Инженерный вестник Дона, №6 (2023)
ivdon.ru/ru/magazine/archive/n6y2023/8460

Применение универсальных состязательных атак в задачах повышения эффективности систем защиты от роботов и спама

В.С. Казанцев, А.О. Мельников, А.М. Русаков, В.В. Филатов, С.С. Долженков

МИРЭА – Российский технологический университет, Москва

Аннотация: В данной статье рассматривается использование универсальных состязательных атак для улучшения эффективности систем защиты от роботов и спама. В частности, рассматриваются ключевые особенности, которые необходимо учитывать для обеспечения оптимального уровня защиты от роботов и спама. Также обсуждается, почему современные методы защиты неэффективны, и как использование универсальных состязательных атак может помочь устранить существующие недостатки. Цель данной статьи - предложить новые подходы и методы защиты, которые могут улучшить эффективность и устойчивость систем защиты от роботов и спама.
Ключевые слова: устойчивость систем защиты, роботы, спам, эффективность защиты, безопасность, нейронные сети, атаки на нейронные сети.

Введение

В современном мире с постоянным развитием информационных технологий и интернет-технологий все более острым становится вопрос защиты от роботов и спама. Ведущая компания в сфере защиты веб-приложений от роботов и спама, Kasada (Australia, Sydney), в 2022 году отчиталась о растущей актуальности данной проблемы (рис. 1).

Рис. 1. – Ожидания компаний по увеличению затрат на защиту от роботов и спама, согласно отчету компании Kasada [1]

Одним из наиболее распространенных методов защиты от роботов и спама является использование автоматизированных тестов Тьюринга, также известных, как «Completely Automated Public Turing test to tell Computers and Humans Apart» (далее CAPTCHA) [2, 3].

Основным принципом работы CAPTCHA является эксплуатация задач, простых для выполнения человеком, но сложных или невозможных для выполнения автоматического. Распространённой задачей, предлагаемой тестами CAPTCHA, является задача распознавания и классификации изображений, не требующая от пользователя специфических знаний, навыков, или принадлежности к определенной языковой группе, при этом достаточно сложная для решения автоматизированными методами [4, 5].

За последние годы нейронные сети достигли ошеломляющего успеха в распознавании и классификации изображений, часто превосходя способности человека в решении подобных задач [6, 7], что привело к необходимости защищать CAPTCHA на основе изображений.

Вследствие постоянного развития технологий, существующие методы защиты ограниченно способны противостоять системам компьютерного зрения, и часто неудобны для человека [8, 9], нарушая основные требования к построению тестов CAPTCHA. В связи с чем, в современном мире сохраняется потребность в разработке методов защиты CAPTCHA, способных соблюдать баланс между удобством решения человеком и сложностью решения автоматизированными методами.

Проблемы построения систем защиты от роботов и спама в контексте развития современных технологий

В общем виде, основной особенностью правильного построения теста CAPTCHA является обоюдное соблюдение следующих условий:

- во-первых, тесты CAPTCHA должны обеспечивать использование задач, требующих обработки информации с участием человека (англ. human-

in-the-loop processing). Генерация теста CAPTCHA должна специализироваться на задачах, ограниченно возможных или невозможных для автоматизированного решения, но доступных для решения человеком. За счет выполнения данного условия, в значительной степени обеспечивается защищенность теста CAPTCHA.

Во-вторых, тесты CAPTCHA должны обеспечивать удобство использования их человеком, их задачи должны минимизировать сложность их решения пользователем, а также быть доступными для людей разных языковых групп, культурных семейств, и для людей с ограниченными физическими возможностями. На удобство использования теста CAPTCHA человеком значительно влияют такие показатели, как средняя скорость прохождения теста CAPTCHA, соотношение числа успешно пройденных проверок и общего количества попыток (сложность CAPTCHA), доступность, и уровень удовлетворенности пользователей теста [10].

С появлением глубоких нейронных сетей стало значительно труднее формировать тесты CAPTCHA, сложные для машинной обработки и простые для человека. Глубокие нейронные сети представляют собой мощный инструмент машинного обучения, который пытается повторить мыслительный процесс человека [11], в связи с чем способен успешно решать ранее признанные сложными для автоматизации задачи, такие, как распознавание и классификация изображений [12, 13].

Доказательством значительного влияния глубоких нейронных сетей на усложнение построения эффективных тестов CAPTCHA является активное развитие исследований, посвященных обходу тех или иных видов CAPTCHA на основе нейросетевых методов и их модификаций [14, 15]. Скачок в развитии глубоких нейронных сетей и методов обхода CAPTCHA на основе нейросетевого подхода, соответственно, привёл к появлению методов защиты CAPTCHA от распознавания нейронными сетями [16, 17].

Рассмотрим некоторые репрезентативные методы защиты от обработки нейросетями, используемые для тестов CAPTCHA на основе изображений:

1) Усложнение семантической сложности задачи. Пользователь классифицирует или распознаёт предложенные контрольные изображения, но с дополнительными условиями, такими, как подбор пар, где представлены реальное и художественное изображение объекта [18], или прохождение небольшой игры [19].

2) Использование дополнительных шумовых элементов, искажений и деформаций изображения. Для усложнения распознавания производится покрытие изображений шумом различного характера в виде синусоидальных колебаний, случайных точек, перекрывающих узоров [20], а также изменение ориентации изображения или представление в виде сложных геометрических фигур [21].

3) Сегментация изображений. CAPTCHA, защищенные с помощью методов сегментации изображений, представляют набор изображений, часто разделенный на отдельные части или сегменты единый объект, который может быть более легко распознан и интерпретирован человеком, нежели компьютером. Часто тесты CAPTCHA с сегментированными изображениями предлагают пользователю задачу правильного расположения сегментов изображения [22], или выбора сегментов, представляющих в совокупности необходимый к распознанию объект или класс.

Несмотря на активное развитие вариаций методов защиты CAPTCHA, основанных на контрольных изображениях, большинство данных методов характеризуется определенными недостатками. Данные методы защиты часто не предоставляют оптимального повышения уровня безопасности, о чем может свидетельствовать, например, значительное количество исследований, посвященных обходу повсеместно используемого теста

Google reCAPTCHA v2, эксплуатирующего большинство современных методов защиты [23-25].

Более того, существенным недостатком текущих методов защиты является отрицательное влияние на удобство прохождения теста CAPTCHA рядовым пользователем системы [9].

