|  |
| --- |
|  |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА –Российский технологический университет»**  **РТУ МИРЭА**  **Институт кибербезопасности и цифровых технологий**  **Кафедра КБ-4 «Интеллектуальные системы информационной безопасности»** |
|  |

Дисциплина «Анализ защищенности систем искусственного интеллекта»

Отчет

о проделанной лабораторной работе №1

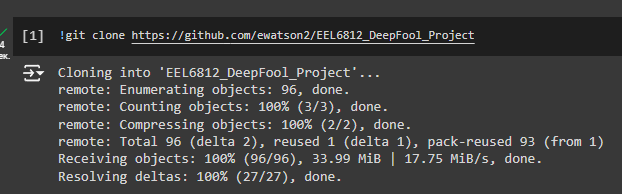
«EEL6812 DeepFool Project»

Выполнил студент 2 курса

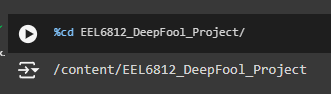
Группы: ББМО-01-23

Чурсинов Герман Сергеевич

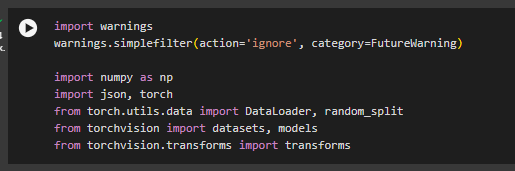
1. Скопировать проект по ссылке в локальную среду выполнения Jupyter (Google Colab)



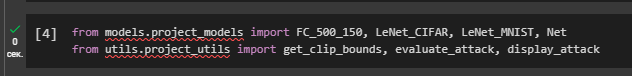
2. Сменить директорию исполнения на вновь созданную папку "EEL6812\_DeepFool\_Project" проекта.



3. Выполнить импорт библиотек



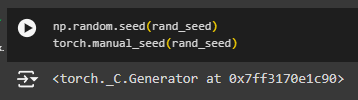
4. Выполнить импорт вспомогательных библиотек из локальных файлов проекта:



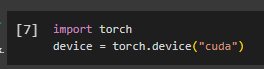
5. Установить случайное рандомное значение в виде переменной rand\_seed



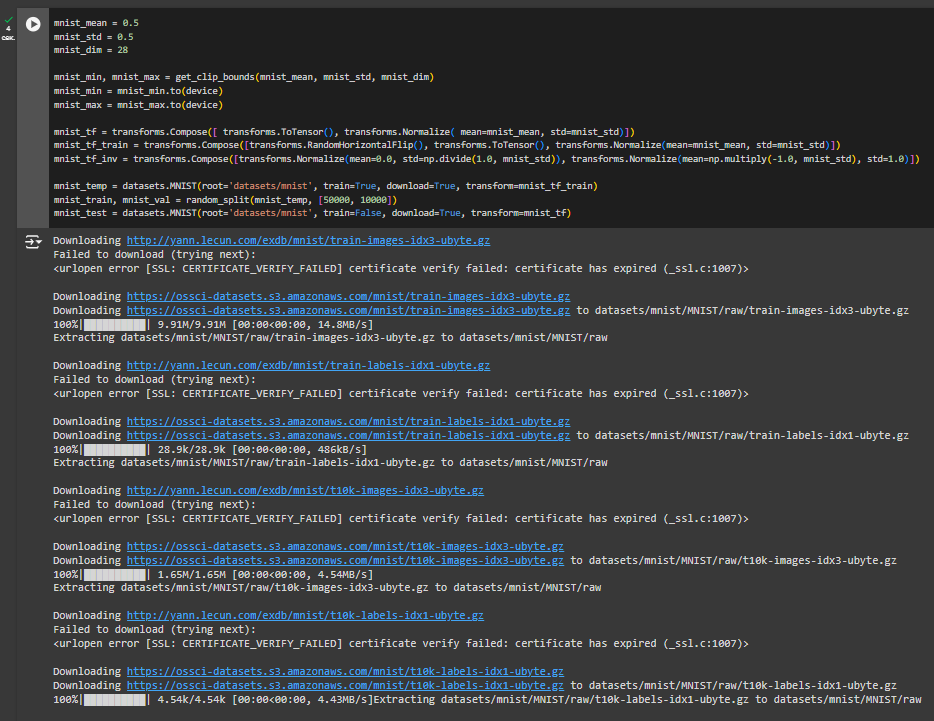
6. Установить указанное значение для np.random.seed и torch.manual\_seed



