|  |
| --- |
| 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

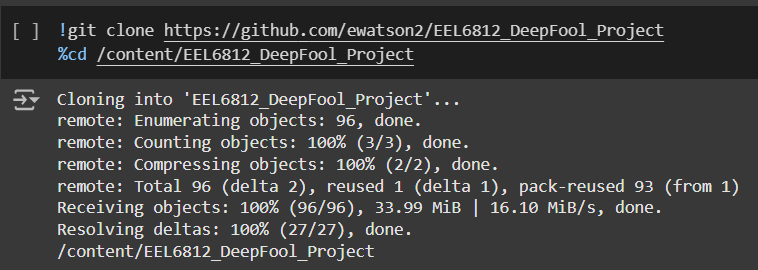
«EEL6812 DeepFool Project»

Выполнил студент 2 курса

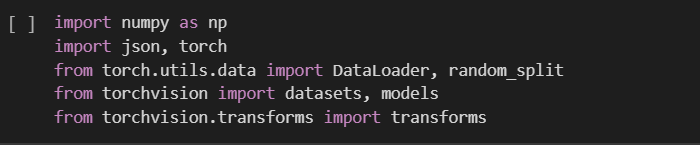
Группы: ББМО-01-23

Асатрян Давид Робертович

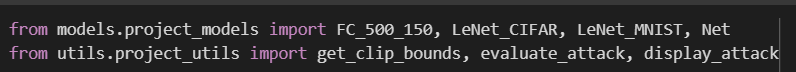
1. Скопировать проект по ссылке в локальную среду выполнения Jupyter



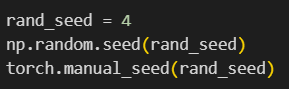
2. Импорт библиотек



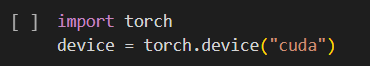
3. Выполнить импорт вспомогательных библиотек из локальных файлов проекта



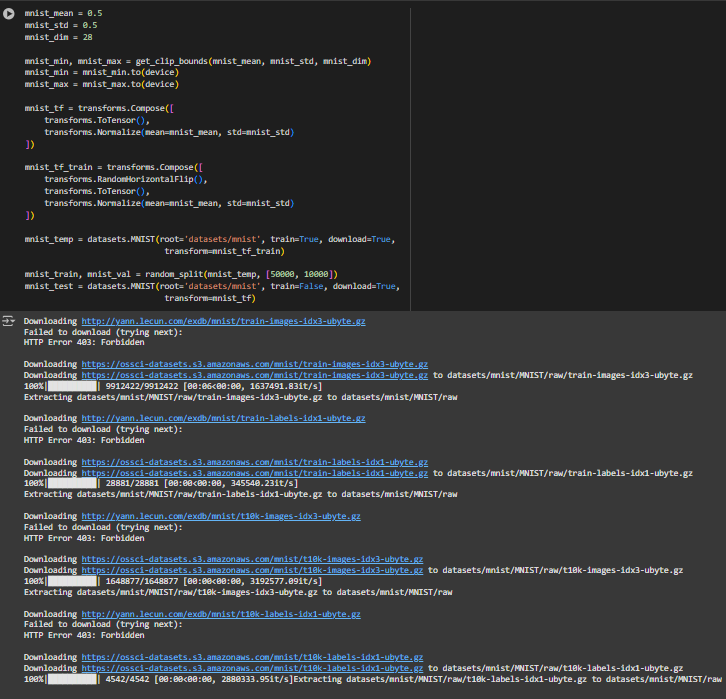
4. Установить случайное рандомное значение в виде переменной rand\_seed и, установить указанное значение для np.random.seed и torch.manual\_seed



