國立台北大學資訊工程學系專題報告

基於對比學習之入侵偵測系統

Intrusion Detection System Based on Contrastive Learning

專題組員:歐陽平、陳孟舜、林致諺、謝佳恩

專題編號: PRJ-NTPUCSIE-113-008

執行期間：2024 年 7 月 至 2025 年 5 月

一、 摘要

本專題實作基於對比學習之入侵偵測系統，透過自行設計的模型架構、數據轉換格式與簡易的使用者介面，以將模型能用於實際的入侵偵測。

二、 簡介

深度學習在分類上取得巨大的成功，利用深度學習進行網路入侵監測已成為巨大的趨勢，然而傳統入侵偵測模型往往需要大量的標籤資料進行訓練，然而由於網路攻擊樣本的標注成本高昂，而無標注資料易於獲取，本研究旨在探索如何利用對比學習的特性，通過大量無標籤資料與少量有標籤資料，提升模型在少量標籤資料下的分類性能。而隨者深入研究發現，近來對比學習的發展也從無標籤的自監督學習[1]發展到有標籤的監督式學習。讓對比學習在少量資料與大量資料上都有實用的方法能應用於入侵偵測。

對此本研究開發一套結合深度學習的入侵偵測系統，運用自行開發的架構、資料處理方法、製圖方法、資料篩選方法。透過架構與資料優化的方式，使其能夠以低成本的方式運行在嵌入式裝置上。並透過即時的數據傳輸分析與模型偵測，將流量進行分類，以便檢測訪問數據是否為安全流量，並與相關的對比學習論文進行比較，展現我們模型的優勢。

三、專題進行方式

本專題的核心目標是利用訓練好的對比學習模型應用於實際的入侵偵測中。架構會分為模型訓練架構與實作運行架構。

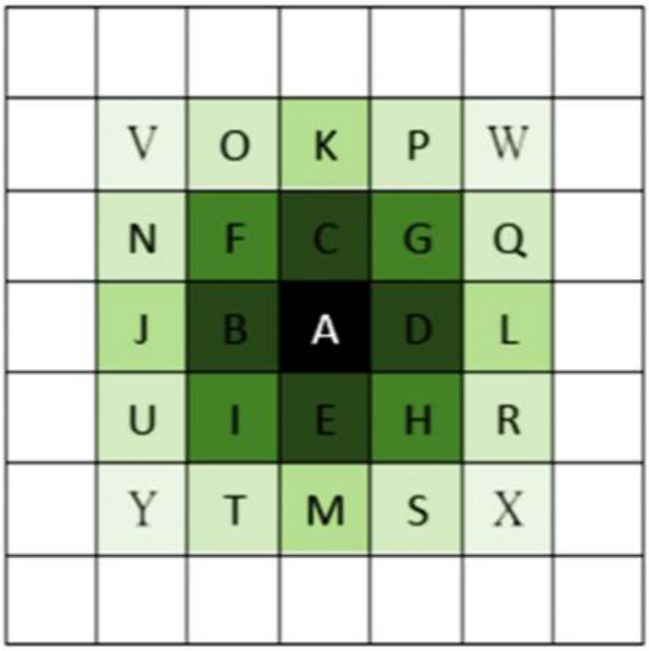
3.1 模型訓練架構

3.1.1 資料前處理

我們使用的是 UNSW_NB15 的官方訓練集與測試集。由於資料類型屬於表格資料(Tabular Data)，無法應用於輸入為圖片的對比學習模型。為此，首先刪除資料中無關特徵並將 NaN 無效值替換成 0，再將非數字的特徵進行 Label encoding，以讓模型能夠進行數學運算。再來透過 Quantile Transformer 將資料分布變成常態分布以對抗偏態分布，再將數值轉成 0 至 255 的區間，對應到圖片的像素值。

3.1.2 圖片製作

這裡我們提出一種基於相關係數的製圖方法，首先計算每個特徵之間的相關係數製成相關係數矩陣。對每個特徵找出與其相關係數前五高的特徵並進行總和，已獲得關係分數。將每個特徵的關係分數進行比較，關係分數最高的特徵會被放置於圖片正中心，如(圖一)的 A。其後 B、C、D、E 則是與 A 相關係數前四高的特徵。再來尋找與 B、C 關係分數最高的特徵填入 F。以此類推填入所有的特徵，其餘填入 0 表示空白值。(圖一)為製圖的示意圖，圖二為製圖方法。



（圖一）製圖示意圖

Algorithm 1 Transform Network Traffic to Images

Input: T= $\{t_1, t_2, \ldots, t_N\}$, $t_i \in \mathbb{R}^F$ represents the i-th network traffic sample with F features.