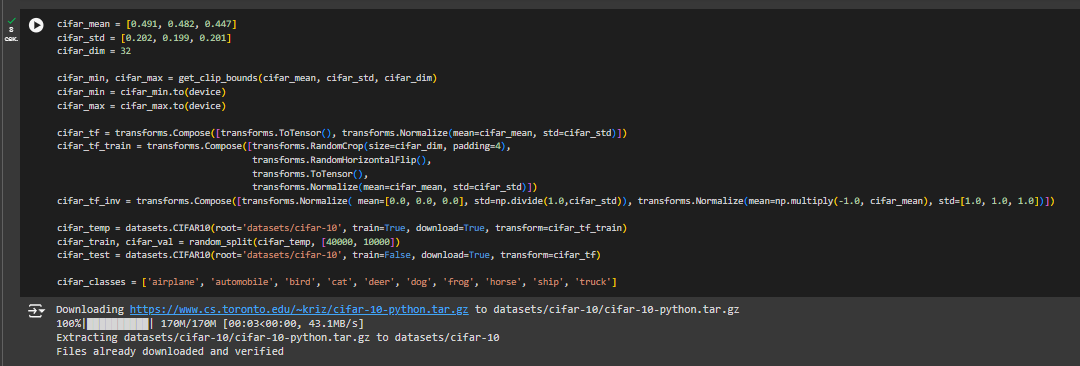
7. Использовать в качестсве устройства видеокарту



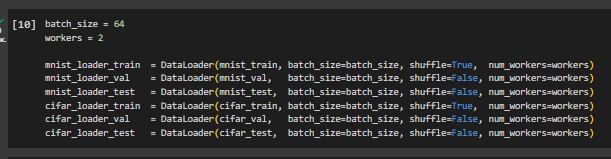
8. Загрузить датасет MNIST c параметрами mnist\_mean = 0.5, mnist\_std = 0.5, mnist\_dim = 28



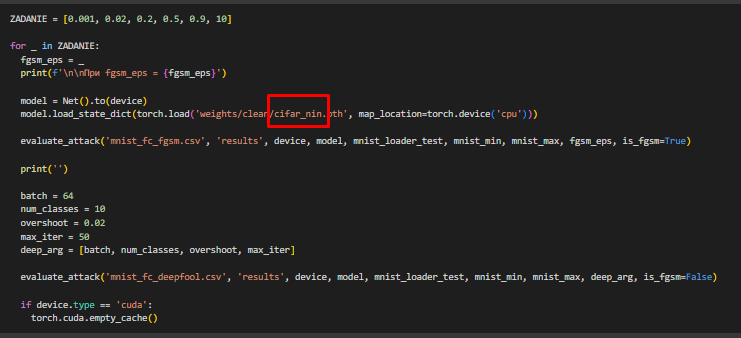
9. Загрузить датасет CIFAR-10 c параметрами cifar\_mean = [0.491, 0.482, 0.447] cifar\_std = [0.202, 0.199, 0.201] cifar\_dim = 32



10. Выполнить настройку и загрузку DataLoader batch\_size = 64 workers = 2



11. Загрузить и оценить стойкость модели Network-In-Network Model к FGSM и DeepFool атакам на основе датасета CIFAR-10