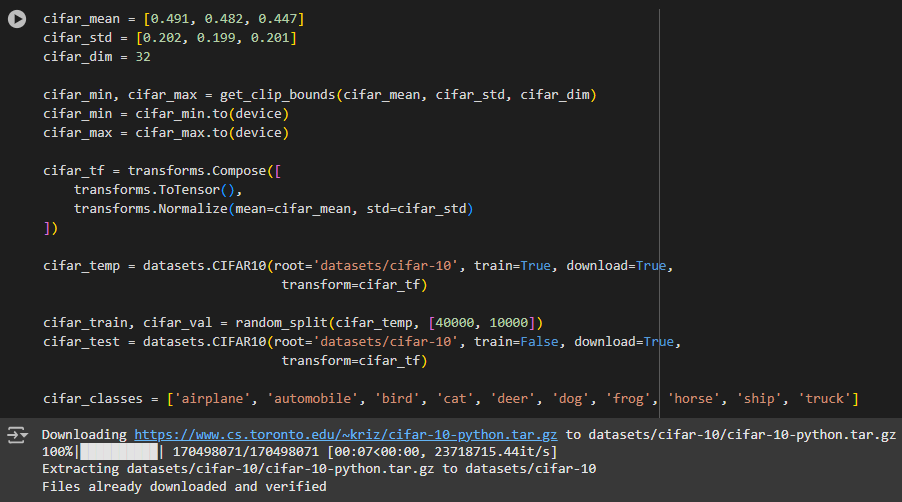
5. Использовать в качестсве устройства видеокарту



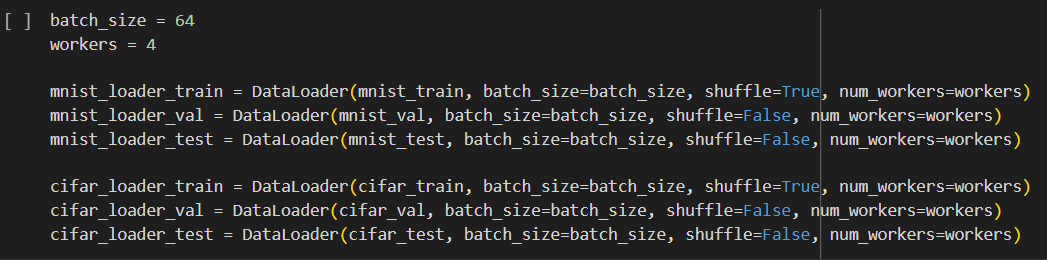
6. Загрузить датасет MNIST c параметрами mnist\_mean = 0.5, mnist\_std = 0.5, mnist\_dim = 28



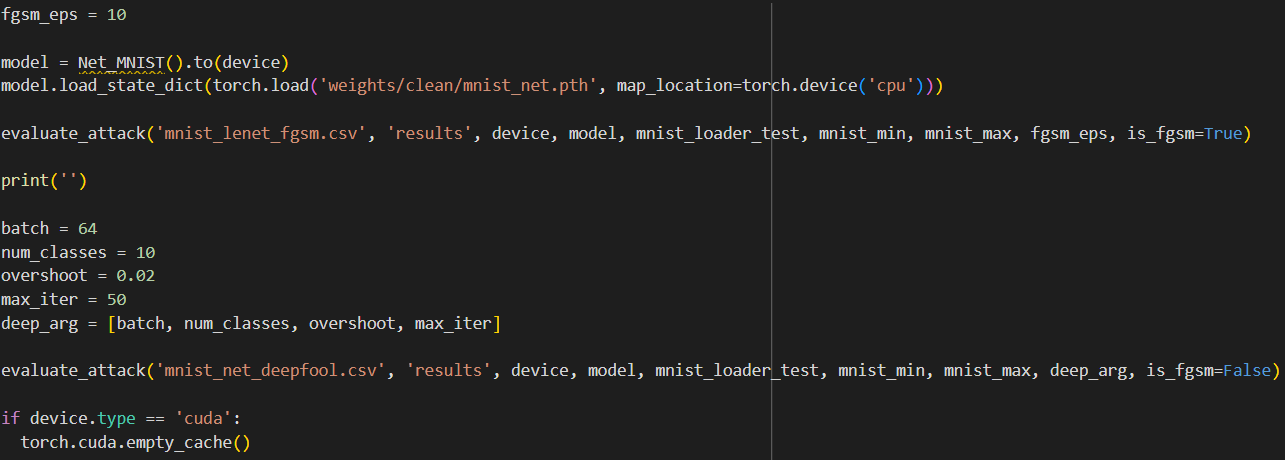
7. Загрузить датасет CIFAR-10 c параметрами cifar\_mean = [0.491, 0.482, 0.447] cifar\_std = [0.202, 0.199, 0.201] cifar\_dim = 32