Output: X = $\{x_1, x_2, \ldots, x_n\}$, x_i represents the image generated from t_i.

for t_i in T **do**
  $t_i' = \text{Normalize}(t_i) \cdot 255, \forall i \in \{1,2,\ldots,N\}$
end for
T′ = $\{t_1', t_2', \ldots, t_N'\}$,
$C_{jk} = \text{Correlation}(t_j', t_k')$, $C \in \mathbb{R}^{FxF}, \forall j,k \in \{1,\ldots,F\}$
Rank all features based on total correlation strength:
  $R(j) = \sum_{k=1}^{F} |C_{jk}|, \forall j \in \{1,2,\ldots,F\}$
Sort all features by R(j) in descending order:
RankedFeatures = $\{f_1, f_2, \ldots, f_F\}$ # f_1 has highest R(j)
LayoutPositions = SpiralPositions(F) # A list of F coordinates in spiral order starting from center
for i from 1 to F **do**
  $P(f_i)$ = LayoutPositions[i] # Assign i-th ranked feature to i-th spiral position
end for
for t_i' in T′ **do**
  $x_i = \text{Image}(t_i', P)$
end for
Return The set of images $X = \{x_1, x_2, \ldots, x_n\}$

（圖二）製圖方法

3.1.3 資料增強

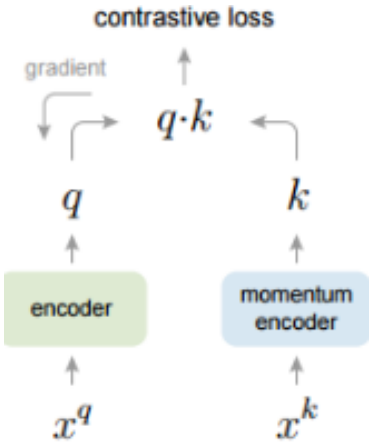
在自監督對比學習中我們會使用資料增強的方式形成正樣本對，透過兩張本質一樣，但是透過對比損失拉近正樣本對的距離，拉遠負樣本對的距離學會「兩張圖片是一樣的」。在此我們比較一些資料增強方法如：高斯模糊、翻轉、隨機裁剪、色彩擾動以及我們設計 Masking 方式。

我們設計的 Masking 方式分成中心區域與外圍區域，如我們的製圖方式所示，中間區域式關聯度高的特徵，也能認為是相對重要的特徵，而周遭特徵則為相對不重要的特徵。遮蔽少部分中央區域的與較多的周遭特徵以達到造成一定的對比差異但不至於損壞整張圖片的語意，相較於隨機對特徵 masking，更有策略的達到更好的效果，也不會造成圖片資訊流失過多。

3.1.4 模型訓練

模型部分我們有分為半監督[4]訓練與有監督訓練。

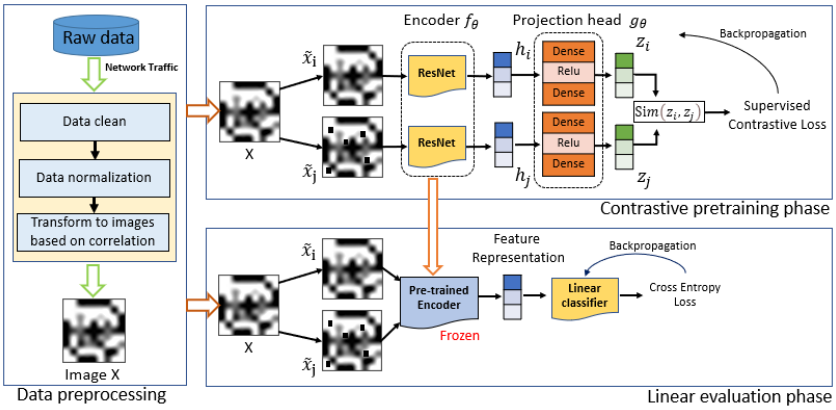
半監督訓練部分，會將訓練集分成兩個部分，分別是大量的無標籤資料與少量的有標籤資料。該部分我們使用 MoCo 對比學習方法，並以 Restnet-18 作為 encoder 架構來進行訓練，MoCo[3]透過 Momentum encoder 的方式解決負樣本不穩定的問題，其架構如(圖三)。