Рассмотрим визуальное представление методов защиты теста CAPTCHA, основанного на изображениях, в сравнении с классической, незащищенной вариацией теста (рис.2):

Рис. 2. – Возможные методы защиты теста CAPTCHA, основанного на распознавании и классификации изображений

Достаточно просто заметить, как текущие методы защиты CAPTCHA на основе изображений искажают и усложняют визуальное представление самого теста, что безусловно отражается в показателях удобства теста для пользователя. Например, по данным маркетингового издательства AcquireConvert (UK, Hull), представленным в 2020 году, использование сложных тестов CAPTCHA может снизить показатель конверсии сайта на 33% [26]. Также, по данным исследования Клемсоновского университета и Центрального университета Флориды, проведенного в 2019 году, среднее время прохождения современных CAPTCHA, основанных на изображениях, занимает значительные 14-16 секунд [27]. Более того, исследование одного из ключевых исследователей рынка информационных технологий Forrester

(USA, Cambridge), проведенное в 2022 году, подтвердило, что в среднем каждый пятый человек в США прекращает заполнение веб-формы из-за сложностей, возникающих в преодолении теста CAPTCHA (рис. 3).

Рис. 3. – Процент пользователей из США, прекращающих заполнение веб-формы при попытке пройти тест CAPTCHA [28]

Более того, усложнение визуальной и когнитивной составляющей тестов CAPTCHA также оказывает значительное влияние на удобство прохождения теста людьми с ограниченными возможностями, в том числе, с нарушениями зрения и когнитивными расстройствами [29]. Например, исследование ресурса WebAIM (USA, Logan), проведенное в 2017 году, показало, что в последние годы одним из наиболее проблемных элементов веб-интерфейса для пользователей с нарушениями зрения являются тесты CAPTCHA (рис. 4).

Можно сделать вывод о том, что в современных реалиях создание эффективных методов защиты CAPTCHA, основанных на классификации изображений, является более комплексной задачей, требующей не столько усложнения задач, предлагаемых в тесте, сколько поиска способов эксплуатировать принципиальные различия в работе нейронных сетей и мышления человека. При этом текущие методы защиты CAPTCHA на основе изображений свидетельствуют о том, что существует необходимость в

меньшей степени понижать удобство прохождения теста CAPTCHA для пользователя.

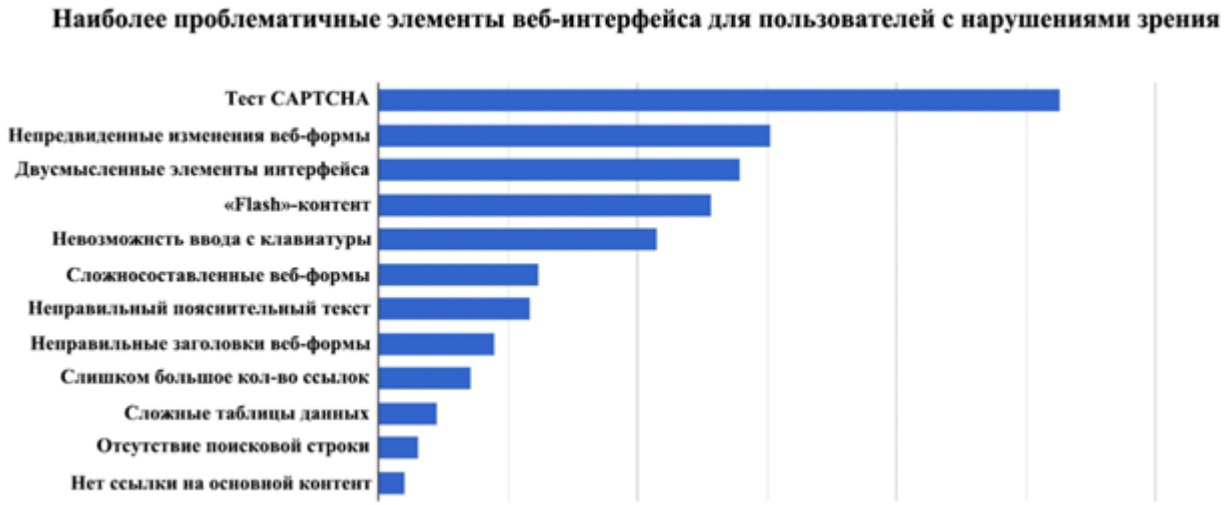

Рис. 4. – Наиболее проблематичные элементы веб-интерфейса для пользователей с нарушениями зрения [30]

Сформируем основные требования к построению более эффективных методов защиты CAPTCHA в контексте современных условий:

1) Предоставляемая технология защиты должна обеспечивать высокую устойчивость теста к обходу с помощью моделей нейронных сетей.

2) Предоставляемая технология защиты должна оказывать минимальное влияние на сложность прохождения теста CAPTCHA для пользователя.

Использование методов универсальных состязательных атак как способа повышения эффективности тестов CAPTCHA

При учете сформированных ранее требований к построению эффективных методов защиты CAPTCHA, потенциально эффективным решением могут стать методы, основанные на эксплуатации уязвимостей глубоких нейронных сетей [31-33]. Исследования в данной области смогли доказать подверженность глубоких нейронных сетей состязательным атакам

– незначительным изменениям во входном изображении, способным привести к неправильной классификации исходного изображения целевой моделью [34, 35].

В отличие от стандартных методов защиты контрольных изображений, используемых в тестах CAPTCHA, состязательные атаки способны учитывать особенности работы нейронных сетей в обнаружении образов и отдельных признаков в изображениях, которые сложно или невозможно заметить человеку, что должно оказывать минимальное влияние на сложность прохождения теста CAPTCHA человеком и значительно усложнять задачу для моделей нейронных сетей.

В общем случае, генерация состязательных атак основывается на применении методов оптимизации для поиска изменения, которое может быть добавлено к исходным данным, чтобы заставить модель нейронной сети неправильно классифицировать изображение.

Допустим, имеется модель нейронной сети с параметрами θ, которая обучена на обучающей выборке D_{train} и используется для классификации множества тестовых изображений X.

Цель состязательной атаки заключается в том, чтобы сгенерировать такое изменение δ, которое будет добавлено к каждому отдельному исходному тестовому изображению x и изменит изображение таким образом, чтобы модель присвоила ему неправильный класс.

