DOI 10.37882/2223-2966.2023.07.14

СУЩЕСТВУЮЩИЕ ИССЛЕДОВАНИЯ, МЕТОДЫ СОЗДАНИЯ ПОДДЕЛЬНЫХ ИЗОБРАЖЕНИЙ И СПОСОБЫ ИХ ОБНАРУЖЕНИЯ

EXISTING RESEARCH, METHODS FOR CREATING FAKE IMAGES AND WAYS TO DETECT THEM

V. Eremuk
V. Romashov

Summary: With the development of advanced technologies in the field of programming, the possibility of creating fake images has arisen, which raises growing concerns about the authenticity of visual content. On the other hand, the detection of fake images has become essential to prevent the spread of misinformation and malicious intent. This thesis paper explores the algorithms used to create fake images and the algorithms used to detect them. Researchers have developed a variety of technical tools, including Photoshop software and deep learning algorithms, to generate fake images with a high degree of realism. At the same time, many detection methods have emerged, including metadata analysis, the use of computer vision algorithms, and the application of blockchain technology to ensure the authenticity of the image. According to recent research, deep learning is a promising approach for developing more accurate anomaly detection algorithms in images. As a consequence, it is critical to develop and continually improve detection methods to ensure the accuracy and authenticity of visual content.

Keywords: fake images, detection methods, computer vision algorithms, blockchain technology, accuracy, deep learning.

Еремук Владимир Вадимович
Аспирант, Университет ИТМО, г. Санкт-Петербург
polar.vl@yandex.ru
Ромашов Виктор Андреевич
Аспирант, Университет ИТМО, г. Санкт-Петербург

Аннотация. С развитием передовых технологий в области программирования возникла возможность создания поддельных изображений, что вызывает растущие опасения относительно подлинности визуального контента. С другой стороны, обнаружение поддельных изображений стало крайне важным для предотвращения распространения дезинформации и злонамеренных намерений. Эта тезисная работа исследует алгоритмы, используемые для создания фейковых изображений, и алгоритмы, используемые для их обнаружения. Исследователи разработали разнообразные технические средства, включая программное обеспечение Photoshop и алгоритмы глубокого обучения, предназначенные для генерации поддельных изображений с высокой степенью реалистичности. В то же время, появилось множество методов обнаружения, включая анализ метаданных, использование алгоритмов компьютерного зрения и применение технологии блокчейн для обеспечения подлинности изображения. Согласно последним исследованиям, глубокое обучение представляет собой перспективный подход для разработки более точных алгоритмов обнаружения аномалий в изображениях. Как следствие, крайне важно разрабатывать и постоянно улучшать методы обнаружения, чтобы обеспечить точность и подлинность визуального контента.

Ключевые слова: поддельные изображения, методы обнаружения, алгоритмы компьютерного зрения, технология блокчейн, точность, глубокое обучение.

Возможность создания поддельных изображений стала все более доступной с появлением новых технологий. Это привело к растущим опасениям относительно подлинности и достоверности визуального контента. Возможность создавать поддельные изображения не только подрывает журналистскую честность, но также может представлять угрозу национальной безопасности в случае распространения таких изображений, ведущей к распространению ложной информации. В то же время, возможность создания убедительных поддельных изображений также расширяет горизонты художественного самовыражения. Однако независимо от мотивов создания поддельных изображений, крайне важно разрабатывать и постоянно улучшать алгоритмы обнаружения таких изображений, чтобы обеспечить точность и подлинность визуального контента.

Для поддержания достоверности визуальных источников и прекращения распространения ложной информации, необходимо сохранять бдительность и готовность к угрозе, вызванной фальшивыми изображениями [5].

Подделка изображений может начинаться с двух точек:

1 — с реального изображения в качестве источника (использование реального изображения в качестве исходного материала, с последующим изменением его для получения поддельного изображения).

2 — из вектора шумовых точек: (Создание поддельного изображения с использованием вектора случайных точек, которые далее обрабатываются с помощью методов машинного обучения, чтобы получить наиболее реалистичное изображение).