(圖三)MoCo 架構

將大量無標籤資料進行自監督對比學習計算，並透過將自身資料增強後的樣本作為正樣本，其他樣本作為負樣本，以學習圖片中的語意，將對比學習訓練完的預訓練模型進行少量有標籤資料的 finetuning 做半監督學習，取得最後的決策模型。架構圖如(圖四)

自監督訓練部分則不會對訓練集進行分割，而會將我們的訓練集重複使用，分別用於第一階段的對比學習自監督訓練與第二部分的 finetuning。有監督對比學習一樣使用 Resnet-18 作為 encoder，與自監督差別在於不是正樣



(圖四)半監督學習架構

本與負樣本的定義是透過 labels 作為區分。在對比學習階段中將大量有標籤資料進行對比學習訓練以得到學習到圖片知識的預訓練模型。在將同樣的資料進行第二輪的 finetuning，最後得到我們的決策模型。架構圖如(圖五)

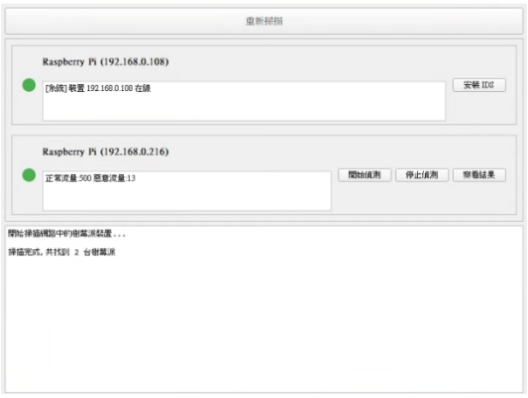
3.1.5 測試結果

使用訓練完成的決策模型對測試集進行預測，獲取準確度、F1-score 等等數據以評斷模型效能。

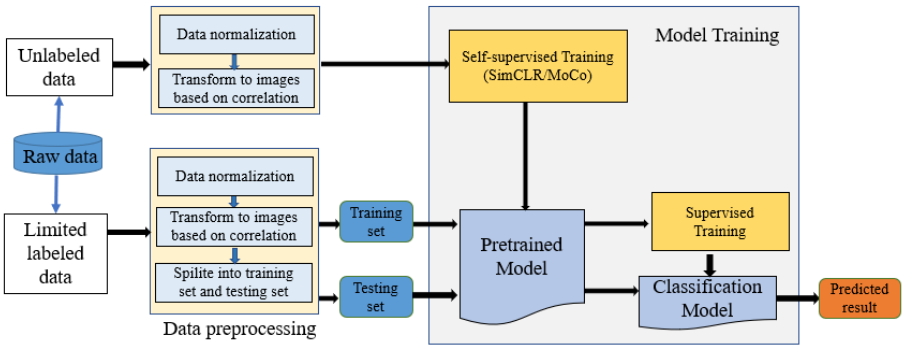
3.2 實作運作架構

實作部分我們使用半監督對比學習訓練完的決策模型進行實作，並使用 Raspberry Pi 5b 作為我們系統的載體。由於只需少許有標籤資料就有不錯的偵測效果，其後也能結合聯邦學習進行系統升級，展現其在邊緣運算裝置上的實用性。

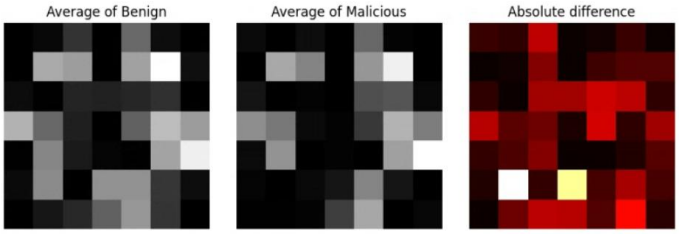
該系統開啟後會先偵測網路中的邊緣運算裝置，偵測完畢後會檢查裝置中是否有偵測系統，沒有的話點擊「安裝 IDS」即可安裝系統。安裝完成後點選「開始偵測」就會運行模型進行偵測。在此我們透過將原本 UNSW-NB15 封包打進安裝系統的邊緣運算裝置中進行偵測，並及時回傳當前的偵測狀況。並會將偵測到的攻擊流量進行儲存，同時儲存同量的正常流量，以便進行展示與後續的模型更新。按下「停止偵測」後即可停止偵測。最後按下察看結果能查看正常流量與攻擊流量的型態與其差異。圖六為偵測系統，圖七為偵測結果。