Формально, задача состоит в том, чтобы найти такое изменение δ, которое минимизирует функцию потерь модели на измененном тестовом изображении $x' = x + \delta$ для неправильной метки y', как описано в формуле:

$$\min_{\delta} L(\theta, x + \delta, y'),$$

где x – исходное тестовое изображение; δ – изменение, вносимое в исходное тестовое изображение x; θ – параметры используемой модели нейронной

сети; y' – метка класса, не соответствующая изначальному классу изображения x; L – функция потерь модели, зависящая от параметров $\theta, x + \delta, y'$.

Визуально процесс защиты контрольного изображения при помощи выполнении состязательной атаки на модель нейронной сети можно отобразить следующим образом (рис.5).

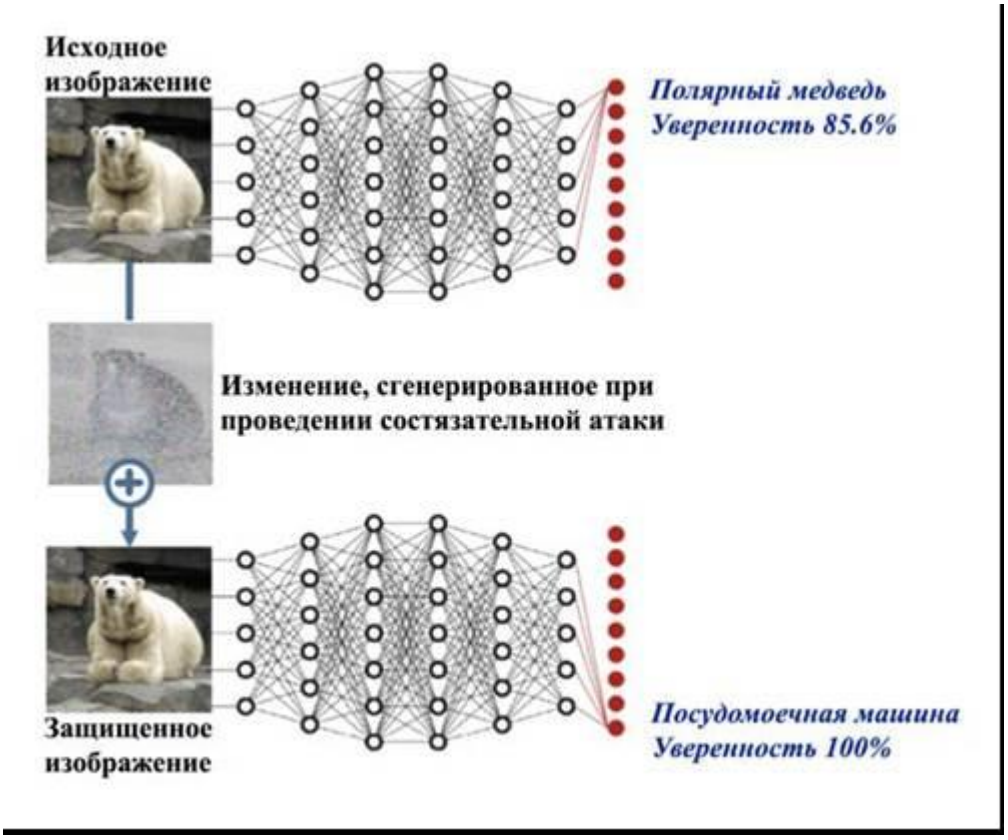

Рис. 5. – Визуальное представление применения состязательной атаки для защиты контрольного изображения

В процессе работы над алгоритмами проведения состязательных атак было сформировано значительное количество способов и вариаций генерации изменений исходных изображений, способных повлиять на ответ классификатора [36, 37]. Исходя из требований к построению эффективного теста CAPTCHA, основным условием подбора алгоритма является сохранение максимального удобства пользователя при обеспечении высокого

уровня защиты. В таком случае, одной из наиболее эффективных вариаций алгоритмов генераций состязательных атак является алгоритм DeepFool [36]. Оптимальность использования данного алгоритма может быть обоснована тем, что алгоритм DeepFool основан на решении задачи поиска минимального изменения входного изображения, что представлено в формуле:

$$\Delta(x;\ \hat{k}) := \min\|\delta^2\| \rightarrow\ \hat{k}(x+\delta)\ \neq \hat{k}(x),$$

где x – исходное тестовое изображение; δ – изменение, вносимое в исходное тестовое изображение x; $x+\delta$ – тестовое изображение x, искаженное изменением δ; $\hat{k}$ – используемая модель нейронной сети; $\hat{k}(x)$ – класс, присвоенный моделью $\hat{k}$ исходному изображению x.

Алгоритм DeepFool обеспечивает минимальное искажение изображения за счет поиска кратчайшего расстояния между гиперплоскостью, разделяющей классы, и исходной точкой. Производится последовательное перемещение точки на данное расстояние вдоль перпендикулярного направления от гиперплоскости. Процесс продолжается до тех пор, пока исходная точка не пересечет гиперплоскость и не будет идентифицирована моделью как другой класс (рис.6).

Данная особенность DeepFool может позволить оказывать минимальное влияние на удобство процесса прохождения теста CAPTCHA пользователем. Для наглядности представления рассмотрим изменение изображения, генерируемое алгоритмом DeepFool, в сравнении с тем же изображением с изменением, генерируемым алгоритмом FGSM, влияющим на характеристики каждого пикселя входного изображения (рис.7).

