

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра компьютерных технологий и систем

ШЕСТАК Татьяна Геннадьевна

МЕТОДЫ, АЛГОРИТМЫ И ТЕХНОЛОГИИ
ОБНАРУЖЕНИЯ ИЗМЕНЕНИЙ СЦЕН И ОБЪЕКТОВ
ВИДЕОПОСЛЕДОВАТЕЛЬНОСТИ

Дипломная работа

Научный руководитель:
старший преподаватель Шолтанюк С.В.

Допущена к защите

« ____ » _____ 20 ____ г.

Заведующий кафедрой компьютерных
технологий и систем, доктор педагогических
наук, кандидат физико-математических наук,
профессор В.В. Казачёнок

Минск, 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	6
ГЛАВА 1 Введение в теорию обработки изображений	7
1.1 Общие понятия обработки изображения	7
1.2 Классификация задач компьютерного зрения	7
1.3 Понятие изменений на кадрах видеопоследовательности	10
ГЛАВА 2 Обзор методов, алгоритмов и технологий для обнаружения изменений на кадрах видеопоследовательности	13
2.1 Методы, основанные на алгебраических вычислениях.....	13
2.2 Методы глубокого обучения	17
ГЛАВА 3 Введение в теорию обработки видеопоследовательностей	24
3.1 Общие понятия о видео.....	24
3.2 Режимы обработки видео и методы, применимые для них: офлайн и онлайн подходы.....	26
3.3 Форматы видеопоследовательностей: подходы к обработке сжатых и несжатых данных	33
ГЛАВА 4 Задача обнаружения изменений сцен и объектов видеопоследовательности	37
4.1 Постановка задачи обнаружения изменений на видеопоследовательностях.....	37
4.2 Выбор методов для решения задачи разделения на кадры	39
4.3 Выбор методов для решения задачи обнаружения ключевых кадров	41
4.4 Выбор методов для решения задачи обнаружения изменений на кадрах	42
ГЛАВА 5 Сравнение результатов выбранных методов для решения поставленной задачи	44
5.1 Результаты решения задачи разделения на кадры	44
5.2 Результаты решения задачи обнаружения кадров со сменой сцен	47
5.3 Результаты решения задачи обнаружения изменений на кадрах	51
ЗАКЛЮЧЕНИЕ	53
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	54

РЕФЕРАТ

Дипломная работа, 56 страниц, 20 рисунков, 2 таблицы, 26 источников

Ключевые слова: ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЬ, СМЕНА СЦЕН, ОБНАРУЖЕНИЕ ИЗМЕНЕНИЙ, ВИДЕОКАДР, СРАВНЕНИЕ ГИСТОГРАММ, РАЗНОСТЬ КАДРОВ.

Объект исследования: методы, алгоритмы и технологии обнаружения изменений сцен и объектов видеопоследовательности.

Предмет исследования: эффективность и особенности применения различных методов, алгоритмов и технологий для обнаружения изменений в сценах и объектах на видеопоследовательностях.

Цель исследования: определение наиболее эффективных методов деления видеопоследовательности на кадры, обнаружения смен сцен и обнаружения изменений на кадрах видеопоследовательностей.

Методы исследования: сравнительный анализ некоторых подходов разбиения видеопоследовательностей на кадры и обнаружения изменений на них.

Полученные результаты и их новизна: были рассмотрены и предложены способы по делению видеопоследовательности на кадры, а также по определению изменений между ними. Проведённое исследование показало, что методы, основанные на глубоком обучении, являются наиболее гибким инструментом для обнаружения различных по характеру изменений в сценах и объектах. В большей степени были выделены алгебраические методы на основе простоты их интерпретации и реализации.

Область возможного практического применения: полученные результаты позволяют рекомендовать рассмотренные методы обнаружения изменений сцен и объектов видеопоследовательности для использования в системах видеонаблюдения, в автоматизированных системах мониторинга пожарной безопасности и других, направленных на раннее обнаружение изменений в местах, за которыми ведется постоянное наблюдение в целях предупреждения чрезвычайных и несовместимых с жизнью человека ситуациях, для которых необходимо экстренное реагирование.

РЭФЕРАТ

Дыпломная праца, 56 старонак, 20 малюнкаў, 2 табліцы, 26 крыніц

Ключавыя словы: ВІДЭАПАСЛЯДОЎНАСЦЬ, ЗМЕНА СЦЭН, ВЫЯЎЛЕННЕ ЗМЯНЕННЯЎ, ВІДЭАКАДР, ПАРАЎНАННЕ ГІСТАГРАМ, РОЗНАСЦЬ КАДРАЎ.

Аб'ект даследавання: метады, алгарытмы і тэхналогіі выяўлення змяненняў сцэн і аб'ектаў відэапаслядоўнасці.

Прадмет даследавання: эфектыўнасць і асаблівасці прымянення розных метадаў, алгарытмаў і тэхналогій для выяўлення змяненняў у сцэнах і аб'ектах на відэапаслядоўнасці.

Мэта даследавання: вызначэнне найбольш эфектыўных метадаў дзялення відэапаслядоўнасці на кадры, выяўлення змен сцэн і выяўлення змяненняў на кадрах відэапаслядоўнасці.

Метады даследавання: параўнальны аналіз некаторых падыходаў разбіцця видеопоследовательностей на кадры і выяўлення змяненняў на іх.

