Изображение выглядит как текст, коллекция картинок

Автоматически созданное описание

МИНОБРНАУКИРОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«МИРЭА**— **Российский технологический университет»**

**РТУМИРЭА**

Институт кибербезопасности и цифровых технологий направление 10.04.01 «Информационная безопасность»

Кафедра КБ-4«Интеллектуальные системы информационной безопасности»

**Лабораторная работа №2**

По дисциплине

«Анализ защищенности систем искусственного интеллекта»

Выполнил:

Суслов Антон Константинович

Группа: ББМО-02-22

Москва 2023

**Задание 1**

Установка adversarial-robustness-toolbox

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Импорт библиотек

Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

Распаковка архива

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Чтение и предварительная обработка изображений

Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание

Первое изображение

Изображение выглядит как снимок экрана, Мультимедийное программное обеспечение, Графическое программное обеспечение, программное обеспечение

Автоматически созданное описание

Создание модели для классификации изображений

Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Автоматически созданное описание

Обучаем модель ResNet50 с заданными параметрами

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Создание второй модели для классификации изображений

Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание

Обучение модели VGG16 с заданные параметрами

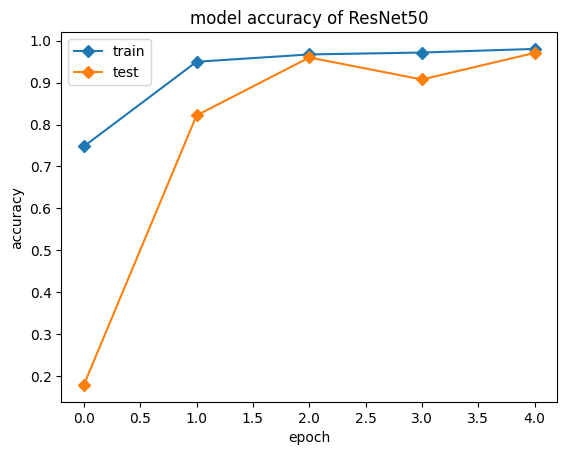
Изображение выглядит как текст, Шрифт, программное обеспечение, снимок экрана

Автоматически созданное описание

Оценка двух моделей

Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание



Изображение выглядит как линия, диаграмма, График, текст

Автоматически созданное описание

Изображение выглядит как текст, линия, снимок экрана, График

Автоматически созданное описание

Изображение выглядит как линия, диаграмма, снимок экрана, График

Автоматически созданное описание

Вывод: как видно из таблицы выше, модели показывают схожие результаты.

**Задание 2**

Загружаем модель из предыдущего задания

Изображение выглядит как текст, Шрифт, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

Создание атаки FGSM

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание

Отображение исходных и адверсариальных изображений

Изображение выглядит как снимок экрана, текст, прямоугольный, пиксель

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, шаблон

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, пиксель

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, текст

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, пиксель

Автоматически созданное описание

Создание атаки PGD

Изображение выглядит как текст, снимок экрана, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

Отображение исходных и адверсариальных изображений

Изображение выглядит как снимок экрана, текст, прямоугольный, пиксель

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, шаблон, пиксель

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, текст, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, текст, Красочность

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, фиолетовый

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, фиолетовый

Автоматически созданное описание

Создание атаки PGD

Изображение выглядит как текст, снимок экрана, программное обеспечение

Автоматически созданное описание

Отображение исходных и адверсариальных изображений

Изображение выглядит как снимок экрана, текст, прямоугольный, пиксель

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, пиксель, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, прямоугольный, шаблон

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, пиксель, шаблон

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, прямоугольный, пиксель

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, фиолетовый

Автоматически созданное описание

Когда пришло время строить графики и смотреть зависимости, обратил внимание на то, что из-за нехватки места в гугл колабе, нужные файлы, видимо, стерлись. Поэтому продемонстрирую то, что есть.

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, линия, диаграмма

Автоматически созданное описание Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание

**Задание 3**

Загрузка тестового набора данных и извлечения изображения с меткой 14

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Реализация атаки FGSM

Изображение выглядит как текст, снимок экрана, Шрифт, программное обеспечение

Автоматически созданное описание

Визуализация последствий атаки

Изображение выглядит как текст, снимок экрана, Шрифт, диаграмма

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, Красочность, пиксель

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, Красочность

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, текст

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, прямоугольный, пиксель

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, Красочность, шаблон

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, прямоугольный, Шрифт

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, Красочность

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, Красочность, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, Красочность, текст, шаблон

Автоматически созданное описание

Реализация атаки PGD

Изображение выглядит как текст, снимок экрана, программное обеспечение, Шрифт

Автоматически созданное описание

Визуализация последствий атаки

Изображение выглядит как текст, снимок экрана, Шрифт, диаграмма

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, Красочность

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, Красочность

Автоматически созданное описание

Изображение выглядит как снимок экрана, шаблон, Красочность, прямоугольный

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, прямоугольный, пиксель

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, Красочность

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, прямоугольный, Шрифт

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, шаблон, Красочность

Автоматически созданное описание

Изображение выглядит как снимок экрана, текст, Красочность, прямоугольный

Автоматически созданное описание

Изображение выглядит как снимок экрана, Красочность, текст, шаблон

Автоматически созданное описание

Таблица со значениями точности при двух атаках

|  |  |  |
| --- | --- | --- |
| **Искажение** | **FGSM** | **PGD** |
| 𝜖=1/255 | 85% | 97% |
| 𝜖=2/255 | 76% | 93% |
| 𝜖=3/255 | 62% | 87% |
| 𝜖=4/255 | 52% | 86% |
| 𝜖=5/255 | 45% | 79% |
| 𝜖=8/255 | 19% | 76% |
| 𝜖=10/255 | 13% | 75% |
| 𝜖=20/255 | 4% | 38% |
| 𝜖=50/255 | 1% | 5% |
| 𝜖=80/255 | 1% | 3% |