обучающие тексты, тем больше их должно быть в обучающем множестве, соответственно, чем фрагменты больше – тем меньшим их количеством можно обойтись.

Литература

1. Леонтьева, Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы [Текст]: учеб. пособие для вузов/ Н. Н. Леонтьева – М.: Издательский центр «Академия», 2006. – 304 с.
2. Machine Learning in Automated Text Categorization [Электронный ресурс] Автоматическая классификация текстов — http://www.math.unipd.it/~fabseb60/Publications/ACMCS02.pdf (дата обращения 14.06.2022).
3. Sebastiani, F.: Machine learning in automated text categorization, ACM Computing Surveys Computing Surveys, vol. 34, pp. 1–47, 2002.
4. Деревья решений – CART математический аппарат [Электронный ресурс]. – https://basegroup.ru/community/articles/math-cart-part1 (дата обращения 15.03.2022).

А.А. Столбова, А.А. Малышев

РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ВЫЯВЛЕНИЯ ФАЛЬСИФИЦИРОВАННЫХ ФРАГМЕНТОВ НА ФОТОИЗОБРАЖЕНИЯХ

(Самарский национальный исследовательский университет
имени академика С.П. Королёва)

В данной работе предлагается разработанная автоматизированная система выявления фальсифицированных фрагментов на фотоизображениях. Система обеспечивает многоэтапную обработку изображений с целью нахождения на них фрагментов, выполненных после создания изображений, в том числе, средствами современных нейронных сетей [1]. Данное решение развивает тему, затронутую в [2], в части развития интеллектуальных возможностей анализа изображений и локализации конкретного фрагмента, подвергшегося редактированию.

Поскольку метаданные цифровых фотографий (EXIF) содержат в себе достаточно исчерпывающую информацию о времени, месте, и устройстве, на которое они были сделаны [3], а сами фотографии сохраняются форматах со сжатием с потерями, разрабатываемая автоматизированная система выявления фальсифицированных фрагментов предполагает к реализации три этапа анализа изображения:

1. Анализ метаданных изображения;
2. Анализ изображения с помощью ELA;
3. Нейросетевой анализ изображения.

На первом этапе система анализирует следующие EXIF-метаданные:

время и дата съемки, место съемки (координаты), модель устройства, диафрагма, выдержка, ISO. Анализ фотоизображения с использованием метаданных проводится на первом этапе, поскольку существуют способы, позволяющие редактировать информацию EXIF и вносить в нее недостоверные данные [4]. Недостоверные сведения в EXIF фиксируются системой.

На втором этапе анализ изображения в системе осуществляется на основе алгоритма оценки уровня ошибок ELA, который задействует схемы сжатия с потерями для выявления манипуляций. Используемые в системе изображения, сжатые по схеме с потерями, повторно сжимаются с известной частотой ошибок, после чего вычисляется разница между исходными изображениями и повторно сжатыми. Исходные изображения имеют достаточно высокие значения уровня ошибок, но каждое последующее сжатие его уменьшает. Таким образом, если изображение подвергается изменению, то в этих измененных областях более высокий уровень ошибок по сравнению с исходным изображением [5, 6]. Найденные подозрительные области (фрагменты) изображения фиксируются системой.

На третьем этапе осуществляется нейросетевой анализ изображения, применяемый, в первую очередь для того, чтобы выявить фрагменты, изменения в которые были внесены с применением нейросетей и редакторов на их основе. На этом этапе исходное анализируемое изображение разбивается на отдельные фрагменты, каждый из которых в цикле обрабатывается нейросетью, классифицирующей фрагмент как имеющий или не имеющий правок. Поскольку нейросети, вносящие изменения в изображения, действуют схожим образом для внесения тех или иных правок, сформирован набор данных для обучения на распространенных типовых изменениях.

Система построена как веб-сервис, что позволяет встраивать ее в другие системы для оценки изображений на наличие фальсифицированных фрагментов. При этом реализован и графический интерфейс пользователя (веб-приложение), позволяющий загрузить фотоизображение и визуально ознакомиться с результатами работы системы: выявленными недостоверными сведениями в EXIF, найденными с помощью ELA фрагментами с отличиями, найденными фрагментами, классифицированными нейросетью как имеющие правки.

Литература

1. Гераськин, А.С. Анализ методов проверки фотоизображений на наличие внесённых изменений / А.С. Гераськин, С.Ю. Желтов // Информационная безопасность регионов. – 2016. – № 4(25). – С. 5-10.

2. Ганеев, Р.М. Разработка программного модуля выявления фейковых фотографий в социальных сетях / Р.М. Ганеев, А.А. Столбова // XVI Королёвские чтения: сборник материалов конф. – Самара: Самарский ун-т, 2021. – С. 458-459.

3. Расширение фотографий: форматы записи цифровых фотографий. – URL: www.fotoprizer.ru/articles/teoria-fotografii/rasshirenie-fotografiy-formati-zapisi-cifrovih-fotografiy/153/?n=153&q=1335 (дата обращения 09.04.2022).

4. Как разоблачить фотоманипуляции // Рамблер Групп Ferra.ru. – URL: www.ferra.ru/review/multimedia/79974.htm (дата обращения 09.04.2022).

5. Jeronymo, D.C. Image forgery detection by semi-automatic wavelet soft-Thresholding with error level analysis / D.C. Jeronymo, Y.C.C. Borges, L.S. Coelho // Expert Systems with Applications. – 2017. – С. 348-356.

6. Sudiatmika, I.B.K. Image forgery detection using error level analysis and deep learning [Текст] / I.B.K. Sudiatmika, F. Rahman // Telecommunication Computing Electronics and Control. – 2019. – С. 653-659.

С.В. Толмачев, И.П. Болодурина, Д.И. Парфёнов, Л.С. Гришина, А.Ю. Жигалов

ИССЛЕДОВАНИЕ УСТОЙЧИВОСТИ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ ПРИ ПРОВЕДЕНИИ СОСТЯЗАТЕЛЬНЫХ АТАК

(Оренбургский государственный университет)

Машинное обучение всё глубже проникает во все сферы нашей жизни, включая виртуальные ассистенты с голосовым управлением, компьютерное зрение и т.п. Глубокие нейронные сети (ГНС) являются один из наиболее распространенных инструментов решения подобных задач, т.к. способны устанавливать закономерности в неструктурированных данных, таких как изображения, видео- и аудиоинформация. Несмотря на то, что современные модели глубокого обучения достаточно надежны и вероятность их ошибки с каждым годом стремится к нулю, они по-прежнему подвержены так называемым состязательным атакам. Состязательный пример это вектор входных данных, для которых модель стабильно выдает предсказания, неверные с точки зрения человека. В связи с тем, что на системы искусственного интеллекта зачастую возлагается функция принятия решения, необходимо обеспечить их устойчивость к уязвимостям данного рода.

С момента первого упоминания на Международной конференции по репрезентационному обучению в 2014 году о наличии состязательной угрозы [1] для алгоритмов глубокого обучения было разработано большое количество методов генерации вредоносных входных данных и способов защиты от них [2]. В рамках настоящей работы попытаемся изучить некоторые типы состязательных атак с различными моделями угроз для построения в дальнейшем устойчивой модели глубокой нейронной сети для решения задач компьютерного зрения.
Рассмотрим следующую постановку задачи классификации изображений.

Пусть дана база данных изображений дорожных знаков для проведения многоклассовой классификации. База данных содержит тренировочный набор из 39209 размеченных изображений и тестовый набор объемом в 12630 преце-