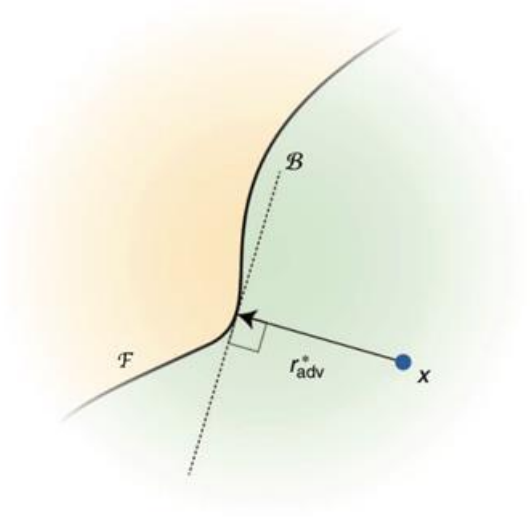

Рис. 6. – Геометрическая интерпретация работы алгоритма DeepFool

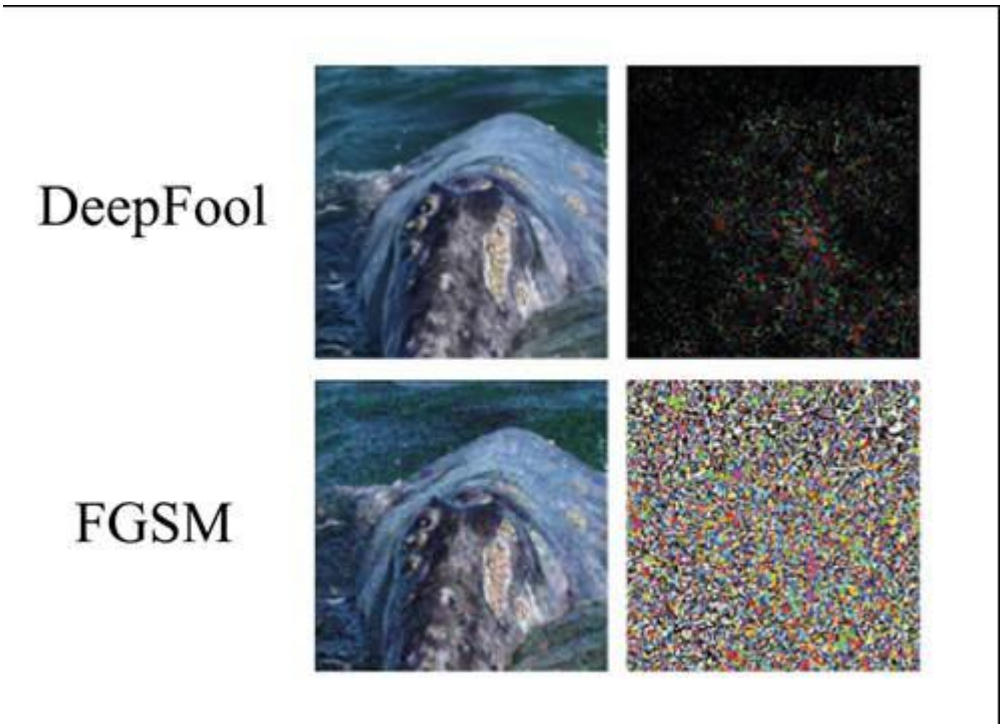

Рис. 7. – Различие в объеме генерируемых изменений изображения на примере алгоритмов DeepFool и FGSM

При минимальном искажении контрольных изображений, алгоритм Deepfool способен также оказывать значительное влияние на снижение точности и вероятность «обмана» классификатора, что доказано в

исследовании университета Циньхуа по сравнению эффективности алгоритмов состязательных атак [38].

Стоит принять во внимание, что существует важная деталь в работе алгоритмов генерации универсальных состязательных атак, значительно осложняющая задачи «встраивания» алгоритма в тест CAPTCHA. Алгоритмы генерации состязательных атак способны формировать изменения изображений, основываясь на выходе конкретной известной модели нейронной сети, что даёт возможность защитить изображения от обработки одним конкретным типом классификатора. В условиях защиты теста CAPTCHA владельцем модели нейронной сети является атакующий, что не даёт возможности понять, на основе выходных значений какой модели должны быть сгенерированы изменения изображений.

Данную проблему можно преодолеть, модифицировав алгоритм DeepFool таким образом, чтобы генерируемое изменение изображений минимально зависело от используемой в процессе работы алгоритма модели. То есть, генерируемое изменение должно стать независимым или частично зависимым от используемой модели (model-agnostic).

Изменения, генерируемые алгоритмом DeepFool, можно сделать незначительно зависимыми от целевой модели, если использовать данный алгоритм как основу алгоритма генерации универсальных состязательных атак – Universal Adversarial Perturbation (далее UAP) [39]. Алгоритм UAP изначально разрабатывался как метод генерации состязательных атак, не зависящих от конкретного используемого изображения, тем самым позволяя применять одно сгенерированное изменение к любому изображению тестового набора данных.

Однако, создатели алгоритма UAP экспериментальным путём смогли доказать, что независимые от используемого исходного изображения состязательные атаки также показывают универсальность генерируемых

изменений относительно разных моделей, что объясняется геометрическими корреляциями между многомерными границами областей принятия решений различных классификаторов [39]. Применительно к алгоритму UAP, алгоритм DeepFool используется для нахождения такого минимального изменения δ, которое переводит исходное изображение x в область классификационной плоскости, где модель классификатора f перестаёт относить изображение к изначальному классу y. Универсальное изменение, генерируемое алгоритмом UAP, представляется, как усреднённое представление минимальных изменений для каждого изображения из тестового набора данных. Таким образом, универсальное изменение δ_u можно описать следующим образом, описанным в формуле:

$$\delta_u = \frac{1}{N} \sum_{i=1}^{N} \delta_i,$$

где δ_u – универсальное изменение; δ_i – минимальное изменение, применимое к тестовому изображению для изменения решения классификатора; N – количество изображений во входном наборе данных.

Так как модели классификаторов используют схожие признаки для классификации изображений, универсальное изменение может быть эффективно для нарушения работы нескольких моделей нейронных сетей одновременно. Это объясняется тем, что универсальное изменение является результатом усреднения переклассификаций для множества разных изображений, что позволяет нарушать работу различных моделей нейронных сетей с высокой точностью и «усреднять» состязательные атаки, проводимые относительно них.

Оценка эффективности применения методов универсальных состязательных атак в тестах CAPTCHA

Для подтверждения гипотезы о том, что универсальные состязательные атаки могут быть применимы в задачах повышения эффективности тестов

CAPTCHA, необходимо произвести дополнительную оценку. Требуется оценить степень влияния универсальных состязательных атак на основные условия разработки эффективного теста CAPTCHA – обеспечение предоставления задач, простых для выполнения пользователем и сложных для выполнения автоматизированными средствами, в частности, моделями нейронных сетей, а также обеспечение максимального удобства прохождения теста для рядового пользователя теста CAPTCHA.

Произведем эмпирическую оценку степени влияния универсальных состязательных атак на удобство визуального представления теста CAPTCHA относительно других ранее рассмотренных методов защиты тестов CAPTCHA, основанных на изображениях (рис.8).