(圖六)偵測系統



(圖五)監督式學習架構



(圖七)偵測結果

三、 訓練結果比較

在本研究中，我們將資料依據模型的訓練方式，區分為半監督式模型與監督式模型兩類進行分析與比較。兩者所使用的數據集將明確區分，以確保比較的準確性與公平性。此外，為了評估本研究模型的效能與實用性，我們也將本研究的結果與其他主流模型進行對照，並納入近年來相關領域的重要學術論文進行深入比較分析，藉此凸顯本研究方法的優勢與潛在應用價值。

4.1 半監督訓練

4.1.1 與其他模型比較

Labeled data	Method	Acc
5000	MoCo (Ours)	91.94%
5000	DNN	90.10%
5000	Resnet18	89.05%
5000	CNN	89.02%

(表一)半監督比較

在此比較中我們與 DNN、Resnet18、CNN 在同樣標籤數量的情況下進行比較，可以發現我們在使用半監督學習時，可以在只有少量標籤資料的情況下有更高的準確率與辨識能力。

4.1.2 與相關論文比較

Model	Acc	Method
BYOL[4]	89.97%	無須負樣本之對比學習
TCG-IDS[5]	91.48%	結合 GCN 之對比學習
Ours	91.94%	MoCo 為基底之半監督對比學習

(表二)半監督相關論文比較

BYOL 的特點是不需要負樣本就能進行對比學習，只靠正樣本來學習，避免學習的不穩定性，但由於其只依靠正樣本會造成其訓練緩慢。而 TCG-IDS 則將流量序列建模為圖結構，強化節點之間得關聯性使用 GCN 進行度對比學習，然而這樣的處理方式會增加資料轉換的成本與訓練時間。

在此我們與其他半監督式對比學習的相關論文 BYOL 與 TCG-IDS 比較，我們的方法相較於 BYOL 有更高的準確率，且由於 BYOL 不參考負樣本，需要更長的訓練時間，而我們能夠在相對少的時間中取得更好的準確率。雖然對比 TCG-IDS 的準確率沒有太大提升，但由於其特殊的資料架構與模型複雜度，會大大增加訓練時間，而我們的方法能夠在更短的時間內達到相同的準確度，於實際應用上也有更好的偵測效率

4.2 監督式訓練

4.2.1 與其他模型比較

Model	Acc	F1-Score
XGBoost	91.18%	93.28%
RF	90.24%	92.36%
CNN_1D	89.67%	91.91%
SVM	88.14%	90.68%
DNN	87.47%	89.99%
Ours	93.18%	95.20%

(表三)有監督比較

在此我們與多種機器學習方法進行比較，其中能看到對於 XGBoost、RF、CNN_1D 等等機器學習方法，我們的監督式對比學習不管是在準確率還是 F1-Score 上都有更好的表現

4.2.2 與相關論文比較

Model	Acc	Method
Yue et al.[6]	92.99%	使用封包遮蔽資料增強之有監督對比學習
Lopez-Martin et al.[7]	88.80%	結合隨機傅立葉特徵之有監督對比學習
Ours	93.18%	結合 Masking 之監督式對比學習

(表四)有監督相關論文比較

Yue 等人的論文透過結合對比損失與預測損失的方式強化訓練效率，並且提出眼閉封包的方式進行資料增強，讓其能有更好的訓練效果並提高模型的魯棒性。而 Lopez-Martin 等人的論文透過將特徵與標籤透過隨機傅立葉特徵映射到同一個向量空間，使其能在空間中拉近正樣本，拉遠負樣本距離，使模型更有解釋力並完成對比學習訓練。