Атрыманыя вынікі і іх навізна: былі разгледжаны і прапанаваны спосабы па дзяленню відэапаслядоўнасці на кадры, а таксама па вызначэнні змяненняў паміж імі. Праведзенае даследаванне паказала, што метады, заснаваныя на глыбокім навучанні, з'яўляюцца найбольш гнуткімі інструментамі для выяўлення розных па характары змяненняў у сцэнах і аб'ектах. У большай ступені былі вылучаныя алгебраічныя метады на аснове прастаты іх інтэрпрэтацыі і рэалізацыі.

Вобласць магчымага практычнага прымянення: атрыманыя вынікі дазваляюць рэкамендаваць разгледжаныя метады выяўлення змяненняў сцэн і аб'ектаў відэапаслядоўнасці для выкарыстання ў сістэмах відэаназірання, у аўтаматызаваных сістэмах маніторынгу пажарнай бяспекі і іншых, накіраваных на ранняе выяўленне змяненняў у месцах, за якімі вядзецца пастаяннае назіранне ў мэтах папярэджання надзвычайных і несумяшчальных з жыццём чалавека сітуацыях, для якіх неабходна экстранае рэагаванне.

SUMMARY

Diploma work, 56 pages, 20 figures, 2 tables, 26 sources

Keywords: VIDEO SEQUENCE, SCENE CHANGE, CHANGE DETECTION, VIDEO FRAME, HISTOGRAM COMPARISON, FRAME DIFFERENCE.

The object of the research: methods, algorithms, and technologies for detecting changes in scenes and objects in a video sequence.

The subject of the research: the effectiveness and application features of various methods, algorithms and technologies for detecting changes in scenes and objects in video sequences.

The purpose of the study: to determine the most effective methods for dividing a video sequence into frames, detecting scene shifts, and detecting changes in video sequence frames.

Methods of research: a comparative analysis of some approaches to splitting video sequences into frames and detecting changes to them.

The results of the work and their novelty: methods for dividing a video sequence into frames, as well as for determining changes between them, are considered and proposed. The conducted research shows that methods based on deep learning are the most flexible tool for detecting changes in scenes and objects of various types. Algebraic methods are singled out to a greater extent based on the simplicity of their interpretation and implementation.

Recommendations on the usage: the results obtained allow us to recommend the considered methods for detecting changes in scenes and objects of the video sequence for use in video surveillance systems, automated fire safety monitoring systems, and others aimed at early detection of changes in places that are constantly monitored in order to prevent emergencies and situations incompatible with human life that require emergency the response.

ВВЕДЕНИЕ

В области компьютерного зрения обнаружение изменений сцен и объектов в видеопоследовательностях представляет собой актуальную задачу, имеющую многочисленные приложения в различных сферах, таких как безопасность, медицина, автономные системы, видеонаблюдение и многие другие. Её важность заключается в возможности последующего моделирования сложных процессов и инфраструктур. Системы со встроенным обнаружением изменений в окружающей среде способны своевременно и быстрее человека реагировать на потенциально значимые или опасные ситуации.

Для решения данной задачи применяются различные классические методы и алгоритмы: алгебраические, статистические и основанные на преобразованиях изображений. Кроме того, активно развиваются и на данный момент считаются самыми продвинутыми алгоритмы глубокого обучения, в которых используются глубокие нейронные сети. Нейронные сети стали одним из наиболее популярных инструментов в искусственном интеллекте, позволяя решать широкий спектр задач, включая обработку изображений, распознавание объектов и анализ видео.

При работе с видеoinформацией необходимо учитывать их характеристики для достижения наиболее качественных результатов. Видеопоследовательности могут находиться как в сжатом, так и в несжатом форме, из-за чего необходимо подбирать соответствующие методы, применимые для конкретной задачи.

Цель данной работы состоит в анализе методов, алгоритмов и технологий, используемых для обнаружения изменений в видеопоследовательностях, а также в оценке их эффективности и применения в реальных ситуациях.

Задачи работы:

- Провести обзор и проанализировать теоретические сведения о теме обнаружения изменений в видеопоследовательностях
- Исследовать и классифицировать различные алгоритмы и методы, применяемые в данной области
- На практике проанализировать существующие технологии, а также оценить их преимущества и недостатки.
- Выявить критерии, от которых зависит точность и применимость методов и алгоритмов обнаружения изменений, и определить их оптимальные значения для использования их в реальных условиях.

ГЛАВА 1 Введение в теорию обработки изображений

1.1 Общие понятия обработки изображения

Обработка изображений является фундаментальной областью цифровых технологий, которая включает в себя методы, алгоритмы и технологии, направленные на работу напрямую с изображениями. Целями применения данной области могут быть, например, изменение изображений, избавление от помех и искажений, извлечение свойств и данных заложенных в них, подготовка изображений к выполнению других задач. Смежной областью с обработкой изображений является компьютерное зрение.

Компьютерное зрение представляет собой область искусственного интеллекта, который позволяет автоматически извлекать, анализировать и обрабатывать информацию из визуальных данных, что значительно облегчает весь процесс. Данные, с которыми приходится работать, могут быть представлены в различных формах, таких как обычные двумерные изображения, трехмерные изображения, видеопотоки и видеопоследовательности. Технологии компьютерного зрения позволяют обнаруживать, идентифицировать и отслеживать объекты в режиме реального времени.

Обработка изображений считается одним из методов, который используется в компьютерном зрении, наряду с другими методами машинного обучения. В настоящее время при решении задач компьютерного зрения все чаще используют современные алгоритмы глубокого обучения, а именно нейронные сети, так как они обрабатывают и сохраняют информацию о каждом проанализированном изображении, что позволяет при накоплении данных показывать более точные результаты.

