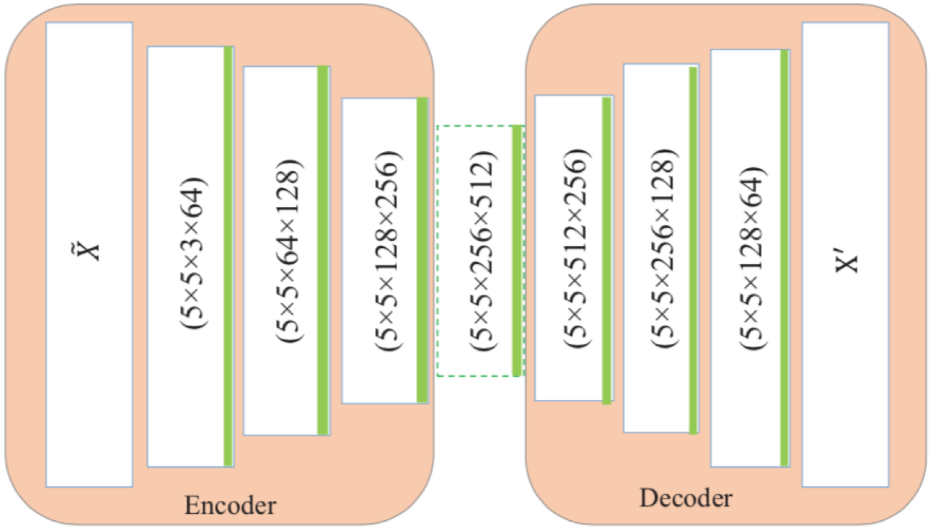
Loss function $\mathcal{L} = \mathcal{L}_{\mathcal{R}+\mathcal{D}} + \lambda\mathcal{L}_{\mathcal{R}}$

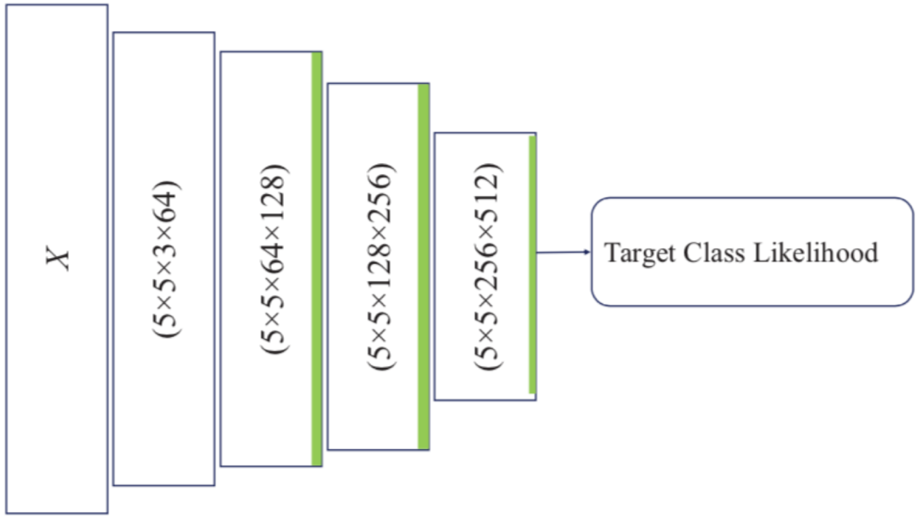
$$\mathcal{L}_{\mathcal{R}} = \|X - X'\|^2.$$

R+D Network Architecture

$$\mathrm{OCC}_2(X) = \begin{cases} \text{Target Class} & \text{if } \mathcal{D}(\mathcal{R}(X)) > \tau, \\ \text{Novelty (Outlier)} & \text{otherwise.} \end{cases}$$

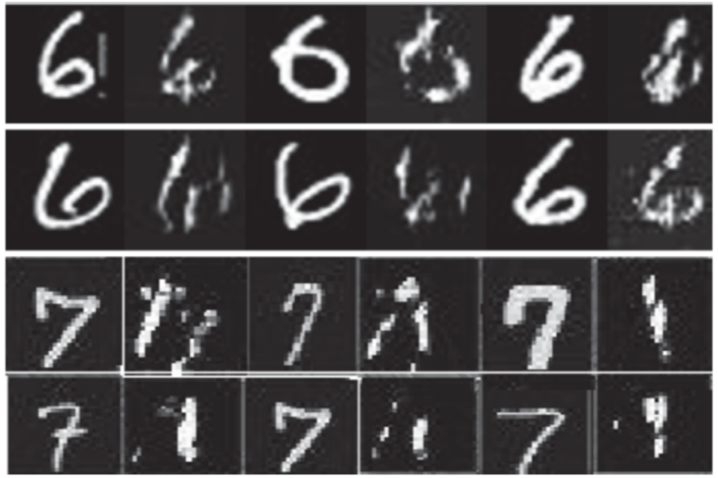
- Autoencoder⟶uses the reconstructed image to train another network for the discrimination task
- R 利用目標類別做訓練，若有異常輸入，就難重建。
- D輸出一個scalar：輸出結果是介於0～1的分數。



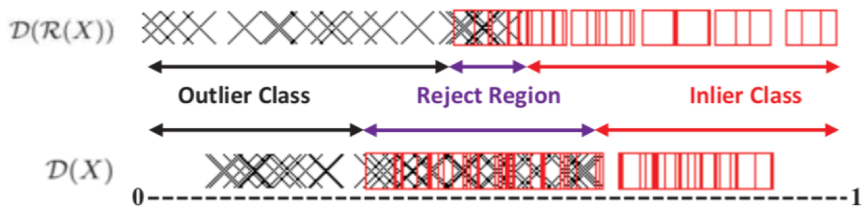


Experiment Results

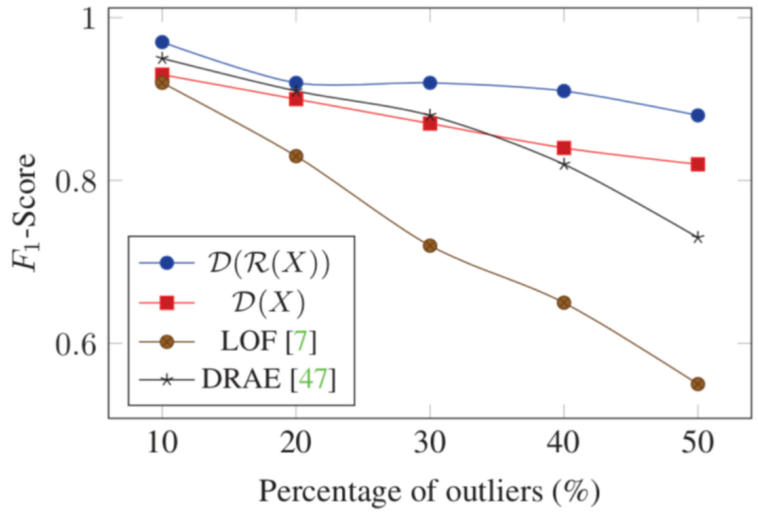
- Data : MNIST
- Each of the ten categories of digits is taken as the target, and we simulate outliers by randomly sampling images from other categories with a proportion of 10% to 50%.



目標要檢測數字1
用數字6、7當作異常輸入



有加 R Network的可讓異常值更分離



用F1-score來評估
異常值的比例增加，模型還是很穩健
檢測能力很好