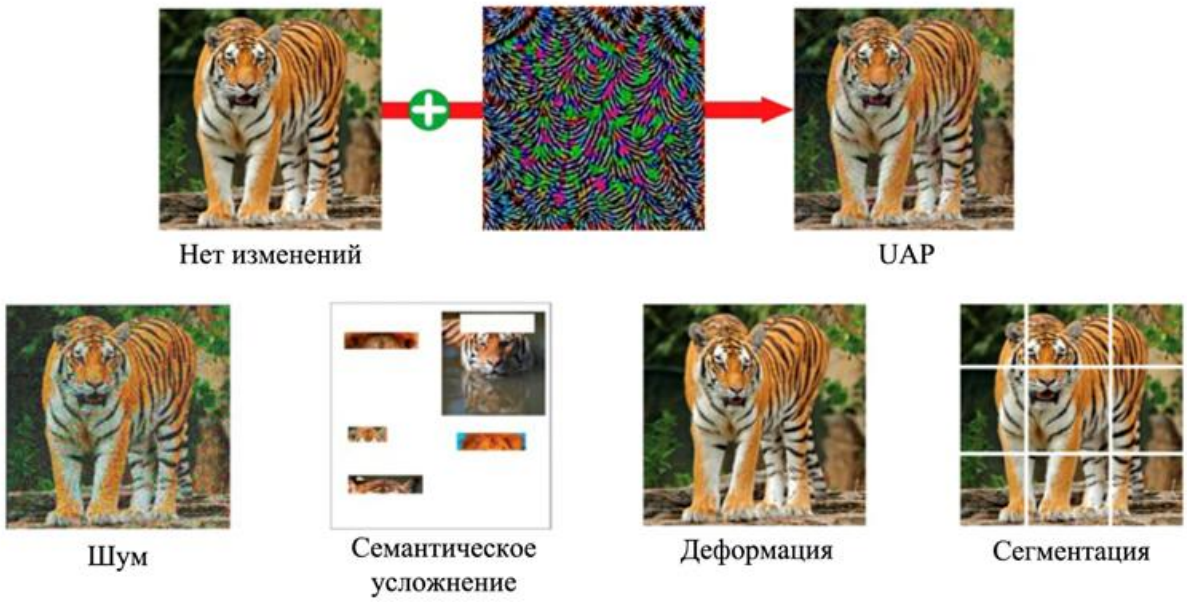

Рис. 8. – Сравнение степени искажения изображений в тесте CAPTCHA при применении текущих алгоритмов защиты и универсальных состязательных атак

Исходя из полученных результатов, можно сделать вывод о том, что применение универсальных состязательных атак для защиты контрольных изображений в CAPTCHA приводит к существенно меньшему изменению исходного изображения в сравнении с текущими методами защиты, что

должно позитивно сказываться на удобстве прохождения теста CAPTCHA конечным пользователем.

Также необходимо оценить, насколько применение универсальных состязательных атак повышает защищенность теста CAPTCHA при попытке преодоления теста на основе эксплуатации моделей нейронных сетей. Для оценки повышения защищенности теста CAPTCHA с помощью универсальных состязательных атак, проведём сравнительный эксперимент.

Построим модель потенциальной атаки злоумышленника на тест CAPTCHA с целью возможности преодоления теста автоматизированными средствами. Сначала будет оценена возможность успешного совершения атаки на стандартный тест CAPTCHA, использующий задачу классификации изображений. Затем будет повторно оценена возможность успешного совершения атаки на тот же тест CAPTCHA, но использующий контрольные изображения, предварительно покрытые шумом, сгенерированным с помощью проведения универсальной состязательной атаки.

Так как злоумышленник может использовать различные модели нейронных сетей для классификации изображений теста CAPTCHA, для точности проведения эксперимента будут использованы пять моделей, основанных на современных типах архитектуры нейронных сетей: ResNet-152, EfficientNet, DenseNet, VGG16, VGG19.

При проведении первого эксперимента, на вход каждой из моделей будет тысячу раз подаваться девять случайных изображений из набора изображений CIFAR-10, моделирующих стандартное количество изображений в тесте CAPTCHA на основе задачи классификации изображений (сетка 3x3). Задача моделируемого теста CAPTCHA будет представлять выбор трёх изображений указанного в задаче класса, поэтому три из девяти изображений будут принадлежать указанному в задаче классу,

а остальные шесть изображений будут принадлежать любому классу, кроме указанного.

При проведении второго эксперимента будет сохранена базовая структура эксперимента, но изменена конфигурация изображений подаваемых на вход – каждое изображение будет защищено с помощью изменения, сгенерированного на основе универсальной состязательной атаки. Так как алгоритм защиты на основе универсальных состязательных атак требует генерации изменения изображений на основе одной конкретной модели, второй эксперимент будет проведён пять раз – изменение изображений будет каждый раз генерироваться на основе одной определенной модели нейронной сети, но подаваться на вход каждой из пяти моделей. Данная конфигурация второго эксперимента используется для оценки возможности защищенных изображений, сгенерированных на основе выхода одной конкретной модели нейронной сети, нарушать в равной или меньшей степени работу любой другой модели нейронной сети, что должно доказать способность метода защиты на основе универсальных состязательных атак противостоять работе различных моделей нейронных сетей.

Метрикой оценки успешности атаки на CAPTCHA будет представлено отношение количества успешных прохождений генерируемого теста моделью нейронной сети к общему количеству попыток. Успешным прохождением генерируемого теста признаётся правильная классификация трёх изображений, относящихся к указанному в задаче классу, и отметка оставшихся шести контрольных изображений любым классом, кроме указанного в задаче. Таким образом, модель нейронной сети сможет подтвердить успешное автоматизированное прохождение теста CAPTCHA, выполнив задачу выбора среди девяти изображений трёх изображений, относящихся к указанному в задаче классу.

Проведём первый эксперимент, подавая оригинальные изображения на вход пяти заранее обозначенных моделей. Результаты измерений успешности прохождения незащищенного теста CAPTCHA представлены в таблице 1, где приведен процент успешно пройденных моделями нейронных сетей тестов CAPTCHA, использующих стандартные изображения в рамках моделируемой атаки.