При fgsm\_eps = 0.001

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.02

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.2

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.5

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.9

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 10

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

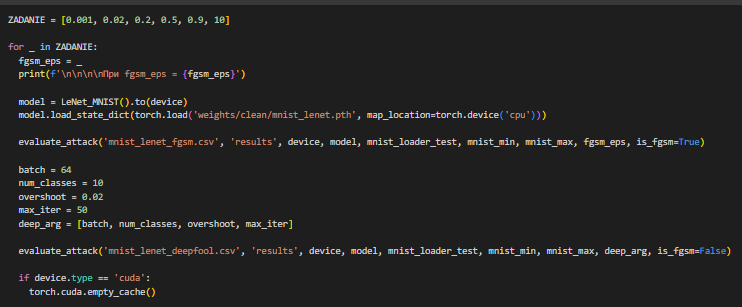
DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

12. Загрузить и оценить стойкость модели LeNet к FGSM и DeepFool атакам на основе датасета CIFAR-10



13. Выполнить оценку атакующих примеров для сетей: LeNet



При fgsm\_eps = 0.001

FGSM Test Error : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.29 s

FGSM Time (Per Image) : 28.86 us

DeepFool Test Error : 98.74%

DeepFool Robustness : 9.64e-02

DeepFool Time (All Images) : 193.32 s

DeepFool Time (Per Image) : 19.33 ms

При fgsm\_eps = 0.02

FGSM Test Error : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.29 s

FGSM Time (Per Image) : 28.86 us

DeepFool Test Error : 98.74%

DeepFool Robustness : 9.64e-02

DeepFool Time (All Images) : 193.32 s

DeepFool Time (Per Image) : 19.33 ms

При fgsm\_eps = 0.2

FGSM Test Error : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.29 s

FGSM Time (Per Image) : 28.86 us