В работе, проведенной J. Yang и соавторами [6], была рассмотрена актуальная проблема обнаружения поддельных изображений, и описана неотложность такого подхода. Генеративно-состязательные сети (GAN), показали свою способность в областях обработки изображе-

ний, звука и речи. Однако, несмотря на преимущества, передовые технологии, используемые в киберпреступлениях, не всегда безопасны и могут представлять угрозу. Технология Deepfake, основанная на генеративно-состязательных сетях (GAN), способна заменять лица различных людей, делая изображения визуально реальными. Deepfake позволяет целенаправленно подменять лицо одного человека другим, что может привести к широкому распространению фальшивых событий в Интернете и иметь серьезные последствия, такие как личные атаки и киберпреступления [6].

Основываясь на передовых исследованиях, авторы предложили разумный судебно-экспертный метод обнаружения Deepfake. Авторы сначала обнаружили тонкие различия в текстуре между реальным и поддельным изображением с помощью карты внимания к изображению, которая показывает различие в текстуре лиц. Для усиления этой разницы авторы использовали направленный фильтр с картой внимания как справочник для повышения артефактов текстуры, вызванных постобработкой, и отображения потенциальных функций фальсификации. Классификационная сеть Resnet18 эффективно учитывает эти различия и наконец реализует обнаружение реальных и поддельных изображений. Авторы оценили производительность метода, и эксперименты подтвердили, что предлагаемый способ позволяет достичь высокой точности обнаружения [10].

L. Choi и соавторы исследовали характеристики поддельных изображений, определяющие их обнаружаемость. Для визуализации областей, которые легко обнаруживает классификатор, был использован сегментационный классификатор. Помимо этого, была предложена техника усиления этих характеристик. Эксперименты показали, что классификатор поддельных изображений способен обнаруживать и классифицировать их даже при оптимальном проектировании генератора. Данное исследование подчеркивает важность обнаружения и классификации поддельных изображений среди других [3].

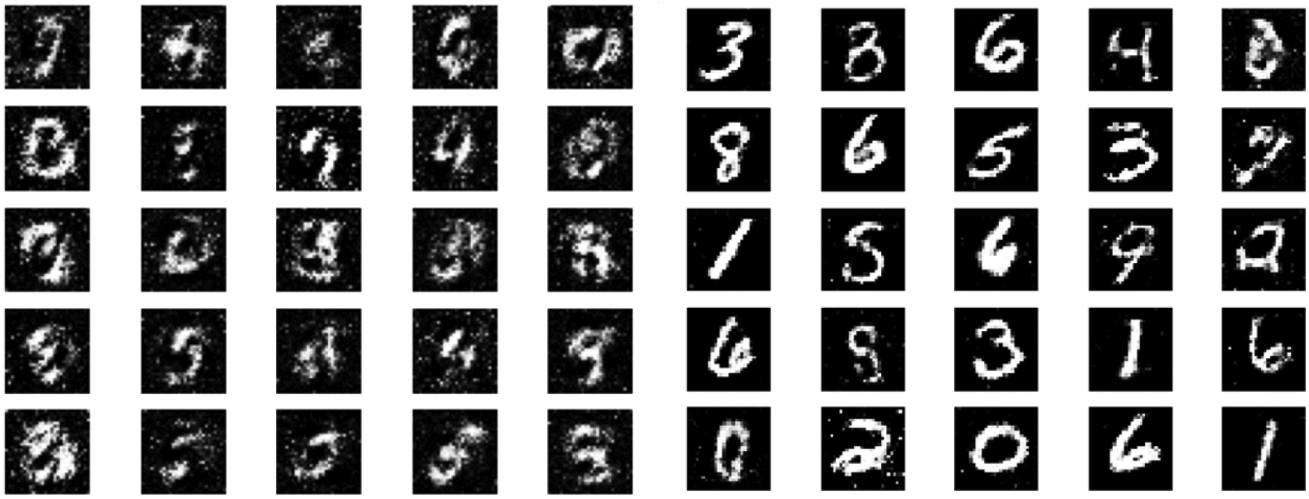
В работе X. Wang. и соавт., авторы сосредоточились на проблеме разработки техник обнаружения лиц на основе GAN, которые могут исследовать и выявлять поддельные лица. Они стремились предоставить всеобъемлющий обзор последних достижений в области обнаружения лиц, созданных с помощью GAN. Они сосредоточились на способах обнаружения поддельных лиц, созданных или синтезированных с помощью моделей GAN. Они разделили свою работу на категории, такие как обучение на основе глубокого обучения, оценка и сравнение с визуальной производительностью человека. Они изучили и кратко изложили особенности и ключевые идеи создания и обнаружения поддельных лиц на основе GAN [9].

Авторы С. Тарик и соавторы исследовали проблему обнаружения поддельных изображений лиц, которая является важной для предотвращения различных видов злоупотреблений. В своей работе они предложили систему форензики изображений под названием FakeFaceDetect, основанную на нейронных сетях, для обнаружения различных поддельных изображений лиц. Особое внимание авторы уделили обнаружению поддельных изображений, созданных с использованием генеративно-состязательных сетей (GAN). Для исследования авторы использовали поддельные изображения, полученные как из GAN, так и от людей. Результаты работы FakeFaceDetect демонстрировали высокую точность в обнаружении поддельных изображений [8].

В работе F.I. Alarsan и соавторов предлагается также использование GAN для выбора оптимальных гиперпараметров при создании поддельных изображений цифр от 1 до 9. Подход основан на обучении генератора и дискриминатора, чтобы определить, являются ли созданные изображения поддельными или подлинными. Техника генетического алгоритма (GA) использовалась для настройки гиперпараметров GAN, и полученный алгоритм назван GANGA: генеративно-состязательная сеть с генетическим алгоритмом. Результатом работы алгоритма стала высокая производительность; удалось получить значение функции потерь для генератора и дискриминатора равное нулю. Для решения данной проблемы была использована среда Anaconda с соответствующими необходимыми библиотеками. Для проверки работоспособности был использован набор данных MNIST. Предлагаемый метод заключается в выборе генетическим алгоритмом наилучших значений гиперпараметров при минимизации функции стоимости, такой как функция потерь, или максимизации точности, используемой для нахождения наилучших значений скорости обучения, нормализации пакетов, количества нейронов и параметра слоя исключения [1].

Начиная с шума, может потребоваться много шагов, чтобы приблизиться к целевому изображению. Для этой цели может использоваться модель, обученная преобразовывать шум в поддельные изображения, пример которой представлен на Рисунке 1; левое изображение находится во время подделки, правое изображение является поддельным изображением на последнем шаге.

Сгенерированные изображения взяты из хорошо известного набора данных под названием MNIST dataset. Набор данных MNIST представляет собой значимый ресурс, который используется для оценки эффективности методов генерации и обнаружения поддельных изображений цифр. Он содержит изображения рукописных цифр, и его использование в данном контексте позволяет проверить разработанные алгоритмы обнаружения поддельных изображений на основе сравнения с подлинными изображениями.



в течение поколения в конце

Рис. 1. Пример поддельных изображений из-за шума [4]

Появление поддельных изображений в эпоху цифровых технологий — это растущая проблема, которая может иметь серьезные последствия. В связи с прогрессом технологий, возможность создания убедительных поддельных изображений с определенными модификациями стала значительно более доступной, чем когда-либо прежде. Это приводит к потенциальному риску обмана для пользователей, поскольку они могут быть введены в заблуждение при восприятии таких изображений. Возможность обнаруживать, и точно классифицировать поддельные изображения становится все более важной для обеспечения целостности и достоверности визуального контента. В отсутствие адекватных методов обнаружения, ложная информация может оперативно распространяться, порождая ошибочные последовательности событий. В то же время, обнаружение поддельных изображений может помочь научным исследованиям и национальной безопасности (например, изображения могут быть подделаны с целью ввода следователей в заблуждение и отклонения от истинного хода расследования). Таким образом, разработка точных и эффективных методов обнаружения является ключевым моментом в обеспечении подлинности и надежности визуальной информации. Укрепляя способность обнаруживать поддельные изображения, можно гарантировать доверие к цифровым носителям, а также защиту от многих негативных последствий мошенничества с изображениями [6].

Новые подходы глубокого обучения для обнаружения поддельных изображений включают [2; 7]:

1. Применение генеративно-состязательных сетей (GAN) для генерации поддельных изображений и обучения сетей-детекторов с целью различения подлинных и поддельных изображений представляет собой активно исследуемый подход в научном сообществе. Этот подход имеет два дополнительных эффекта: генерация поддельных изображений и обнаружение поддельных изображений.
2. Применение алгоритмов машинного обучения, включая случайный лес (Random Forest) и XGBoost, для разработки классификаторов, способных определять поддельные изображения на основе значений их гиперпараметров. Для этого используются характеристики или признаки изображений, которые могут быть извлечены и использованы для обучения классификационных моделей.
3. Использование сверточных нейронных сетей (Convolutional Neural Networks) для автоматического извлечения признаков изображений и обнаружения поддельных элементов в поддельных изображениях.
4. Использование нескольких сетей для обнаружения поддельных изображений, таких как вариационные автоэнкодеры (Variational Autoencoder) и сетей Гауссовских смесей (Gaussian Mixture Networks), которые могут обнаруживать поддельные элементы, которые нельзя обнаружить при использовании только одной сети (например, лица и эмоции).

Указанные подходы могут улучшить возможность обнаружения поддельных изображений и защитить людей и информацию от потенциальных злоупотреблений.

Исследования по обнаружению поддельных изображений

Несмотря на некоторый прогресс в обнаружении поддельных изображений с помощью машинного обучения, все еще существуют проблемы, требующие решения и анализа [9], такие как:

— Недостаточное количество поддельных изображений в обучающих наборах данных. Это может

привести к тому, что модели будут недостаточно эффективными при обнаружении поддельных изображений.
— Сложность детектирования поддельных видео. Для обнаружения поддельных видео требуются дополнительные методы, такие как анализ движения или анализ звука.
— Возникает проблема обнаружения поддельных изображений с использованием генеративно-состязательных сетей. Такие поддельные изображения могут быть трудно отличимыми от настоящих, поскольку они могут обладать высокой степенью реалистичности.

Анализ и решение этих проблем требуют дополнительных исследований и разработок новых подходов и методов для улучшения существующих решений по обнаружению поддельных изображений.

В целом, обнаружение поддельных изображений является сложной задачей, которая требует использования различных методов и технологий. С ростом доступности мощных вычислительных ресурсов и развитием алгоритмов машинного обучения исследования в этой области стали более активными. Тем не менее, обнаружение поддельных изображений является важной задачей в нашем мире, где фальсификации могут повредить исходную информацию и привести к непредсказуемым последствиям, в том числе, к утечкам данных, фальсификации документов, политическим манипуляциям и т.д. Предоставление надежной и эффективной защиты от фальсификаций является ключевым шагом в обеспечении нашей безопасности и доверия в Интернете.

ЛИТЕРАТУРА

1. Alarsan F.I., Younes M. Best selection of generative adversarial networks hyper-parameters using genetic algorithm // SN Computer Science. — 2021. — Т. 2. — №. 4. — С. 283
2. Birunda S.S. et al. Fake Image Detection in Twitter using Flood Fill Algorithm and Deep Neural Networks // 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence). — IEEE, 2022. — С. 285–290.
3. Chai L. et al. What makes fake images detectable? understanding properties that generalize // Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16. — Springer International Publishing, 2020. — С. 103–120.
4. Cheng K. et al. An analysis of generative adversarial networks and variants for image synthesis on MNIST dataset //Multimedia Tools and Applications. — 2020. — Т. 79. — С. 13725–13752.
5. Galbally J., Marcel S. Face anti-spoofing based on general image quality assessment // 2014 22nd international conference on pattern recognition. — IEEE, 2014. — С. 1173–1178.
6. Qi P. et al. Exploiting multi-domain visual information for fake news detection // 2019 IEEE international conference on data mining (ICDM). — IEEE, 2019. — С. 518–527.
7. Tariq S. et al. Detecting both machine and human created fake face images in the wild // Proceedings of the 2nd international workshop on multimedia privacy and security. — 2018. — С. 81–87.
8. Tariq S. et al. Gan is a friend or foe? a framework to detect various fake face images // Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing. — 2019. — С. 1296–1303.
9. Wang X. et al. Gan-generated faces detection: A survey and new perspectives //arXiv preprint arXiv:2202.07145. — 2022.
10. Yang J. et al. Detecting fake images by identifying potential texture difference // Future Generation Computer Systems. — 2021. — Т. 125. — С. 127–135.

Журнал «Современная наука: актуальные проблемы теории и практики»