БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра биомедицинской информатики

Аннотация к дипломной работе

«Разработка экспериментальных методов защиты от adversarial атак»

Гимбицкий Виталий Владимирович

Научный руководитель – кандидат технических наук, доцент кафедры биомедицинской информатики ФПМИ Ковалёв В. А.

Минск, 2024

Реферат

Дипломная работа, 51 страница, 25 рисунков, 13 таблиц, 12 формул, 6 источников.

Ключевые слова: НЕЙРОННАЯ СЕТЬ, ЗАЩИТА, ВРАЖДЕБНАЯ АТАКА, FAST SIGN GRADIENT METHOD, ADVERSARIAL TRAINING.

Объектом исследования является защита нейронных сетей от adversarial атак.

Предметом исследования являются модели нейронных сетей а также алгоритмы для защиты от adversarial атак.

Целью работы является разработка и оптимизация методов защиты от adversarial атак.

В ходе работы были изучены и реализованы существующие методы защиты от adversarial атак. Предложены новые архитектуры нейронных сетей для защиты от adversarial атак. Проведены эксперименты для оценки качества защиты для различных методов и наборов данных. По данным исследованиям проведены наблюдения о зависимости качества защиты от метода, продолжительности обучения и набора данных.

Полученные в результате работы модели можно применять для защиты классификационных нейронных сетей в задачах компьютерного зрения.

Abstract

Diploma thesis,51 pages, 25 figures, 13 tables, 12 formulas, 6 sources.

Keywords: NEURAL NETWORK, PROTECTION, ADVERSARIAL ATTACK, FAST SIGN GRADIENT METHOD, ADVERSARIAL TRAINING.

The object of research is the defense of neural networks against adversarial attacks.

The subject of study is is neural network models and algorithms for defense against adversarial attacks.

The aim of this work is to develop and optimize methods of protection against adversarial attacks.

In the course of the work, the existing methods of defense against adversarial attacks were studied and implemented. New architectures of neural networks for protection against adversarial attacks have been proposed. Experiments have been conducted to evaluate the quality of defense for different methods and datasets. Observations on the dependence of defense quality on method, training duration and dataset are made on these studies.

*The resulting model*s can be applied to the defense of classification neural networks in computer vision tasks.