1.2 Классификация задач компьютерного зрения

Выделяют несколько основных задач, которые решаются в компьютерном зрении:

1. Классификация изображения

Задача бинарной классификации заключается в том, чтобы ответить на вопрос «присутствует ли объект на изображении?» В данном случае всего два возможных ответа: «да» или «нет». Многоклассовая классификация распределяет объекты на изображении по категориям, а исходя из того, какие нам изначально известны метки классов, мы приписываем классы объектам на изображении. Так можно отделить человека от животного, кота от собаки.

2. Сегментация изображений

Сегментация изображений имеет много общего с задачей распознавания объектов, однако вместо того, чтобы обрамлять обнаруженные объекты рамками, ее цель состоит в том, чтобы выделить все пиксели, формирующие объекты. То есть в данной случае изображение разбивается на сегменты, которые полностью его покрывают. Выделяют три вида сегментации: семантическая сегментация, сегментация экземпляров и паноптическая сегментация. Семантическая сегментация разделяет объекты на классы, присваивает метки и выделяет их пиксели. В то время как сегментация экземпляров берет на себя еще и задачу отделить каждый объект своего класса, присвоив каждому уникальный номер что позволяет отличать друг от друга каждый объект класса. Паноптическая сегментация объединяет собой семантическую сегментацию и сегментацию по экземплярам, чтобы полностью классифицировать все пиксели изображения. Для решения данной задачи используют как классические методы, которые напрямую взаимодействуют с пикселями, так и глубокие нейронные сети. Популярным методом решения задачи сегментации является полносвязная сверточная нейронная сеть (FRN) и другие более сложные архитектуры как UNet, DenseNet, PSPNet, ResNet. Примеры отработки алгоритмов сегментации приведены на рисунке 1.1.

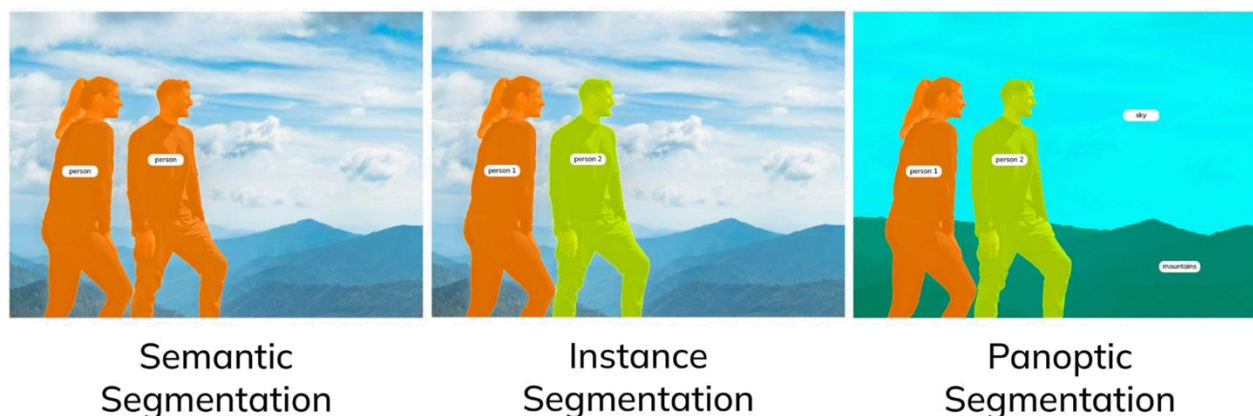


Рисунок 1.1 – Результат решения задачи сегментации

3. Локализация объекта

Локализация или, другими словами, оценка положения представляет собой задачу определения положения объекта относительно камеры. Необходимо найти объект, выделить его каркас и определить его расположение на изображении. Выделяют два вида локализации: одиночную и множественную, в зависимости от количества рассматриваемых объектов. В

случае множественной оценки положения необходимо учитывать взаимное наложение объектов.

Построенный каркас может быть использован в дальнейшем для прогнозирования направления движения объекта.

Процесс решения задачи локализации начинается с удаления фона, далее для каждого объекта выделяются области, которые обозначают линиями, которые соединяются в единую структуру «скелета». Для подзадачи выделения области-«признака» используются сверточные сети, так как в процессе своей работы они переходят от простых особенностей изображения к более абстрактным и неочевидным. На рисунке 1.2 наглядно показано, как выглядит каркас при оценке положения на примере трех изображений человека.



Рисунок 1.2 – Результат решения задачи локализации

4. Обнаружение (детекция) объектов

Цель задачи обнаружения объектов в поиске и идентификации конкретных объектов на изображении. Данный процесс состоит из двух составных частей: локализация объекта и классификация объектов, которые в совокупности и составляют процесс обнаружения объекта. К видео также возможно применять данную задачу, однако в большинстве случаев для этого видео представляют как серию кадров и работают с каждым кадром отдельно. В сравнении с задачей сегментации изображения, обнаружение объектов на изображении является менее трудозатратой задачей.

1.3 Понятие изменений на кадрах видеопоследовательности

Видеопоток представляет собой последовательность изображений, захваченных одной камерой с минимальными временными интервалами между кадрами, при этом обработка видеопотока осуществляется в реальном времени. Видеокадром называют отдельное изображение из видеопотока или видеопоследовательности, которое состоит из большого количества объектов. Кадры в видеопотоке появляются с частотой примерно 30 раз в секунду, что представляет собой малое ограниченное для обработки каждого изображения. В связи с этим, алгоритмы, работающие в реальном времени, часто отличаются своей простотой. Ключевой характеристикой видеопотока является временная согласованность между кадрами, так как за время приблизительно равное $1/30$ секунды происходят незначительные изменения.