在此我們與其他監督式對比學習的相關論文 Yue 等人與 Lopez-Martin 等人的論文比較，我們採取並修改 Yue 等人 masking 的資料增強方式，並取得比 Yue 等人的論文更高的準確率。而相較於 Lopez-Martin 等人的論文，我們的方法更體現了在監督式對比學習上，我們對於資料的處置與增強方法於網路流量的辨識更有效果。

4.3 資料增強效果

Augmentation	Acc	Effectiveness
Feature Masking	**88.61%**	**High**
Random Horizontal Flip	**88.24%**	**High**
Gaussian Blur	**88.15%**	**High**
Random Vertical Flip	87.75%	Low
Color Jitter	87.07%	Low
Random Resized Crop	86.71%	Low

(表五)資料增強比較

在此我們測試多種資料增強方式對於對比學習的影響，經過測試能發現翻轉、高斯模糊與 masking 有較好的效果，而隨機添加像素、隨機裁減等等會不利於網路流量對比學習。在此我們能夠發現對於網路流量的資料來說，我們可以對特徵進行模糊翻轉等等不流失資料且保持其結構的資料增強方式能讓我們有有效的對比學習。但對於裁減、添加像素等等破壞圖片結構與語意的資料增強方式不利於用於網路流量的對比學習上。

五、結語與展望

學術方面，我們提出了基於相關係數的網路流量製圖方法，提出了基於我們製圖方法的 masking 方式，對對比學習於 IDS 領域提供些許貢獻，也顯示了對比學習能夠有效運用於 IDS 領域。此專題結合的半監督與監督式學習方法，不管有標籤資料的多寡，都能有有效的對比學習方式去應對，提升對於環境的適應能力。而且透過圖型化表格資料，能使艱澀難懂的網路流量能以圖片的方式展現，更為直觀且容易理解，能一眼看出正常流量與攻擊流量的差異。未來希望能結合聯邦學習，在不同裝置應對不同攻擊時能透過聯邦學習互相整合，升級 IDS 系統，提高模型偵測準確率，並打造更完善的應用系統，能讓基於深度學習的 IDS 系統更貼近使用者。

六、銘謝

衷心感謝指導教授在本專題研究期間給予的悉心指導與寶貴建議，以及慷慨提供實驗室的各項資源與設備。無論是在研究方向的確立、方法論的選擇，抑或是算法優化的關鍵環節，教授專業而深入的見解都為本研究提供了不可或缺的指引；同時，特別感謝實驗室的同學們在研究過程中的無私協助與經驗分享，您們的熱心支持使得本專題研究能夠克服諸多挑戰，最終如期圓滿完成。

七、參考文獻

[1] X. Liu, F. Zhang, Z. Hou, M. Li, Z. Wang, J. Zhang and J. Tang, "Self-Supervised Learning: Generative or Contrastive," arXiv preprint arXiv:2006.08218, Jun. 2020.

[2] K. He, H. Fan, Y. Wu, S. Xie and R. Girshick, "Momentum Contrast for Unsupervised Visual Representation Learning," arXiv preprint arXiv:1911.05722, Nov. 2019.

[3] X. Zhu and A. B. Goldberg, "A Survey on Deep Semi-Supervised Learning," in ACM Computing Surveys, vol. 54, no. 6, pp. 1 – 37, Jul. 2021.

[4] Z. Wang, Z. Li, J. Wang and D. Li, "Network intrusion detection model based on improved BYOL self-supervised learning",Secur. Commun. Netw., vol. 2021, pp. 1-23, Oct. 2021.

[5] C. Wu, J. Sun, J. Chen, M. Alazab, Y. Liu and Y. Xiang, "TCG-IDS: Robust Network Intrusion Detection via Temporal Contrastive Graph Learning," inIEEE Transactions on Information Forensics and Security, vol. 20, pp. 1475-1486, 2025

[6] Y. Yue, X. Chen, Z. Han, X. Zeng and Y. Zhu, "Contrastive Learning Enhanced Intrusion Detection," inIEEE Transactions on Network and Service Management, vol. 19, no. 4, pp. 4232-4247, Dec. 2022

[7] M. Lopez-Martin, A. Sanchez-Esguevillas, J. I. Arribas and B. Carro, "Contrastive Learning Over Random Fourier Features for IoT Network Intrusion Detection," in IEEE Internet of Things Journal, vol. 10, no. 10, pp. 8505-8513, 15 May15, 2023