影像分類攻擊方法模組圖形化介面 使用說明書

撰寫日期：2022/04/25

目錄

一、 執行說明

1. 直接執行 EXE 檔案

「ART_Tool_CPU.exe」為影像分類攻擊方法模組程式，可至 GitHub Release 裡下載最新版本。請注意，本程式只能運行在 **Windows 64 位元** 作業系統上，並且不支援 GPU 運行，只能運行在 CPU 上。

執行 EXE 檔方式支援後接參數設定，詳情請至「執行參數說明」查看。

2. 透過 Python 執行

本程式可以運行在 Python 相關虛擬環境中，像是 Anaconda。請注意，本程式只支援 **Python 3.8** 版本，在安裝相關環境前，請先注意 Python 版本是否正確。「requirements.txt」為安裝 Python 相關環境需求，可透過在終端機中執行「pip3 install -r requirements.txt」 來安裝。

安裝環境完成後， 執行「python ART_Tool_RGB.py」就可以執行圖形化介面。執行 Python 檔方式也支援後接參數設定，詳情請至「執行參數說明」查看。

3. 執行結果

實驗完成後會產生兩個實驗數據 CSV 檔，檔名為「目標模型_origin_Accuracy.csv」(代表未經攻擊準確率)、「目標模型_攻擊方法_Accuracy.csv」(代表歷次攻擊後準確率)，例如：LeNet5_origin_Accuracy.csv、LeNet5_FGSM_Accuracy.csv。

在「目標模型_攻擊方法_Accuracy.csv」(代表未經攻擊準確率)裡，資料內容欄位為攻擊模型、攻擊方法、準確率(%)、參數、程式執行結束日期時間。每一列為每一次執行攻擊實驗後的準確率。

二、 圖形化介面說明

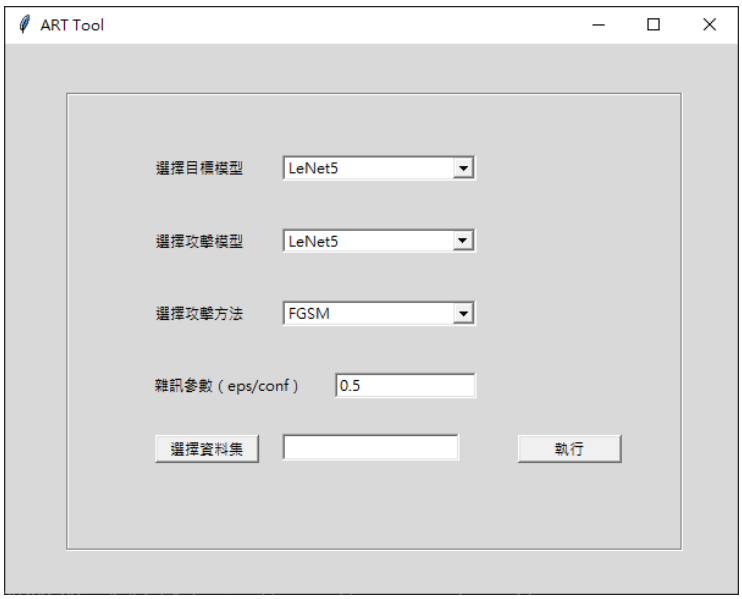

圖 1、圖形化介面

「選擇目標模型」、「選擇攻擊模型」、「選擇攻擊方法」為透過下拉選單來選擇，「雜訊參數」需自行輸入，否則為預設值「0.5」。如果為輸入 eps 擾動參數，建議輸入值介於 0 到 1 之間的浮點數，如果為輸入 conf 置信度，則建議輸入值介於 0 到 16 之間的整數。按下「選擇資料集」後，會進入選擇資料夾視窗，接受的資料夾格式請參見「資料集說明」。

三、 資料集說明

選擇一個資料夾路徑，該資料夾底下必須包含名稱為「train」及「test」的兩個子資料夾，各自代表訓練資料集及測試資料集。

其中資料集的架構，同一類別的圖片必須存在同一資料夾，且資料夾名稱為該類別的名稱，名稱形式不限。圖片檔名不影響程式運作。可接受的圖片檔案類型為 PNG 檔及 JPEG 檔。

1. MNIST 手寫辨識資料集 範例

以下為 MNIST 手寫辨識資料集架構，以作為可接受資料集架構範例：

```
mnist /      <=== 圖形介面選擇的資料集路徑
├── train/         <=== 訓練資料集
│    ├── 0/      <=== 類別為「0」的圖片資料夾
│    │    ├── train_0_1.png   <=== 檔名不影響
│    │    ├── train_0_2.png
│    │    └── ...
│    │
│    ├── 1/      <=== 類別為「1」的圖片資料夾
│    ├── 2/
│    ├── ...
│    └── 9/
│
└── test/          <=== 測試資料集
     ├── 0/
     ├── ...
     └── 9/
```

四、 執行參數說明

本程式支援參數設定，指令格式為

```
$ {ART_Tool_GPU.exe 或 python ART_Tool_RGB.py} [-h] [--interface INTERFACE]
    [--cuda CUDA] [--dataset-path DATASET_PATH] [--num-workers NUM_WORKERS]
    [--predict-model PREDICT_MODEL] [--attack-model ATTACK_MODEL]
    [--white-box] [--attack-func ATTACK_FUNCTION] [--max-iter MAX_ITER]
    [--eps EPS] [--conf CONFIDENCE] [--epochs EPOCHS] [--batch-size BATCH_SIZE]
    [--optim OPTIM] [--lr LR] [--momentum MOMENTUM]
```

範例

```
$ ART_Tool_GPU.exe --cuda 1 --epochs 5 --lr 0.01
$ python ART_Tool_RGB.py --num-workers 4 –norm
```

表 1、個別參數說明

參數	後接參數型態	預設值	說明
-h, --help	無	無	顯示參數說明
--interface	整數	1	選擇設定參數方式 (1:GUI 介面, 2:標準輸入, 3:執行參數)
--cuda	整數	0	設定運行 GPU 的 id (-1:CPU, ≥0:GPU CUDA id)
--dataset-path	資料夾路徑	無	設定資料集路徑
--num-workers	整數	8	設定執行緒數量
--predict-model	整數	1	選擇目標模型 (1:LeNet5, 2:CNN, 3:AlexNet, 4:GoogLeNet, 5:VGG19, 6:ResNeXt101)
--attack-model	整數	1	選擇攻擊模型 (1:LeNet5, 2:CNN, 3:AlexNet, 4:GoogLeNet, 5:VGG19, 6:ResNeXt101)
--white-box	無	否	設定白盒實驗
--attack-func	整數	1	選擇攻擊方法 (1:FGSM, 2:BIM, 3:PGD, 4:C&W L2, 5:C&W Linf)
--max-iter	整數	20	設定最大迭代次數
--eps	浮點數	0.1	設定擾動參數
--conf	整數	無	設定置信度
--epochs	整數	20	設定訓練回合數
--batch-size	整數	32	設定 batch 值
--optim	優化器名稱(字串)	SGD	設定優化器
--lr	浮點數	0.001	設定學習速率
--momentum	浮點數	0.9	設定 SGD 的 Momentum 值