Таблица № 1

Результаты прохождения незащищенного теста

ResNet-152	EfficientNet	DenseNet	VGG16	VGG19
74.6%	75.6%	72.7%	61.3%	64.5%

Исходя из результатов первого эксперимента, можно отметить, что современные модели нейронных сетей способны с высокой точностью классифицировать незащищенные контрольные изображения, позволяя автоматизировать решение стандартных тестов CAPTCHA, основанных на задаче классификации изображений, что нарушает основные принципы построения эффективных тестов CAPTCHA.

Проведём второй эксперимент, подавая изображения, защищенные с помощью универсальной состязательной атаки, на вход пяти заранее обозначенных моделей. Эксперимент проводится пять раз, так как генерация изменений исходных изображений каждый раз проводится на основе выходных данных одной конкретной модели нейронной сети из пяти выбранных, но подаётся на вход каждой из пяти моделей. Результаты эксперимента представлены в таблице 2, где представлен процент успешно пройденных моделями нейронных сетей тестов CAPTCHA, использующих изображения, защищенные на основе изменений, сгенерированных универсальной состязательной атакой.

В столбцах таблицы представлены модели, производившие атаку на моделируемый тест CAPTCHA, а в строках представлены модели, генерировавшие изменения для контрольных изображений теста.

Таблица № 2

Результаты прохождения защищенного теста

	ResNet-152	**EfficientNet**	**DenseNet**	**VGG16**	**VGG19**
ResNet-152	**5.3%**	14.7%	12.5%	17.1%	16.7%
EfficientNet	14.2%	**3.5%**	11.3%	21.5%	19.3%
DenseNet	11.8%	12.2%	**4.8%**	17.4%	15.3%
VGG16	18.1%	16.6%	17.2%	**3.1%**	9.3%
VGG19	18.3%	14.5%	15.9%	8.5%	**2.3%**

Исходя из данных второго эксперимента, можно сделать вывод о том, что применение изменений, сгенерированных с помощью универсальных состязательных атак, к контрольным изображениям теста CAPTCHA, способно кратно понизить вероятность преодоления теста CAPTCHA моделями нейронных сетей.

Использование универсальных состязательных атак также доказало возможность применения состязательных атак в разрезе защиты контрольных изображений теста CAPTCHA от обработки различными моделями нейронных сетей. В случаях, когда атака проводилась при помощи модели, отличной от той, на основе выходных значений которой генерировалось защитное изменение изображений, были представлены менее впечатляющие, но все еще высокие показатели защищенности теста CAPTCHA.

Выводы

Была доказана возможность применения универсальных состязательных атак в задачах повышения эффективности и устойчивости

тестов CAPTCHA, основанных на задаче классификации изображений. Проведённый анализ текущих проблем, связанных с поддержанием высокого уровня безопасности и эффективности тестов CAPTCHA, позволил определить основные направляющие векторы в улучшении работоспособности тестов CAPTCHA, и на основе данных требований разработать новый метод повышения устойчивости и эффективности систем, используемых в задачах защиты от роботов и спама. Однако, несмотря на положительные результаты применения универсальных состязательных атак, в задачах повышения эффективности и устойчивости систем защиты от роботов и спама, все еще существует перечень потенциально возможных проблем. Необходимо исследовать устойчивость изображений, защищенных с помощью универсальных состязательных атак, к обработке моделями нейронных сетей, использующих слои, позволяющие незначительно искажать входное изображение для нарушения работы ранее примененных к изображению алгоритмов защиты. Потенциальными мерами противодействия универсальным состязательным атакам могут быть встроенные в атакующую модель нейронной сети слои, позволяющие производить значительное изменение размеров входного изображения, или покрытие изображения незначительным количеством пикселей. Также необходимо исследовать возможность обучения нейронных сетей на образцах изображений, защищенных с помощью универсальных состязательных атак. Существует потенциальный шанс получения нейронной сети, способной адаптировать логику распознавания изображений к существованию в изображениях изменений, сгенерированных на основе универсальных состязательных атак. Поэтому дальнейшее совершенствование и исследование применения универсальных состязательных атак в задачах повышения эффективности и устойчивости систем защиты от роботов и спама является актуальной задачей.

Литература

1. 2022 State of Bot Mitigation // Kasada URL: kasada.io/2022-state-of-bot-mitigation/ (date accessed: 02.05.2023).

2. Igbekele E.O., Adebiyi A.A., Ibikunle F.A., Adebiyi M.O., Olugbara O.O. Research trends on CAPTCHA: A systematic literature // International Journal of Electrical and Computer Engineering. 2021. T. 11. № 5. pp. 4300-4312.

3. Dinh N.T., Hoang V.T. Recent advances of Captcha security analysis: a short literature review // Procedia Computer Science. 2023. T. 218. pp. 2550-2562.

4. Abdalla K., Kaya M. An evaluation of different types of CAPTCHA: effectiveness, user-friendliness, and limitations // International Journal of Scientific Research in Information Systems and Engineering. 2017. T. 2. № 3. pp. 12-19.

5. Явна Д.В. Компьютерное моделирование зрительных механизмов группирования, избирательных к пространственным модуляциям контраста // Инженерный вестник Дона, 2013, № 4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2009.

6. He K. Zhang X., Ren Sh., Sun J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification // Proceedings of the IEEE international conference on computer vision. 2015. pp. 1026-1034.

7. Цыганков В.А., Шабалина О.А., Катаев А.В. Использование генетических алгоритмов для повышения скорости обучения нейронных сетей // Инженерный вестник Дона, 2023, № 2. URL: ivdon.ru/ru/magazine/archive/n2y2023/8207.

8. Bursztein E., Bethard S., Fabry C., Mitchell J.C. Jurafsky D. How good are humans at solving CAPTCHAs? A large scale evaluation // 2010 IEEE symposium on security and privacy. IEEE, 2010. pp. 399-413.

9. Shirali-Shahreza S., Shirali-Shahreza M.H. Accessibility of CAPTCHA methods // Proceedings of the 4th ACM workshop on security and artificial intelligence. 2011. pp. 109-110.

10. Beheshti S.M.R.S., Liatsis P. CAPTCHA Usability and Performance, How to Measure the Usability Level of Human Interactive Applications Quantitatively and Qualitatively? // 2015 International Conference on Developments of E-Systems Engineering (DeSE). IEEE, 2015. pp. 131-136.

11. Rudra S. An AI completes an unfinished composition 115 years after composer's death // Vice URL: vice.com/en/article/neakqm/an-ai-completes-an-unfinished-composition-115-years-after-composers-death (date accessed 02.05.2023).