DeepFool Test Error : 98.74%

DeepFool Robustness : 9.64e-02

DeepFool Time (All Images) : 193.32 s

DeepFool Time (Per Image) : 19.33 ms

При fgsm\_eps = 0.5

FGSM Test Error : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.29 s

FGSM Time (Per Image) : 28.86 us

DeepFool Test Error : 98.74%

DeepFool Robustness : 9.64e-02

DeepFool Time (All Images) : 193.32 s

DeepFool Time (Per Image) : 19.33 ms

При fgsm\_eps = 0.9

FGSM Test Error : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.29 s

FGSM Time (Per Image) : 28.86 us

DeepFool Test Error : 98.74%

DeepFool Robustness : 9.64e-02

DeepFool Time (All Images) : 193.32 s

DeepFool Time (Per Image) : 19.33 ms

При fgsm\_eps = 10

FGSM Test Error : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.29 s

FGSM Time (Per Image) : 28.86 us

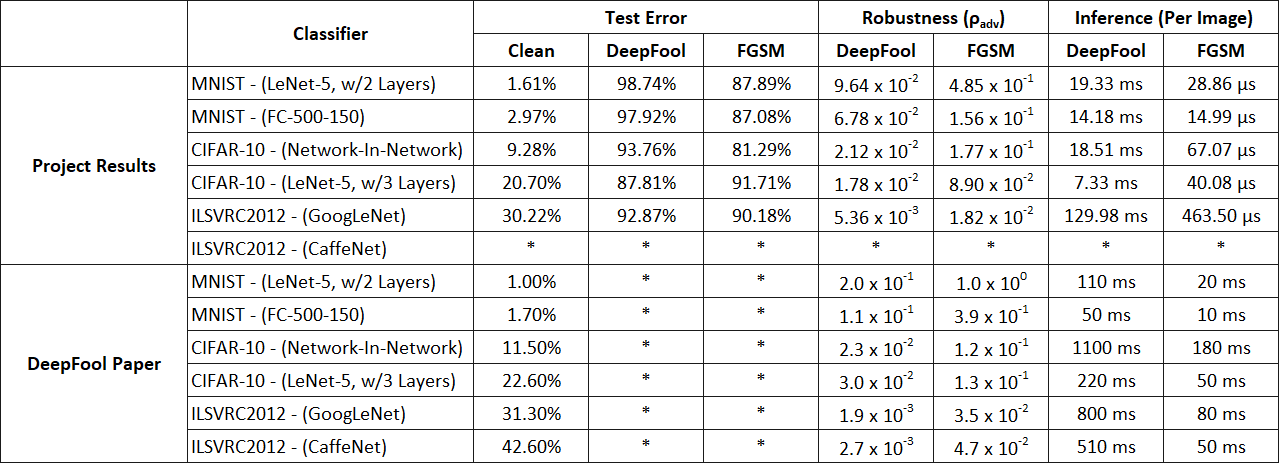
DeepFool Test Error : 98.74%

DeepFool Robustness : 9.64e-02

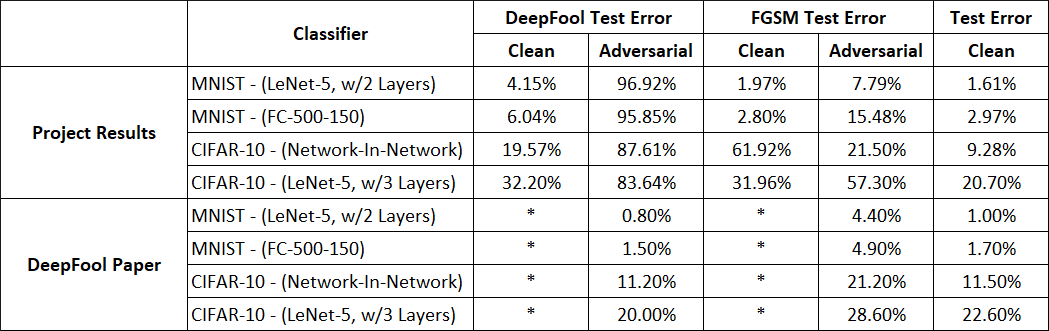
DeepFool Time (All Images) : 193.32 s

DeepFool Time (Per Image) : 19.33 ms

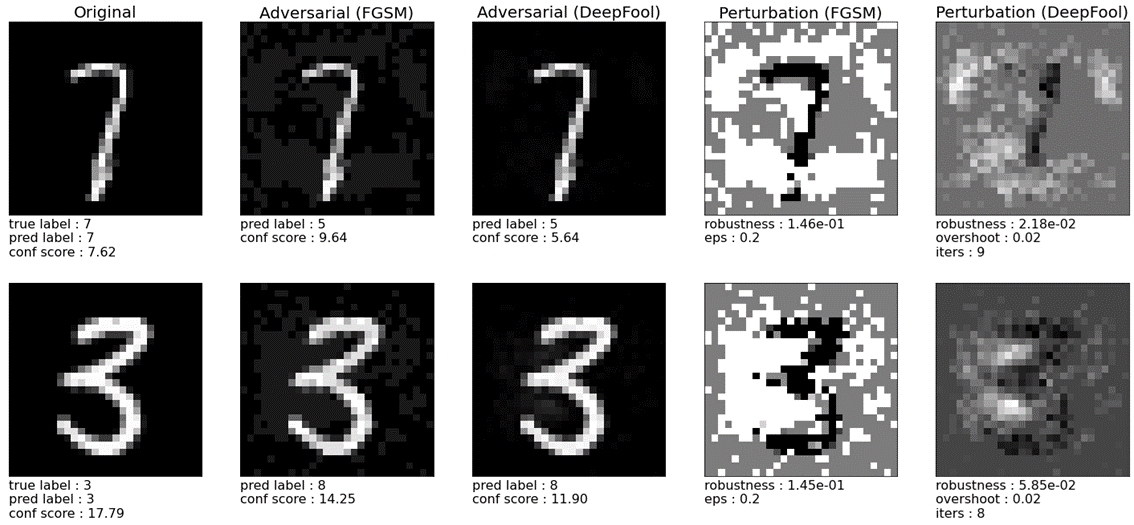
14. Подготовить отчет в формате pdf (отразить отличия для fgsm\_eps=(0.001, 0.02, 0.5, 0.9, 10) и выявить закономерность/обнаружить отсутсвие влияние параметра eps для сетей FC LeNet на датасете MNIST, NiN LeNEt на датасете CIFAR