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



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



При fgsm\_eps = 0.01

FGSM Test Error : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 98.74%

DeepFool Robustness : 9.64e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

При fgsm\_eps = 0.02

FGSM Test Error : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 98.74%

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 : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 98.74%

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 : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 98.74%

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 : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 98.74%

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 : 87.89%

FGSM Robustness : 4.58e-01

FGSM Time (All Images) : 0.15 s

FGSM Time (Per Image) : 14.99 us

DeepFool Test Error : 98.74%

DeepFool Robustness : 6.78e-02

DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

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



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

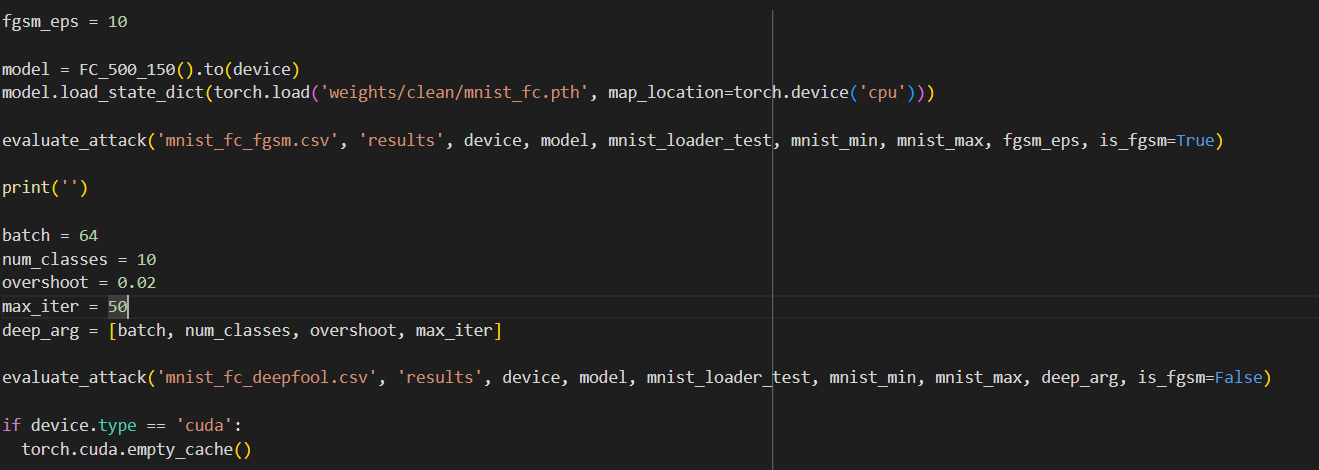
FGSM Time (Per Image) : 28.86 us

DeepFool Test Error : 98.74%

DeepFool Robustness : 9.64e-02

DeepFool Time (All Images) : 193.32 s

DeepFool Time (Per Image) : 19.33 ms

11. Загрузить и оценить стойкость модели FC к FGSM и DeepFoo 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