Видеопоследовательность также представляет собой последовательную группу из статичных, неподвижных изображений, которые создают представление о движении во времени. Вместе все кадры формируют динамическую картину: каждые два соседних кадра в последовательности имеют незначительные отличия из-за малого различия во времени, позиции камеры и объектов во время съемки. Необходимо отметить, что в отличие от видеопотока, видеопоследовательность имеет ограниченный размер и позволяет выполнять обработку по всей длине последовательности.

Другое важное понятие – динамическая сцена. Динамическая сцена определяется как пространство, фон на котором происходит появление, изменение формы, перемещение, исчезновение объектов. Для исследования динамической сцены необходим анализ последовательности изображений.

Изменения на кадрах видеопоследовательности имеют ключевое значение для создания общей картины о движении объектов и для анализа событий.

Устойчивыми изменениями на кадрах и снимках считаются те, которые перевели динамическую сцену в новое состояние, причем новое состояние длится не менее заданного промежутка времени. Важно отметить, что время, затраченное на перевод динамической системы в новое состояние, должно не превышать данного промежутка времени, иначе изменение не считается устойчивым.

Изменения, наблюдаемые на кадрах и изображениях, могут быть обусловлены рядом причин:

- 1) Динамические, движущиеся объекты, находящиеся в поле зрения камеры будут изменять положение на изображениях. Примером является передвижение человека или автомобиля.

- 2) Движение камеры, которая делать запись видео. Видеооператор может изменять положение камеры.
- 3) Изменение окружающей среды. Изменения погоды или освещения могут привести к смене сцен. Так как временной промежуток между двумя сделанными экземплярами изображений может быть различным, может измениться освещение, которое повлияет на качество изображения ввиду затемнения или освещения объекта. Существует проблема определения реальных форм объекта из-за присутствия тени на кадрах.
- 4) Технические особенности снимаемых объектов. Они могут приносить значительные изменения, которые важны для исследований или на производстве. К примеру, извержение вулкана обусловлено не его движением или движением камеры, снимающей его, а появлением из вулкана лавы. Другим примером может быть возгорание вычислительного прибора, если его оставили без надлежащего контроля.

Системы обнаружения изменений сцен и объектов видеопоследовательности решают проблему поиска визуальных паттернов на основе предыдущих кадров. Приложения, в которых необходимо найти изменения разнообразны и зависят, в основном, от контекста. Например, выявление изменений на видео необходимо в таких случаях как: безопасность на дороге для выявления нарушения правил дорожного движения, таких как переход дороги пешеходом в неположенном месте; мониторинг общественных мест для обнаружения подозрительной активности или оставленных вещей в метро или в местах скопления людей; пожарной безопасности для обнаружения потенциальных источников пожара или открытого огня. Таким образом, в некоторых случаях критически важно иметь системы видеонаблюдения и обнаружения изменений на основе видео, так как они могут помочь предотвратить серьезные последствия, которые можно избежать при своевременном обнаружении и устранении. Также очень важное применение технологий обнаружения изменений заключается в использовании его при дистанционном зондировании.

Дистанционное зондирование – это получение информации об объекте на расстоянии от него. Часто этот термин используется более конкретно для обозначения наблюдения за поверхностью Земли сверху со спутника или самолета.

В настоящее время доступное большое количество изображений, полученных посредством дистанционного зондирования, так как информация в этих изображениях представляет интерес для ряда областей, таких как

картография, наблюдение за климатическими изменениями, мониторинг стихийных бедствий и городское планирование.

Оптические изображения, которые получают при дистанционном зондировании, разделяют на три группы: гиперспектральные, многоспектральные и панхроматические. Многоспектральные изображения имеют несколько (обычно 3–15) спектральных диапазонов, в то время как гиперспектральные изображения выбирают спектр гораздо более точно с десятками или сотнями узких диапазонов. Панхроматические изображения собирают весь видимый свет в один единственный диапазон. Многоспектральные датчики обычно работают в визуальном, ближнем инфракрасном и коротковолновом инфракрасном диапазоне спектра. Красные, зеленые и синие (RGB) изображения являются часто используемым и хорошо известным подмножеством многоспектральных изображений.

Гиперспектральные изображения имеют самое высокое спектральное разрешение, в то время как панхроматические имеют самое низкое. Гиперспектральные изображения с их высоким спектральным разрешением могут затем в полной мере использовать информацию, предоставленную тонкой выборкой спектральной сигнатуры объектов. Большое количество полос делает гиперспектральную визуализацию хорошо подходящей для обнаружения аномалий [22], но также допускает применение в обнаружении изменений [14]. Однако полная спектральная выборка не всегда необходима. В зависимости от приложения существует ряд спектральных диапазонов, которые обеспечат наибольшую полезность, а добавление большего количества диапазонов часто приводит к снижению результирующего качества [11].

Многоспектральные изображения очень распространены, просты и доступны для получения. Таким образом, главным источником информации в дистанционном зондировании, об изменении объектов на поверхности Земли, являются разновременные мультиспектральные снимки. Их легко интерпретировать и, в этом отношении, их легче вручную аннотировать. Многоспектральные изображения наследуют некоторые проблемы как от панхроматических, так и от гиперспектральных изображений. Из-за большего количества спектральных диапазонов необходимо эффективно использовать эту дополнительную информацию. С другой стороны, хотя их высокое пространственное разрешение предоставляет ценную информацию, которую можно использовать для обнаружения изменений, оно также фиксирует больше изменчивости, деталей и высокочастотного шума, что, в свою очередь, делает автоматическое обнаружение изменений более сложным.