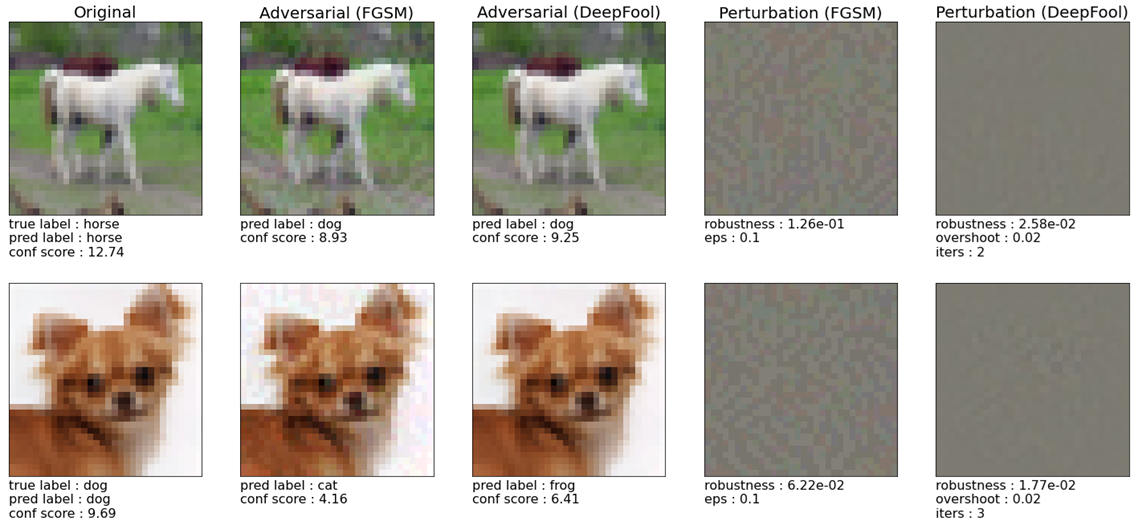
Таблица, демонстрирующая экспериментальные результаты, состязательный вывод и надежность



Обучение состязательности



Примеры состязательных программ MNIST (FC-500-150)



Примеры состязательных программ CIFAR-10 (LeNet-5)

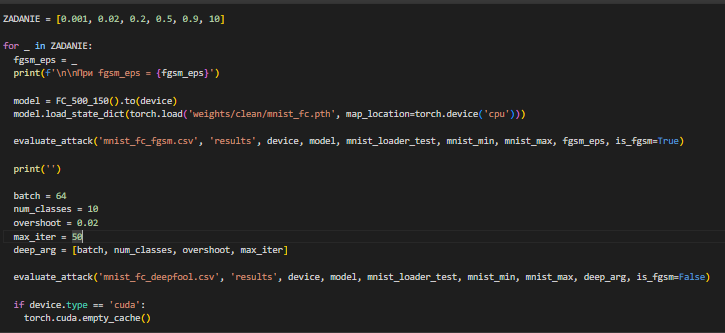


Примеры состязательных атак ILSVRC2012 (GoogLeNet)

Для этого мы указывали заранее классы, по которым будет проходить распределение моделью изображений.



**eps** – эпсилон, параметр, отвечающий за максимальную степень изменчивости отдельно взятого пикселя



При fgsm\_eps = 0.001

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.02

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.2

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.5

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.9

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 10

FGSM Test Error : 87.08%

FGSM Robustness : 1.56e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms