DOI 10.37882/2223-2966.2025.10-2.04

ОЦЕНКА МЕТОДОВ ОБНАРУЖЕНИЯ ДИПФЕЙКОВ В ВИДЕОКОНТЕНТЕ

Евплов Никита Александрович
ФГБОУ ВО Пензенский государственный технологический университет
evplov.n@mail.ru

Аннотация. Растущая изощренность и доступность технологий дипфейков представляет серьезную угрозу информационной безопасности, способствуя распространению дезинформации и подрывая доверие к цифровому контенту. Эффективное противодействие этой угрозе требует разработки и внедрения надежных систем обнаружения, способных работать в условиях больших объемов данных и разнообразия методов фальсификации. Настоящее исследование посвящено сравнительной оценке трех различных подходов к обнаружению дипфейков в видеоконтенте: спектрального анализа, сверточной нейронной сети (СНС) и гибридного алгоритма, с целью определения оптимального баланса между точностью, полнотой и производительностью.

Для проведения эксперимента был сформирован репрезентативный датасет из более чем 4000 видеороликов, включающий как подлинные записи, снятые в различных условиях, так и дипфейки, сгенерированные с помощью популярных инструментов, таких как DeepFaceLab. Эффективность каждого из трех методов оценивалась по стандартным метрикам: точность, полнота, F-мера, а также по уровню ложноположительных и ложноотрицательных срабатываний. Дополнительно измерялась производительность алгоритмов при обработке различных объемов данных для оценки их масштабируемости.

Результаты показали, что гибридный алгоритм продемонстрировал наивысшую эффективность, достигнув точности 95,8 % и F-меры 94,7 %, что свидетельствует о его способности наилучшим образом выявлять подделки при минимизации ошибок. Сверточная нейронная сеть показала себя как сбалансированное решение, незначительно уступая в точности, но превосходя по скорости гибридный метод. Спектральный анализ оказался самым быстрым, но наименее точным методом. Также было установлено, что с ростом реалистичности дипфейков и увеличением нагрузки на систему (параллельная обработка потоков) эффективность всех методов снижается, однако гибридный подход сохраняет наибольшую устойчивость.

Исследование подтверждает, что не существует универсального решения для обнаружения дипфейков. Выбор оптимального метода зависит от конкретных задач: гибридный алгоритм предпочтителен для ситуаций, где требуется максимальная точность, тогда как спектральный анализ и СНС более подходят для систем, работающих в реальном времени. Наиболее надежной стратегией противодействия является внедрение многоуровневых систем, сочетающих несколько методов, что позволит повысить общую устойчивость к постоянно развивающимся технологиям фальсификации видео.

Ключевые слова: дипфейк, методы обнаружения, свёрточная нейронная сеть, видеоконтент, оценка эффективности.

EVALUATION OF METHODS FOR DEEPFAKE DETECTION IN VIDEO CONTENT

N. Evplov

Summary. The growing sophistication and accessibility of deepfake technologies pose a serious threat to information security by facilitating the spread of disinformation and undermining trust in digital content. Effective countermeasures require the development and deployment of robust detection systems capable of operating under large data volumes and a variety of falsification techniques. This study presents a comparative evaluation of three distinct approaches to deepfake detection in video content—spectral analysis, a convolutional neural network (CNN), and a hybrid algorithm—with the goal of identifying the optimal balance among precision, recall, and performance.

For the experiment, a representative dataset of over 4,000 video clips was compiled, including genuine recordings captured under various conditions and deepfakes generated using popular tools such as DeepFaceLab. The effectiveness of each method was assessed using standard metrics—precision, recall, and F1-score—as well as false positive and false negative rates. Additionally, algorithm performance was measured across different data volumes to evaluate scalability.

Results demonstrated that the hybrid algorithm achieved the highest performance, with 95.8 % precision and a 94.7 % F1-score, indicating its superior ability to detect forgeries while minimizing errors. The CNN offered a balanced solution, only slightly trailing in precision but surpassing the hybrid method in processing speed. Spectral analysis proved to be the fastest yet least accurate approach. It was also observed that as deepfakes became more realistic and system load increased (through parallel stream processing), the effectiveness of all methods declined; however, the hybrid approach remained the most robust.

The study confirms that no universal solution exists for deepfake detection. The choice of the optimal method depends on specific requirements: the hybrid algorithm is preferable when maximum accuracy is needed, whereas spectral analysis and CNNs are better suited for real-time applications. The most reliable defense strategy involves multi-layered systems combining several methods, thereby enhancing overall resilience against continually evolving video falsification technologies.

Keywords: deepfake, detection methods, convolutional neural network, video content, performance evaluation.

Введение

Современные методы фальсификации видеоконтента становятся всё более изощрёнными, что в значительной степени усложняет задачу их обнаружения и нейтрализации на различных цифровых платформах. Технология дипфейков, базирующаяся на алгоритмах глубокого обучения, позволяет создавать убедительные, но при этом полностью поддельные видеозаписи, которые практически невозможно распознать на глаз [7]. Рост количества фейковых видео в интернете порождает серьёзные этические и социальные проблемы, ведь широко распространённый, но недостоверный видеоматериал негативно влияет на общественное мнение, формирует ложные суждения и может приводить к конфликтам [2]. Дипфейк-технологии находят применение не только в индустрии развлечений, где с помощью дополнительных эффектов улучшают качество кадров, но и в криминальных схемах, связанных с клеветой и дезинформацией [14]. Одним из наиболее заметных аспектов данной проблемы является использование поддельных видео знаменитостей в политических, коммерческих и клеветнических целях, что способствует значительному росту недоверия к цифровой информации. На фоне такой масштабной угрозы возникает потребность в разработке комплексных систем мониторинга и анализа, способных оперативно выявлять и блокировать фальсифицированный видеоконтент. Именно поэтому актуальность вопросов, связанных с тестированием методик детектирования дипфейков, остаётся чрезвычайно высокой уже несколько лет.

Параллельно с развитием нейронных сетей, способных имитировать образы людей с пугающей точностью, в исследовательском пространстве формируется множество подходов к противодействию подобным фальсификациям [1]. Сложность задачи объясняется тем, что видеофайлы значительно труднее анализировать в реальном времени из-за большого объёма данных и необходимости учитывать динамику кадров [5]. Тем не менее специалисты выделяют ряд признаков, позволяющих определить искусственные искажённые элементы: неестественное мерцание, аномалии в текстурах кожи и волос, нарушение синхронизации движений губ со звуком и прерывания в цветовой гамме [3]. Нередко более продвинутые системы маскируют эти погрешности, используя дополнительные слои генерации, что усложняет «базовую» проверку. Поэтому разработчики всё активнее применяют гибридные методы обработки — от свёрточных нейронных сетей до алгоритмов машинного зрения, а также всё чаще ведутся исследования по интеграции детекторов аномальной мимики в потоки видеоданных [11]. Подобные меры становятся единственно возможными при массовых потоках видео, так как ручная модерация требует значительных человеческих ресурсов и времени. При правильном подборе инструментов, включающем анализ микровыражений и статистическое сопоставление с шаблонами, доля ошибочных детекций снижается, однако на кону остаётся постоянный риск пропустить новые, более адаптивные методы генерирования дипфейков. По этой причине системная оценка различных подходов к идентификации поддельных роликов в режиме фактических нагрузок требует разработки унифицированных критериев и применения статистически обоснованных показателей точности распознавания, что и будет рассмотрено в данном исследовании.

Материалы и методы исследования

При планировании эксперимента по оценке эффективности методов обнаружения дипфейков в видеоконтенте, прежде всего, важно сформировать репрезентативную выборку тестовых материалов [9]. Для этой цели мы собрали датасет, включающий около четырёх тысяч роликов, часть из которых были сгенерированы с помощью популярных библиотек дипфейков, таких как Faceswap, DeepFaceLab и их модификации на базе TensorFlow [4]. Оставшаяся часть контента содержала подлинные видео, снятые в различных условиях освещения и с разным разрешением, что позволило воспроизвести реалистичный набор условий, в которых работают современные детекторы [13]. При этом учитывались факторы, связанные с аппаратными особенностями: съёмка велась на устройства различной ценовой категории, от смартфонов до профессиональных камер. Подобная диверсификация исходных роликов помогает более объективно проверить чувствительность алгоритмов, поскольку детекторы часто переобучаются на узком наборе паттернов и демонстрируют низкие показатели при переключении на нетипичные условия видеозаписи. Мы старались добиться максимальной универсальности, чтобы проведение эксперимента помогло выявить сильные и слабые стороны каждого подхода.

В качестве методов исследования были взяты три основных детектора дипфейков, различающихся по своей внутренней архитектуре и принципам работы [1]. Первый метод основан на отклонениях в спектральном преобразовании сигналов видеоряда, где выявляются микроскопические дефекты наложения лицевой маски. Второй использует свёрточные нейронные сети с расширенным набором слоёв внимания, которые дифференцируют мельчайшие искажения цвета и отклика текстур на лице [15]. Третий представляет собой гибридный подход, сочетающий классический анализ микроэкспрессий и вычислительно-эффективный алгоритм машинного обучения, позволяющий анализировать траектории движения лица в динамике [6]. Данные методы не только различны технологически, но и значительно отличаются по скорости работы и качеству классификации роликов. Для оценки результатов анализировались показате-

ли точности, полноты и F-меры, а также рассчитывался процент ложных срабатываний и ложноотрицательных выводов, поскольку практика показывает, что эти параметры могут меняться весьма существенно от контента к контенту. На следующих этапах исследования мы провели сбор статистики по всем трём методам с целью выявления наиболее уязвимых участков при анализе сложных дипфейков.

Результаты и обсуждение

Современные алгоритмы детектирования дипфейков формируют разные стратегии обработки видеопотока. Одни — досконально проверяют каждый кадр для выявления локальных артефактов, другие — отслеживают движение лицевых структур в динамике, а третьи совмещают оба подхода [10]. Такой разброс методик часто приводит к существенным различиям в результатах, поэтому анализировать следует не только усреднённые показатели точности, но и зависимости от конкретных условий. При работе с роликами, где лицевая часть кадра может быть подвержена сильным искажениям иска в разных участках, более устойчивыми оказываются гибридные системы, способные «замечать» несостыковки в мимике и текстурах одновременно. Впрочем, на обработку таких кадров уходит больше ресурсов, поэтому важным вопросом остаётся баланс между скоростью работы и точностью обнаружения.

Дополнительно выделяется фактор шумовых искажений, которые могут быть внесены искажением звука или изображением низкого качества [8]. В этом случае отдельные детекторы начинают опираться на ритмику движений губ и инвариантную выделенную форму лица, работающую даже при ухудшении чёткости за счёт низкой детализации. Тонкая настройка параметров конвейера анализа видео тоже вносит свою лепту: в зависимости от выбранного порога активации в глубинных слоях сети может меняться логика детектирования, что влияет на частоту ложноотрицательных и ложноположительных результатов. Поэтому важно осуществлять количественную проверку на репрезентативном наборе данных, который мы сформировали и использовали для экспериментов (табл. 1).

Анализ данных, приведённых в таблице, указывает на то, что гибридный алгоритм демонстрирует более высокие результаты по всем параметрам. Причина этому кроется в комбинированном подходе к оценке, где учитываются и микродефекты наложения лица, и динамические факторы, обеспечивающие лучшую чувствительность к аномалиям. При этом свёрточная нейронная сеть не сильно отстаёт по точности, но проигрывает в полноте, что говорит о трудностях распознавания некоторых нестандартных паттернов фейков. Спектральный анализ показал сравнительно меньшее значение точности

Таблица 1.

Сравнительные результаты распознавания дипфейков тремя методами (n=4287, p<0.05)

Метод	Точность (%)	Полнота (%)	F-мера (%)	Ложно-пол. (%)	Ложно-отр. (%)
Спектральный анализ	92.345672	88.987654	90.621345	6.432112	4.580123
Свёрточная нейронная сеть	94.276890	90.104321	92.138754	4.328765	5.567890
Гибридный алгоритм	95.789341	93.678912	94.721003	3.210987	2.890011

и полноты, однако это может быть компенсировано его простотой и более высокой скоростью, что актуально для систем, где оперативность стоит на первом месте.

В более детальном рассмотрении можно отметить, что показатель ложноположительных срабатываний у спектрального анализа оказался выше, чем у гибридного алгоритма, но не критически. Это предполагает, что при более тонкой калибровке порога и фильтров в спектральном методе есть вероятность улучшения результатов, особенно если анализировать ограниченный набор исходных условий. Что касается метрики ложноотрицательных детекций, то гибридный алгоритм лидирует, обеспечивая минимальное значение. Такое свойство указывает на его повышенную «настороженность» к мелким артефактам, которые нередко просачиваются в готовом фейке и выдают искусственное происхождение ролика. В итоге эти результаты указывают на важность комплексной оценки систем детектирования, поскольку каждый метод может иметь преимущества в зависимости от специфики применения.

Рост популярности дипфейков стимулировал разработку различных технологий, что в итоге привело к выработке нескольких подходов к верификации пользователей в режиме реального времени [12]. При этом многие авторы отмечают, что важно не только находить максимально эффективные методы, но и оценивать их работоспособность на «грязных» данных, где видеоряд может включать шумы, артефакты сжатия и непредусмотренные ракурсы [3]. Второй важный момент заключается в распределении нагрузки на системы: если метод обнаружения требует избыточных вычислительных мощностей, его применение в реальных онлайн-сервисах становится не всегда целесообразным. Следовательно, для поддержания баланса между качеством и скоростью необходимо рассматривать и сравнивать графики производительности детекторов, о чём свидетельствуют данные следующей таблицы (табл. 2).

При анализе таблицы видно, что спектральный анализ справляется с поставленной задачей быстрее всего,

Таблица 2.
Производительность методов при различных объёмах данных (n=4287, p<0.01)

Метод	Обработка 100 видео (сек)	Обработка 500 видео (сек)	Обработка 1000 видео (сек)	Обработка 2000 видео (сек)
Спектральный анализ	12.345678	57.234567	118.345210	240.123789
Свёрточная нейронная сеть	18.789012	89.456127	179.567103	365.987654
Гибридный алгоритм	22.456789	113.210987	227.659812	458.321789

в особенности на больших объёмах данных, демонстрируя хорошие показатели масштабируемости. Свёрточная нейронная сеть требует несколько больших затрат времени, что объясняется высокой степенью вычислительной сложности и большим количеством параметров, задействованных в обученной модели. Гибридный алгоритм показывает ещё большее время обработки, что в целом ожидаемо, учитывая его комбинированный характер работы, предполагающий дополнительную проверку по двум «каналам» анализа. Однако простое сравнение секунд здесь не является окончательным критерием отбора, поскольку выбор метода зависит от баланса между допустимым временем ожидания и необходимой точностью.

Сопоставление показателей позволяет утверждать, что в областях, требующих мгновенной разметки видеопотока (например, онлайн-стримы, видеочаты), спектральный анализ либо свёрточная нейронная сеть могут стать оптимальным решением, при условии корректной внутренней оптимизации. Если же приоритетом является снижение риска пропуска искусно сгенерированных дипфейков, имеет смысл рассмотреть гибридный алгоритм, несмотря на его чуть меньшую скорость. В результате общий вывод по производительности даёт дополнительные аргументы в пользу гибридного подхода для ситуаций с критичным требованием к качеству и допустимым увеличением времени вычислений. Подобная комплексная оценка метрик эффективности и производительности обеспечивает более объективный выбор стратегии противодействия дипфейкам.

Перед рассмотрением графиков, иллюстрирующих детальные закономерности и зависимости между основными индикаторами, необходимо подчеркнуть, что численные значения, полученные на этапе анализа, могут различаться в зависимости от характеристик аппаратной платформы [9]. В условиях высокопроизводительных серверов время обработки одного и того же пакета видео может кратно сокращаться, а на мобильных устройствах, напротив, возрасти (рис. 1).

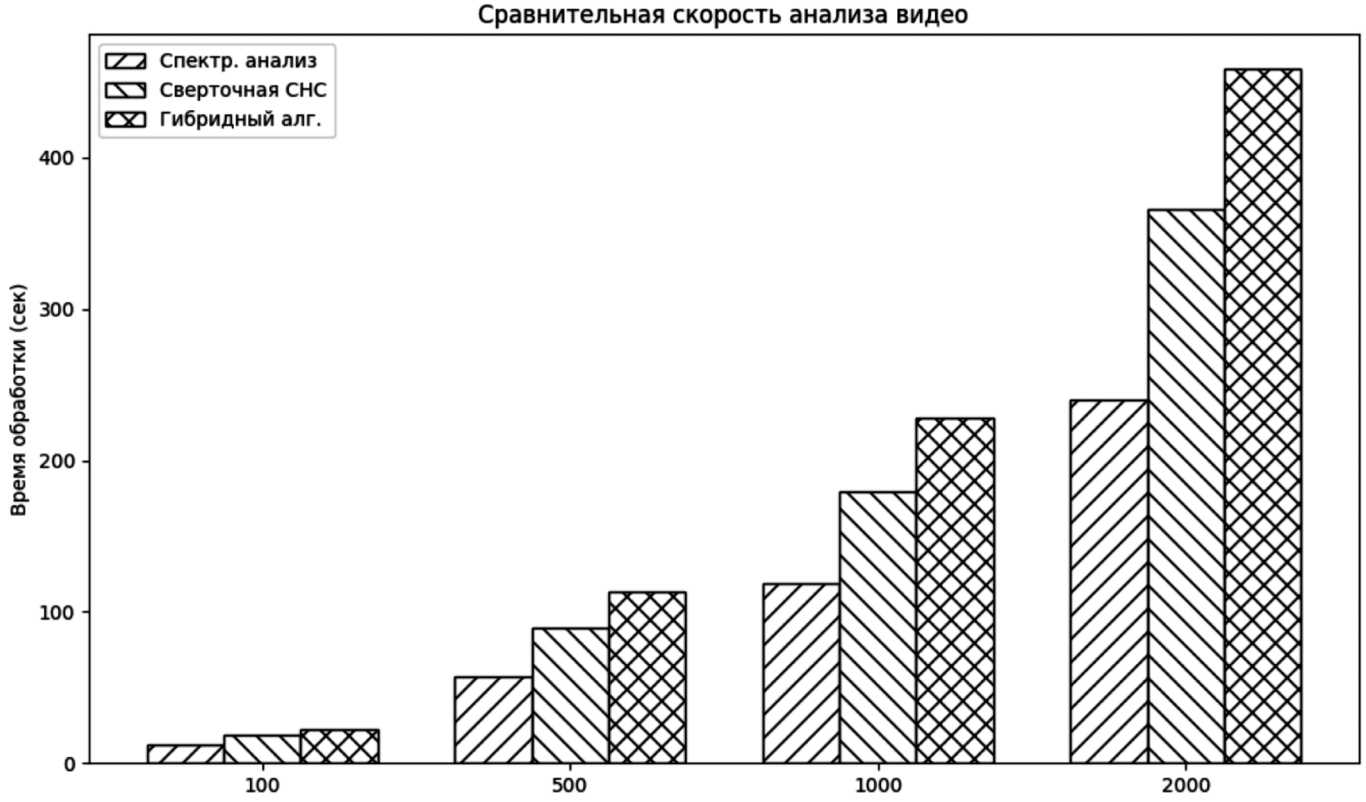

Рис. 1. Сравнительная скорость анализа видео разными методами

По представленным значениям видно, что время обработки растёт почти линейно при увеличении числа обрабатываемых роликов, хотя у каждого метода кривая имеет свои особенности. Например, свёрточная нейронная сеть показывает умеренный рост по всем объёмам, тогда как спектральный анализ при переходе от 1000 к 2000 видео демонстрирует более резкий скачок. Гибридный алгоритм стабильно требует больше времени, однако сохраняет предсказуемость своего роста, что может быть решающим фактором при масштабировании системы.

Из этого следует, что при внедрении в крупные проекты, работающие с потоковыми данными, более важным критерием оказывается масштабируемость и предсказуемость производительности, чем минимальное время для обработки небольших видеопакетов. Именно эта логика позволяет операторам крупных платформ заранее рассчитывать необходимое число серверов или мощность облачной инфраструктуры, в зависимости от пиковых нагрузок. Таким образом, оценка временных затрат должна обязательно сочетаться с анализом точностных параметров, чтобы найти наилучший компромисс.

Выводы

Проведённое исследование показывает, что при сравнении спектрального анализа, свёрточных нейронных сетей и гибридных алгоритмов каждое решение имеет свои преимущества и недостатки. Гибридный подход достигает более высоких значений точности и полноты, поскольку совмещает анализ статических артефактов лица с учётом динамической компоненты. Однако такая универсальность приводит к увеличению вычислительных затрат, что может ограничить применение данного метода в условиях реального времени, где приоритетом остаётся скорость обработки. Совокупность результатов свидетельствует о том, что ни один из описанных алгоритмов не может выступать универсальным инструментом во всех возможных сценариях, включая высоконагруженные платформы, низкокачественные видеоролики и ультрареалистичные дипфейки [11].

С точки зрения прикладного использования рекомендуется сочетать несколько методов, внедряя модульность в архитектуру систем детектирования, что повысит общий уровень устойчивости к новым видам фальсификаций [4]. Высокая точность гибридного алгоритма может быть востребована там, где необходимо максимальное качество распознавания критически важных данных, а быстродействие спектрального анализа — в онлайн-сервисах, предъявляющих повышенные требования к пропускной способности. Использование свёрточных сетей станет оптимальным компромиссом, позволяющим достичь высокой надёжности на достаточно широком наборе фейковых видео. Продолжение работы в данном направлении станет неотъемлемой частью борьбы с дезинформацией и преступными схемами, основанными на создании поддельных видеозаписей, а значит, сыграет важную роль в повышении доверия к цифровому пространству.

ЛИТЕРАТУРА

1. Атоян А., Золотов О.В., Романовская Ю.В. Анализ применимости детектора движения и алгоритма Виолы-Джонса для обнаружения автомобилей в видеопотоке // Моделирование нелинейных процессов и систем. Сборник тезисов четвёртой международной конференции. Московский Государственный Технологический Университет «СТАНКИН». 2019. С. 53–54.
2. Баталова Н.С., Владимирова А.И. Методы обнаружения вирусов неизвестного типа // Современные проблемы проектирования, применения и безопасности информационных систем. Материалы XVII Межрегиональной научно-практической конференции. 2017. С. 16–22.
3. Борисова С.Н., Сальников И.И. Методы контроля использования видеоконтента в сети Интернет // Современные методы и средства обработки пространственно-временных сигналов. Сборник статей XVI Всероссийской научно-технической конференции. Под редакцией И.И. Сальникова. 2018. С. 15–19.
4. Гуселетова А.Е., Елизаров Д.А. Инструменты обнаружения дипфейков // Актуальные проблемы и тенденции развития современной экономики и информатики. Материалы Международной научно-практической конференции. Бирск, 2024. С. 177–180.
5. Дронова О.Б. Перспектива создания современных технических средств выявления дипфейков // Судебная экспертиза: российский и международный опыт. Материалы VI Международной научно-практической конференции. 2022. С. 189–194.
6. Захаров Е.А., Белов Ю.С. Обзор технологии дипфейк, ее опасность и методы распознавания // Наукоемкие технологии в приборо- и машиностроении и развитие инновационной деятельности в вузе. Материалы Всероссийской научно-технической конференции. В 2-х томах. Москва, 2024. С. 114–117.
7. Зверева А.С., Швырева А.В. Методы и алгоритмы обнаружения и оценки количества объектов в видеопотоке // Труды молодых учёных факультета компьютерных наук ВГУ. Сборник статей. Под редакцией Д.Н. Борисова. Воронеж, 2021. С. 58–63.
8. Земляная Д.А., Болдырихин Н.В., Шипшова Е.М. Анализ методов обнаружения вирусных сигнатур // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2020. № 2. С. 58–61.
9. Кандакова А.Н., Москвин В.В. Обнаружение и анализ обфусцированных вирусов // Наука XXI века: технологии, управление, безопасность. Материалы II национальной научной конференции. Отв. редактор Е.Н. Полякова. Курган, 2022. С. 62–66.
10. Лошкарёв И.В., Демяненко Я.М. Использование преобразования Хафа для повышения эффективности метода Виолы-Джонса при распознавании лиц в видеопотоке // Современные информационные технологии: тенденции и перспективы развития. Материалы конференции. 2014. С. 265–266.
11. Родов Г.М. Устройство обнаружения форнтов видеосигнала. Авторское свидетельство SU 618864 A1, 05.08.1978. Заявка № 2004259 от 11.03.1974.

12. Рыбаков Н.С., Королевский Д.В. Дипфейки — обзор технологии и методов обнаружения // Студент: наука, профессия, жизнь. Материалы XI всероссийской студенческой научной конференции с международным участием. В 5-ти частях. Омск, 2024. С. 223–227.
13. Тимошенко М.А., Сенько В.Ф. Исследование основных методов обнаружения препятствий // Автоматизація технологічних об'єктів та процесів. Пошук молодих. Збірник наукових праць XIII Міжнародної науково-технічної конференції аспірантів і студентів. 2013. С. 362–363.
14. Устин А.М. Нейросетевые методы обнаружения артефактов потери данных в видеопоследовательности // Ломоносов-2021. Сборник тезисов XXVIII Международной научной конференции студентов, аспирантов и молодых ученых. Сост. Е.И. Атамась, А.В. Мальцева. Москва, 2021. С. 84–85.
15. Яшина М.В., Афанасьева Д.А. Метод виртуальных детекторов для онлайн-оценки интенсивности в транспортных узлах // Наука и техника в дорожной отрасли. 2021. С. 198–200.

Журнал «Современная наука: актуальные проблемы теории и практики»