ГЛАВА 2 Обзор методов, алгоритмов и технологий для обнаружения изменений на кадрах видеопоследовательности

2.1 Методы, основанные на алгебраических вычислениях

Точность результатов обнаружения изменений на кадрах видеопоследовательности зависит от многих факторов, включая:

- a) точную геометрическую регистрацию между разновременными изображениями,
- b) калибровку или нормализацию между многовременными изображениями,
- c) сложность ландшафта и окружающей среды исследуемой области,
- d) используемые методы или алгоритмы обнаружения изменений,
- e) временные и финансовые ограничения.

Универсальных методов анализа видеопоследовательностей, которые могли бы соперничать с эффективностью восприятия и анализом информации человеком, не существует. Данная проблема обусловлена сложностью человеческого восприятия, которое включает в себя не только распознавание объектов, но и глубокое понимание контекста ситуации, что в целом делает задачу автоматического анализа видеоданных особенно сложной. Разработка алгоритмов, способных учитывать такие аспекты, является активной областью исследований в сфере компьютерного зрения и машинного обучения.

В каждом кадре видео содержится много информации, которую можно выявить на различных уровнях. Используя содержащуюся в видеоизображениях информацию, такую как цвет, пространственные соотношения, форма и движение, появляется возможность определить местоположение отдельных объектов. Методы обнаружения изменений в основном полагаются на автоматический анализ данных об изменениях, полученных из разновременных изображений.

Алгебраические методы обычно состоят из двух этапов. На первом этапе строится разностное изображение, например, путем взятия разницы или соотношения значений пикселей двух изображений. На втором этапе используется форма порогового значения для создания карты изменений.

Карта изменений – это двумерный массив или изображение, которое иллюстрирует изменения между двумя или более изображениями. Она служит для визуализации областей, где произошли изменения, и часто используется в задачах, связанных с обнаружением изменений, мониторингом объектов, анализом сцен и другими приложениями.

Некоторые примеры методов обнаружения изменений на основе алгебры включают в себя дифференциацию изображений, регрессию изображений, пропорционирование изображений и другие. Эти методы могут обнаруживать изменения, которые превышают выбранный порог. Выбор порогового значения имеет решающее значение для производительности этих алгоритмов, поскольку он определяет их специфичность и чувствительность, и его необходимо корректировать для каждого уникального набора данных и каждой цели. Рассмотрим подробнее некоторые из методов.

1) Разность изображений

Работа метода заключается в вычитании первого изображения из второго, пиксель за пикселем. Данный алгоритм простой и понятный, его легко интерпретировать результаты. Однако невозможно предоставить подробную матрицу изменений, так как требуется выбор пороговых значений.

2) Вычитание фона

Введем новое понятие: фоновая модель представляет собой статическое изображение, предполагается, что на нем нет объектов. Пиксели помечаются как объект (значение 1) или не объект (значение 0) на основе порогового значения абсолютной разницы интенсивности между текущим кадром и фоном. Вычитание фона позволяет достаточно точно извлечь форму объекта, при условии, что интенсивность/цвет объекта достаточно сильно отличается от фона. Легко реализовать, но низкая точность, так как вычитание фона чувствительно к изменению освещения и незначительному движению фона.

3) Анализ вектора изменений (CVA)

Метод анализа вектора изменений является наиболее часто используемым методом. CVA обычно применяется к мультиспектральным изображениям, так как при рассмотрении всех спектральных каналов используется вся доступная информация о рассматриваемых изменениях [4].

4) Медленный анализ функций (SFA)

Существует метод обнаружения изменений, основанный на медленном анализе признаков, целью которого является поиск наиболее инвариантного компонента в многовременных изображениях для выделения измененных областей [24]. Помимо обнаружения изменений, данный метод также используется для радиометрической коррекции [26] и обнаружения изменения сцены [25].

Выделяют группы методов на основе преобразования, они основаны на первом преобразовании изображений таким образом, чтобы усилить изменения и подавить очевидные различия неизмененных областей.

Примерами таких методов являются анализ главных компонент (РСА), многомерное обнаружение изменений (MAD), метод Грамма-Шмидта, преобразование «шапки с кисточкой» – Tasseled Cap. Некоторые из этих методов, такие как РСА, также можно отнести к классическому машинному обучению. Эти методы подчеркивают интересующую информацию. Однако они по-прежнему полагаются на соответствующий выбор порогового значения для обнаружения изменений. Также важно отметить, что интерпретация измененных областей может быть более сложной при использовании преобразованных изображений. Далее будут представлены краткие характеристики вышеперечисленных методов.

1. Многомерное обнаружение изменений (MAD)

Многомерное обнаружение изменений является алгоритмом обнаружения изменений, основанным на каноническом корреляционном анализе, целью которого является максимизация дисперсии различий в проекционных характеристиках [17].

2. Преобразование «шапки с кисточкой»

Преобразование Tasseled Cap является частным случаем анализа главных компонент, который преобразует данные изображений с новую систему координат с новым набором ортогональных осей. Основная ось, яркость, вычисляется статистически как взвешенная сумма отражений всех спектральных каналов и описывает наибольшее разнообразие на изображении. Это преобразование называется Tasseled Cap (шапка с кисточкой) из-за формы графического распределения данных.