12. Wang Z., Shi P. CAPTCHA recognition method based on CNN with focal loss // Complexity. 2021. T. 2021. pp. 1-10.

13. Stark F., Hazırba¸s C., Triebel R., Cremers D. Captcha recognition with active deep learning // Workshop new challenges in neural computation. 2015. T. 2015. pp. 95-102.

14. Noury Z., Rezaei M. Deep-CAPTCHA: a deep learning based CAPTCHA solver for vulnerability assessment // URL: doi.org/10.48550/arXiv.2006.08296 (date accessed 02.05.2023).

15. Kimbrough T., Tian P., Liao W., Blasch E., Yu W. Deep CAPTCHA Recognition Using Encapsulated Preprocessing and Heterogeneous Datasets // IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2022. pp. 1-6.

16. Dinh N., Nguyen T., Truong V. zxCAPTCHA: New Security-Enhanced CAPTCHA // 2023 15th International Conference on Knowledge and Smart Technology (KST). IEEE, 2023. pp. 1-6.

17. Dushyant K., Muskan G., Annu, Gupta A., Pramanik S. Utilizing Machine Learning and Deep Learning in Cybesecurity: An Innovative Approach // Cyber Security and Digital Forensics. 2022. pp. 271-293.

18. Vikram S., Fan Y., Gu G. SEMAGE: a new image-based two-factor CAPTCHA // Proceedings of the 27th Annual Computer Security Applications Conference. 2011. pp. 237-246.

19. Yu H., Riedl M.O. Automatic Generation of Game-based CAPTCHAs // Proceedings of the FDG workshop on Procedural Content Generation URL: pcgworkshop.com/archive/yu2015generation.pdf (date accessed 02.05.2023).

20. Zhang Y. Gao H., Pei G., Luo S., Chang G., Cheng N. A survey of research on captcha designing and breaking techniques // 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, 2019. pp. 75-84.

21. Kim S.Y., Kim S., Cho H.G. A New Human Interactive Proof System Using Arbitrary and Fractal Polygon Image // 2011 IEEE 11th International Conference on Computer and Information Technology. IEEE, 2011. pp. 261-268.

22. Ali F.A.B.H., Karim F.B. Development of CAPTCHA system based on puzzle // 2014 International Conference on Computer, Communications, and Control Technology (I4CT). IEEE, 2014. pp. 426-428.

23. Wang D., Moh M., Moh T.S. Using deep learning to solve google recaptcha v2's image challenges // 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM). IEEE, 2020. pp. 1-5.

24. Hossen M.I., Tu Y., Rabby F., Islam N., Cao H., Hei X. An object detection based solver for google's image recaptcha v2 // 23rd International Symposium on Research in Attacks, Intrusions and Defenses URL: usenix.org/system/files/raid20-hossen.pdf (date accessed 02.05.2023).

25. Hossen M. I. Tu Y., Rabby F., Islam N., Cao H. Bots Work Better than Human Beings: An Online System to Break Google's Image-based reCaptcha v2 // URL: regmedia.co.uk/2019/09/11/recaptcha.pdf (date accessed: 02.05.2023).

26. Giles T. [Form Conversion] An In-Depth Guide To Form Optimization // URL: acquireconvert.com/conversion-form/ (date accessed: 02.05.2023).

27. Madathil K.C., Greenstein J.S., Horan K. Empirical studies to investigate the usability of text-and image-based CAPTCHAs // International Journal of Industrial Ergonomics. 2019. T. 69. pp. 200-208.

28. Carrieli S., DeMartine A., Raposo I., Dostie P. We All Hate Captchas, Except When We Don't // Forrester URL: forrester.com/report/we-all-hate-captchas-except-when-we-dont/RES177777 (date accessed 02.05.2023).

29. Berton R., Gaggi O., Kolasinska A., Palazzi C.E., Quadrio G. Are captchas preventing robotic intrusion or accessibility for impaired users? // 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC). IEEE, 2020. pp. 1-6.

30. Screen Reader User Survey #7 Results // WebAim URL: webaim.org/projects/screenreadersurvey7/ (date accessed 02.05.2023).

31. Ye G., Tang Zh., Fang D., Zhu Zh., Feng Y., Xu P., Chen X., Wang Zh. Yet another text captcha solver: A generative adversarial network based approach // Proceedings of the 2018 ACM SIGSAC conference on computer and communications security. 2018. pp. 332-348.

32. Du F.L., Li J.-X., Yang Zh., Chen P., Wang B., Zhang J. CAPTCHA recognition based on faster R-CNN // Intelligent Computing Theories and Application: 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part II 13. Springer International Publishing, 2017. pp. 597-605.

33. Shu Y., Xu Y. End-to-End Captcha Recognition Using Deep CNN-RNN Network // 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). IEEE, 2019. pp. 54-58.

34. Fawzi A., Moosavi-Dezfooli S.M., Frossard P. The robustness of deep networks: A geometrical perspective // IEEE Signal Processing Magazine. 2017. T. 34. № 6. pp. 50-62.

35. Cubuk E.D., Zoph B., Schoenholz S.S., Le Q.V. Intriguing properties of adversarial examples // URL: doi.org/10.48550/arXiv.1711.02846 (date accessed 02.05.2023).

36. Abdollahpourrostam A., Abroshan M., Moosavi-Dezfooli S.M. Revisiting DeepFool: generalization and improvement // URL: doi.org/10.48550/arXiv.2303.12481 (date accessed 02.05.2023).

37. Papernot N., McDaniel P., Jha S., Fredrikson M., Celik Z.B., Swami A. The limitations of deep learning in adversarial settings // 2016 IEEE European symposium on security and privacy (EuroS&P). IEEE, 2016. pp. 372-387.

38. Dong Y., Fu Q.-A., Yang X., Pang T., Su H., Xiao Z., Zhu J. Benchmarking adversarial robustness on image classification // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. pp. 321-331.

39. Moosavi-Dezfooli S.M., Fawzi A., Fawzi O., Frossard P. Universal adversarial perturbations // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 1765-1773.

References

1. 2022 State of Bot Mitigation URL: kasada.io/2022-state-of-bot-mitigation/ (date accessed: 02.05.2023).