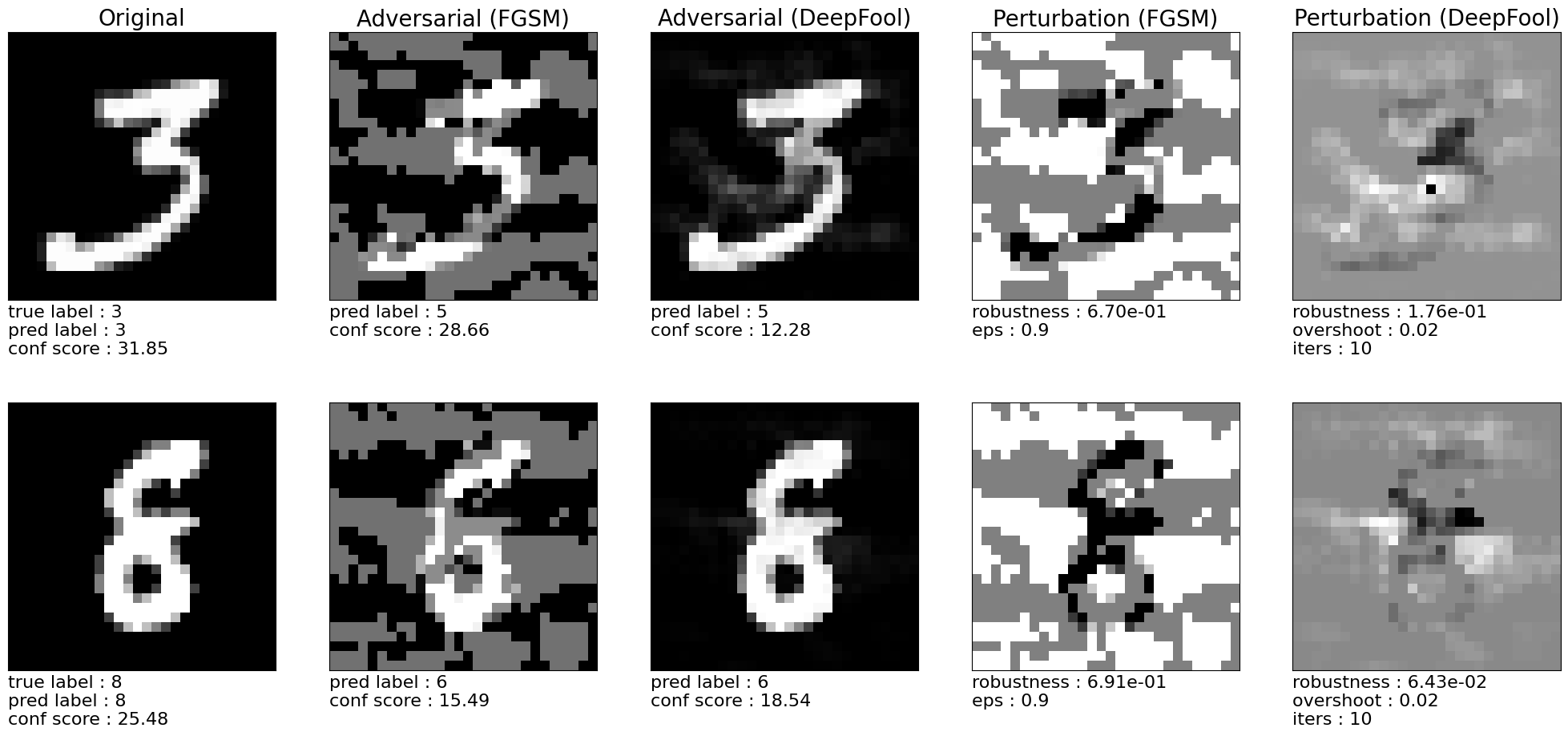
DeepFool Test Error : 97.92%

DeepFool Robustness : 6.78e-02

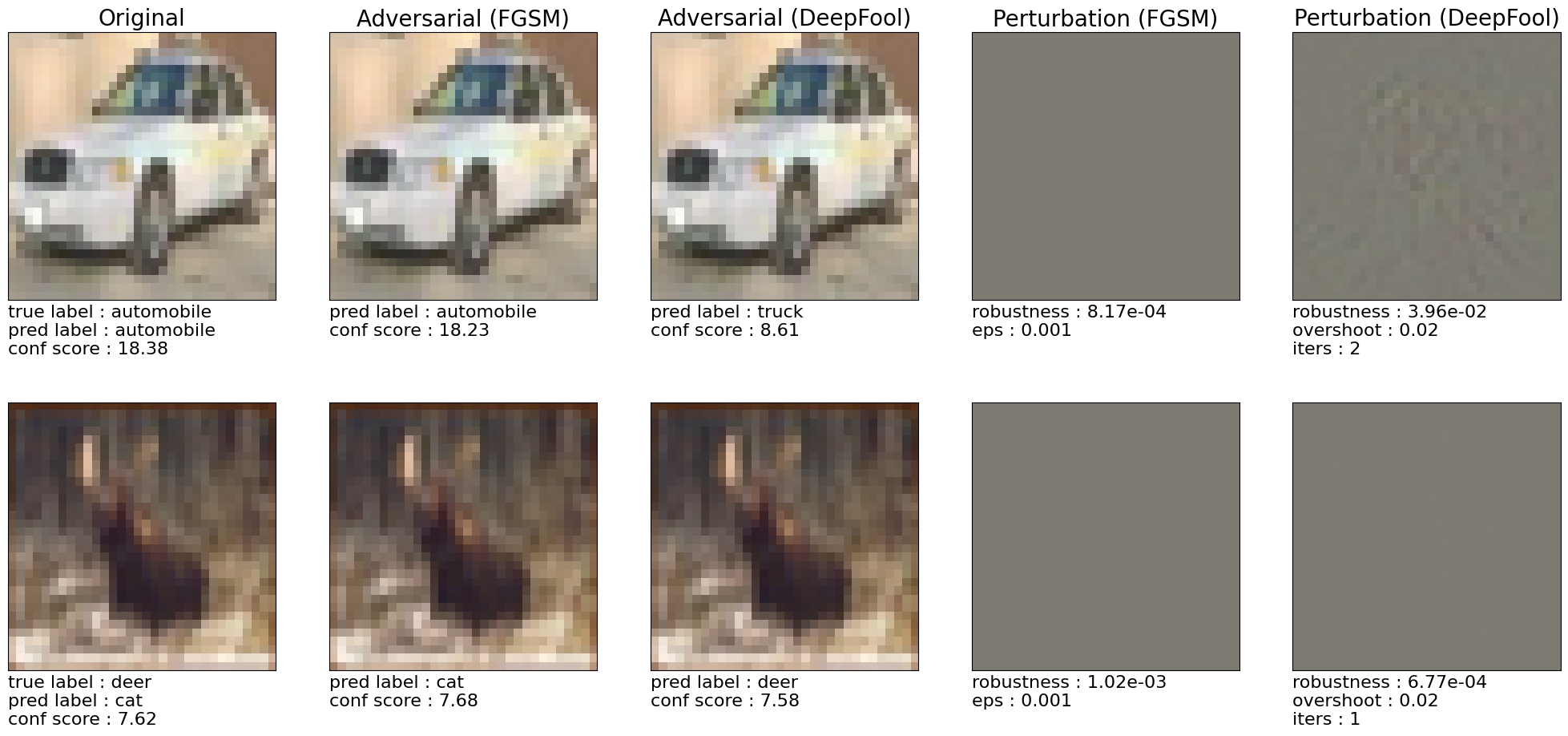