3. Метод анализа главных компонент и К-средних (РСА-K-means)

Данный метод разбивает разностное изображение на непересекающиеся блоки. Ортонормированные собственные векторы извлекаются посредством анализа главных компонент из набора блоков для создания пространства собственных векторов. Каждый пиксель в разностном изображении представлен S-мерным вектором признаков, который представляет собой проекцию данных разностного изображения на сгенерированное пространство собственных векторов. Обнаружение изменений достигается путем деления векторного пространства признаков на два кластера с использованием К-средних, каждому признаку ставится в соответствие один из кластеров по Евклидовому расстоянию [5]. Схематическая последовательность шагов изображена на рисунке 2.1.



Рисунок 2.1 – Схема работы метода анализа главных компонент и К-средних

Выделяют группы методов, основанных на статистических свойствах либо всего изображения, либо его частей. Распределение пикселей и их свойства используются для обнаружения различных или аномальных объектов между изображениями. Статистические методы чаще всего используются для гиперспектральных изображений, поскольку они, как правило, демонстрируют меньше видимых изменений, чем многоспектральные изображения с высоким разрешением.

Пиксельная классификация изображений, предшествующая обнаружению изменений, также может быть включена в категорию методов, основанных на преобразовании. Она преобразует изображение в новую упрощенную его версию — карту классов, которая упрощает идентификацию изменений. Объекты на изображениях сначала классифицируются по категориям посредством семантической сегментации, затем карты классификации разных изображений сравниваются, чтобы выявить изменения. Однако этот процесс не чувствителен к изменениям, которые происходят без изменения категории, например, изменениям в одном здании.

2.2 Методы глубокого обучения

Последняя группа методов, которая будет рассмотрена, это обнаружение изменений на основе глубокого обучения. Глубокое обучение — это подмножество машинного обучения. Оно использует нейронные сети, состоящие из большого количества слоев (отсюда и название «глубокое»), для изучения и представления данных. С развитием и растущей популярностью методов глубокого обучения в компьютерном зрении, естественно также применять их к проблеме обнаружения изменений на изображениях. Модели глубокого обучения способны представлять сложные и иерархические особенности в данных, что делает их востребованными для обнаружения изменений в дистанционном зондировании.

Одной из самых больших проблем обнаружения изменений является наличие очевидных изменений. Каждый пиксель с изображения, снятого в более раннее время, может быть изменен на изображении, снятом в более позднее время, без каких-либо изменений в изображенных на них объектах. Это создает проблему для классических алгоритмов (алгебраических, статистических, основанных на преобразованиях), которые являются более жесткими и не способны представлять сложные признаки, что приводит к множеству ложных срабатываний и необходимости тонкой настройки порога обнаружения для каждого приложения. Благодаря гибкости, масштабируемости и иерархической структуре моделей глубокого обучения они имеют потенциал для обучения представлению данных, которые слишком сложны для описания более простыми моделями. Таким образом, в алгоритмах глубокого обучения каждый пиксель рассматривается в контексте окружающих его пикселей, поскольку для принятия решений используются признаки более высокого уровня.

Классические подходы к использованию многоспектральных изображений часто опираются на спектральные индексы [3], которые объединяют различные диапазоны спектров для подчеркивания желаемых свойств. Выбор правильного индекса или комбинации диапазонов является задачей, требующей экспертных знаний, и, в конечном итоге, используется только часть информации, содержащейся в выбранных диапазонах. Глубокое обучение позволяет использовать все доступные диапазоны без необходимости предварительного выбора под руководством эксперта. Важность и вклад каждого диапазона изучаются моделью самостоятельно в процессе обучения. Модели обнаружения изменений с глубоким обучением разделяют на две категории: полностью контролируемые модели и модели, которые не полностью контролируемы, то есть полуконтролируемые и

неконтролируемые. Полностью контролируемые методы почти всегда требуют большого количества размеченных данных для обучения сети, в то время как полуконтролируемые и неконтролируемые методы уменьшают или устраняют необходимость в размеченных данных.

В отличие от многих основных задач в компьютерном зрении (таких как классификация или семантическая сегментация), обнаружение изменений получает в качестве входных данных два (или более) изображения, а не одно, вместе с одним наземным истинным изображением.

Два входных изображения могут быть обработаны различными способами, которые обобщенно разделяют на однопоточные и двухпоточные структуры:

- I. В однопоточной структуре, также называемой ранним слиянием, два изображения объединяются до того, как они будут поданы в сеть, либо путем конкатенации, либо другими процедурами, такими как дифференциация. Среди современных моделей наиболее распространены сети на основе архитектуры UNet [20]. Общая структура однопоточной сети на основе UNet показана на рисунке 2.2. Входные изображения сначала объединяются, а затем вводятся в сеть типа UNet с кодером, декодером и пропусками для семантической сегментации. Выход модели — карта изменений.

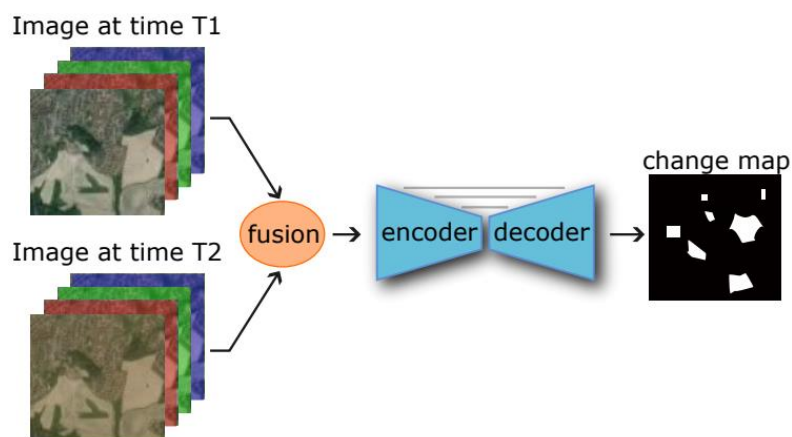


Рисунок 2.2 – Структура однопоточной нейронной сети типа UNet

- II. Двухпоточная структура основана на обработке каждого из изображений по отдельности, прежде чем они снова будут объединены и сравнены. Двухпоточные архитектуры включают две идентичные подсети, которые работают параллельно друг другу, причем каждая из них принимает одно из изображений в качестве входных данных. Этот тип структуры обычно называют сиамской, если подсети имеют общие веса, и псевдосиамской, если они не имеют общих весов. На рисунке 2.3