2. Igbekele E.O., Adebiyi A.A., Ibikunle F.A., Adebiyi M.O., Olugbara O.O. International Journal of Electrical and Computer Engineering. 2021. V. 11. № 5. pp. 4300-4312.

3. Dinh N.T., Hoang V.T. Procedia Computer Science. 2023. V. 218. pp. 2550-2562.

4. Abdalla K., Kaya M. International Journal of Scientific Research in Information Systems and Engineering. 2017. V. 2. № 3. pp. 12-19.

5. Yavna D.V. Inzhenernyj vestnik Dona, 2013, № 4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2009.

6. He K. Zhang X., Ren Sh., Sun J. Proceedings of the IEEE international conference on computer vision. 2015. pp. 1026-1034.

7. Tsygankov V.A., Shabalina O.A., Kataev A.V. Inzhenernyj vestnik Dona, 2023, № 2. URL: ivdon.ru/ru/magazine/archive/n2y2023/8207.

8. Bursztein E., Bethard S., Fabry C., Mitchell J.C. Jurafsky D. 2010 IEEE symposium on security and privacy. IEEE, 2010. pp. 399-413.

9. Shirali-Shahreza S., Shirali-Shahreza M.H. Proceedings of the 4th ACM workshop on security and artificial intelligence. 2011. pp. 109-110.

10. Beheshti S.M.R.S., Liatsis P. 2015 International Conference on Developments of E-Systems Engineering (DeSE). IEEE, 2015. pp. 131-136.

11. Rudra S. An AI completes an unfinished composition 115 years after composer's death URL: vice.com/en/article/neakqm/an-ai-completes-an-unfinished-composition-115-years-after-composers-death (accessed: 02.05.2023).

12. Wang Z., Shi P. Complexity. 2021. V. 2021. pp. 1-10.

13. Stark F., Hazırba¸s C., Triebel R., Cremers D. Workshop new challenges in neural computation. 2015. V. 2015. pp. 95-102.

14. Noury Z., Rezaei M. Deep-CAPTCHA: a deep learning based CAPTCHA solver for vulnerability assessment. URL: doi.org/10.48550/arXiv.2006.08296 (accessed: 02.05.2023).

15. Kimbrough T., Tian P., Liao W., Blasch E., Yu W. IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2022. pp. 1-6.

16. Dinh N., Nguyen T., Truong V. 15th International Conference on Knowledge and Smart Technology (KST). IEEE, 2023. pp. 1-6.

17. Dushyant K., Muskan G., Annu, Gupta A., Pramanik S. Cyber Security and Digital Forensics. 2022. pp. 271-293.

18. Vikram S., Fan Y., Gu G. Proceedings of the 27th Annual Computer Security Applications Conference. 2011. pp. 237-246.

19. Yu H., Riedl M.O. Automatic Generation of Game-based CAPTCHAs URL: pcgworkshop.com/archive/yu2015generation.pdf (accessed: 02.05.2023).

20. Zhang Y. Gao H., Pei G., Luo S., Chang G., Cheng N. 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, 2019. pp. 75-84.

21. Kim S.Y., Kim S., Cho H.G. 2011 IEEE 11th International Conference on Computer and Information Technology. IEEE, 2011. pp. 261-268.

22. Ali F.A.B.H., Karim F.B. 2014 International Conference on Computer, Communications, and Control Technology (I4CT). IEEE, 2014. pp. 426-428.

23. Wang D., Moh M., Moh T.S. 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM). IEEE, 2020. pp. 1-5.

24. Hossen M.I., Tu Y., Rabby F., Islam N., Cao H., Hei X. 23rd International Symposium on Research in Attacks, Intrusions and Defenses URL: usenix.org/system/files/raid20-hossen.pdf (accessed: 02.05.2023).

25. Hossen M. I. Tu Y., Rabby F., Islam N., Cao H. Bots Work Better than Human Beings: An Online System to Break Google’s Image-based reCaptcha v2 URL: regmedia.co.uk/2019/09/11/recaptcha.pdf (accessed: 02.05.2023).

26. Giles T. [Form Conversion] An In-Depth Guide To Form Optimization URL: acquireconvert.com/conversion-form/ (accessed: 02.05.2023).

27. Madathil K.C., Greenstein J.S., Horan K. International Journal of Industrial Ergonomics. 2019. V. 69. pp. 200-208.

28. Carrieli S., DeMartine A., Raposo I., Dostie P. We All Hate Captchas, Except When We Don’t URL: forrester.com/report/we-all-hate-captchas-except-when-we-dont/RES177777 (accessed: 02.05.2023).

29. Berton R., Gaggi O., Kolasinska A., Palazzi C.E., Quadrio G. 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC). IEEE, 2020. pp. 1-6.

30. Screen Reader User Survey #7 Results. URL: webaim.org/projects/screenreadersurvey7/ (accessed: 02.05.2023).

31. Ye G., Tang Zh., Fang D., Zhu Zh., Feng Y., Xu P., Chen X., Wang Zh. Proceedings of the 2018 ACM SIGSAC conference on computer and communications security. 2018. pp. 332-348.

32. Du F.L., Li J.-X., Yang Zh., Chen P., Wang B., Zhang J. Intelligent Computing Theories and Application: 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part II 13. Springer International Publishing, 2017. pp. 597-605.

33. Shu Y., Xu Y. 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). IEEE, 2019. pp. 54-58.

34. Fawzi A., Moosavi-Dezfooli S.M., Frossard P. IEEE Signal Processing Magazine. 2017. V. 34. № 6. pp. 50-62.

35. Cubuk E.D., Zoph B., Schoenholz S.S., Le Q.V. Intriguing properties of adversarial examples URL: doi.org/10.48550/arXiv.1711.02846 (accessed: 02.05.2023).

36. Abdollahpourrostam A., Abroshan M., Moosavi-Dezfooli S.M. Revisiting DeepFool: generalization and improvement URL: doi.org/10.48550/arXiv.2303.12481 (accessed: 02.05.2023).

37. Papernot N., McDaniel P., Jha S., Fredrikson M., Celik Z.B., Swami A. 2016 IEEE European symposium on security and privacy (EuroS&P). IEEE, 2016. pp. 372-387.

38. Dong Y., Fu Q.-A., Yang X., Pang T., Su H., Xiao Z., Zhu J. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. pp. 321-331.

39. Moosavi-Dezfooli S.M., Fawzi A., Fawzi O., Frossard P. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 1765-1773.