DeepFool Time (All Images) : 141.81 s

DeepFool Time (Per Image) : 14.18 ms

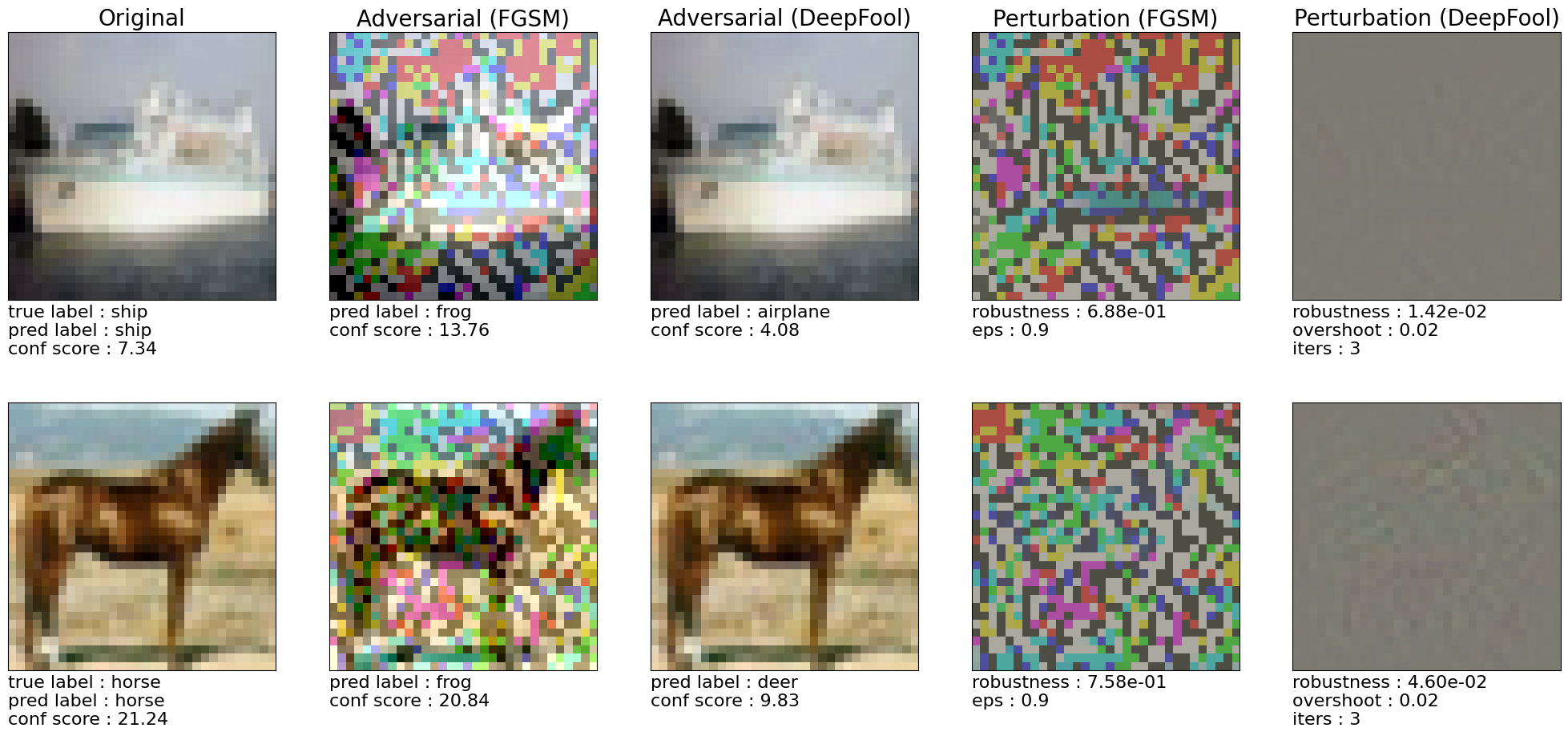
Обучение состязательности



Примеры состязательных программ MNIST (FC-500-150)



Примеры состязательных программ CIFAR-10 (LeNet-5)



Примеры состязательных атак ILSVRC2012 (GoogLeNet)