представлены общие структуры нейронных сетей данных архитектур. Для двухпоточных нейронных сетей типа UNet каждое изображение вводится отдельно в кодер. Выход двух кодеров объединяется на нескольких уровнях и подается на один декодер, который создает карту изменений. Для нейросетей на основе сиамского извлечения признаков каждое изображение сначала вводится в свои идентичные сиамские (общие веса) или псевдо-сиамские (разные веса) подсети для извлечения признаков. Затем извлеченные признаки объединяются вместе, и для создания карты изменений используется либо автоматическая модель принятия решений, либо классификатор на основе машинного обучения.

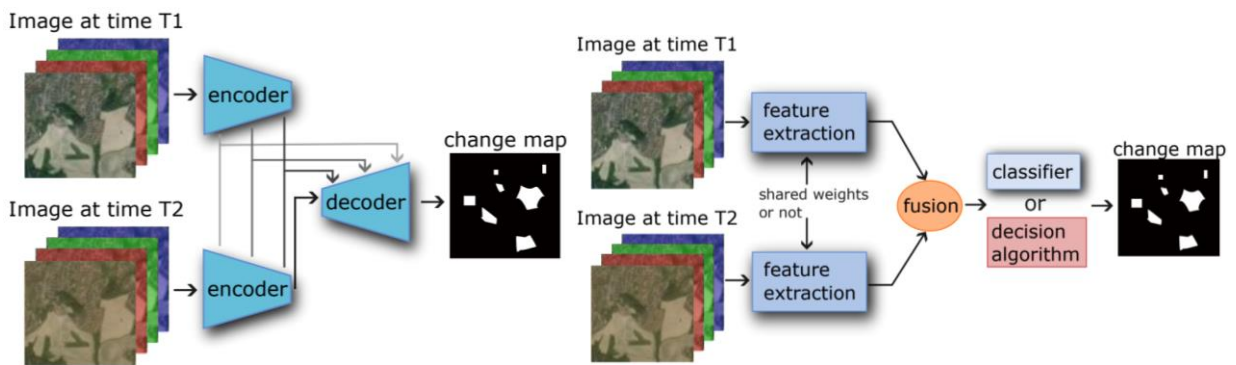


Рисунок 2.3 – Слева – структура двухпоточной нейронной сети типа UNet, справа – общая структура сети на основе сиамского извлечения признаков

Рассмотрим некоторые примеры алгоритмы глубокого обучения подробнее:

1. Глубокий анализ медленных функций (DSFA)

Основываясь на теории глубокого сетевого и медленного анализа признаков, был предложен алгоритм обнаружения изменений для многовременных изображений дистанционного зондирования, называемый Deep Slow Feature Analysis (DSFA) – Глубокий анализ медленных функций. DSFA — это модель обнаружения изменений, которая использует двухпоточную глубокую нейронную сеть для изучения нелинейных функций и выделения изменений с помощью линейного SFA.

В модели DSFA две симметричные глубокие сети используются для проецирования входных данных двух разновременных изображений. Затем данные проходят через модуль SFA для подавления неизменных компонентов и выделения измененных компонентов преобразованных функций. Предварительное обнаружение CVA используется для поиска неизменных пикселей с высокой достоверностью в качестве обучающих выборок. Интенсивность изменений рассчитывается с помощью расстояния хи-квадрат, а изменения определяются с помощью пороговых алгоритмов.

Эксперименты показали, что DSFA может превзойти другие современные алгоритмы, включая другие методы глубокого обучения на основе SFA [9]. Схематическая структура алгоритма DSFA представлена на рисунке 2.4.

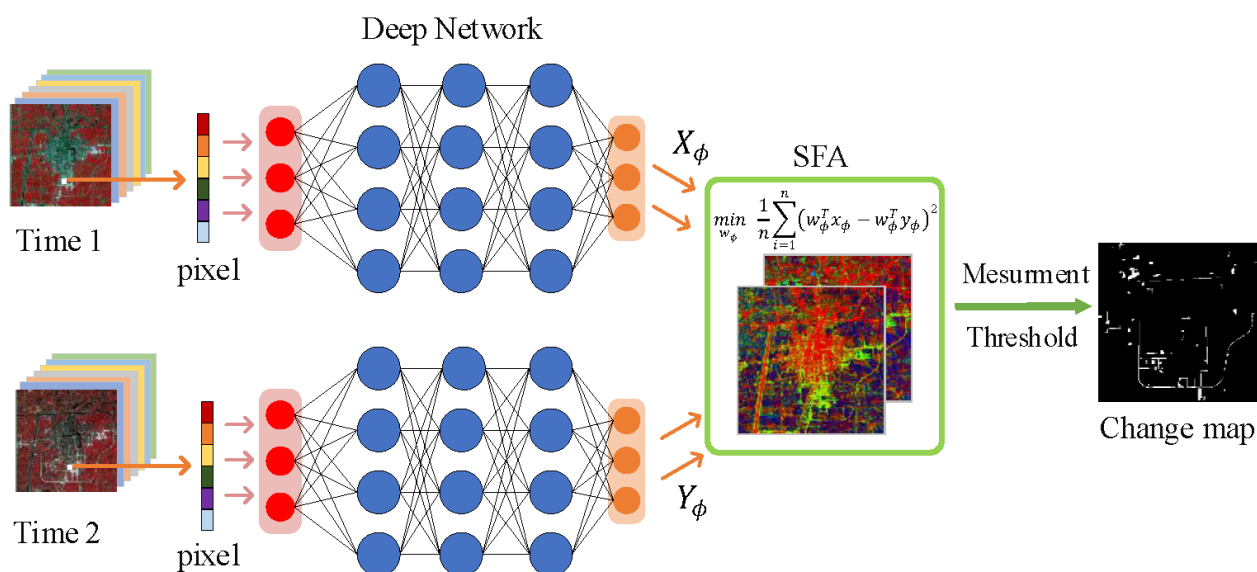


Рисунок 2.4 – Схема алгоритма глубокого анализа медленных функций

2. Глубокая сиамская сверточная многоуровневая рекуррентная нейронная сеть (SiamCRNN)

SiamCRNN — это сквозная общая архитектура обнаружения изменений с несколькими источниками, которая состоит из трех подсетей: глубокой сиамской сверточной нейронной сети (DSCNN), многоуровневой RNN (MRNN) и полносвязных слоев (FC). Глубокая сиамская сверточная нейронная сеть имеет гибкую структуру для изображений из нескольких источников и способен извлекать пространственно-спектральные характеристики из участков изображения. Многоуровневая RNN, объединенная блоками долговременной памяти (LSTM), отвечает за отображение пространственно-спектральных признаков, извлеченных DSCNN, в новое пространство скрытых признаков и анализ информации об изменениях между ними. Кроме того, последняя часть SiamCRNN, полносвязный слой используется для прогнозирования вероятности изменений. Схематическая структура глубокой сиамской сверточной многоуровневой рекуррентной нейронной сети представлена на рисунке 2.5.

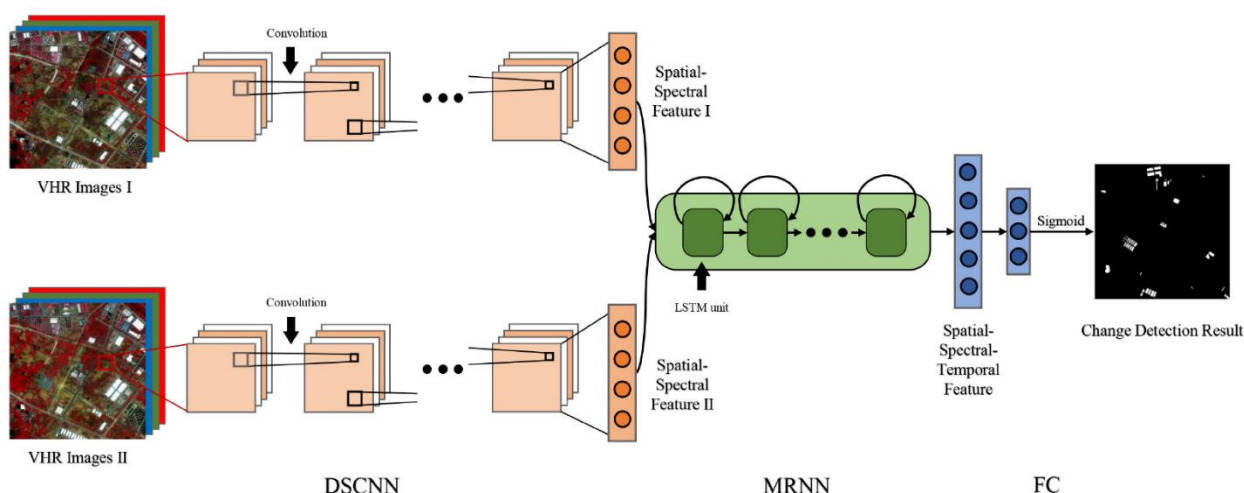


Рисунок 2.5 – Схема глубокой сиамской сверточной многоуровневой рекуррентной нейронной сети

3. Архитектура UNet

UNet и различные модифицированные версии UNet часто используются для задачи обнаружения изменений. Эта полностью сверточная нейронная сеть была разработана для семантической сегментации и, таким образом, хорошо подходит для задачи обнаружения изменений по пикселям. Чтобы обеспечить предсказание класса для каждого пикселя в изображении, UNet состоит из кодировщика и декодирующей части. За сжимающим экстрактором признаков, кодировщиком, следует расширяющийся декодер, который использует повышающую дискретизацию вместо объединения для повышения разрешения выходных данных. Таким образом, UNet названа так, так как состоит из энкодера и декодера, организованных в форме латинской буквы "U".

Siamese UNet — это модификация архитектуры U-Net, которая также часто встречается в задачах семантической сегментации и обнаружения изменений. Siamese UNet использует две (или более) идентичные подархитектуры UNet, которые принимают разные входные данные (например, разные изображения или преобразования одного и того же изображения).

Типичная архитектура UNet представлена на рисунке 2.6.

4. Архитектуры полностью сверточных нейронных сетей (FCNN)

Данные архитектуры были специально разработаны для обнаружения изменений на разновременных парах изображений наблюдений Земли. К ним относятся FC-EF, FC-Siam-conc, FC-Siam-diff.

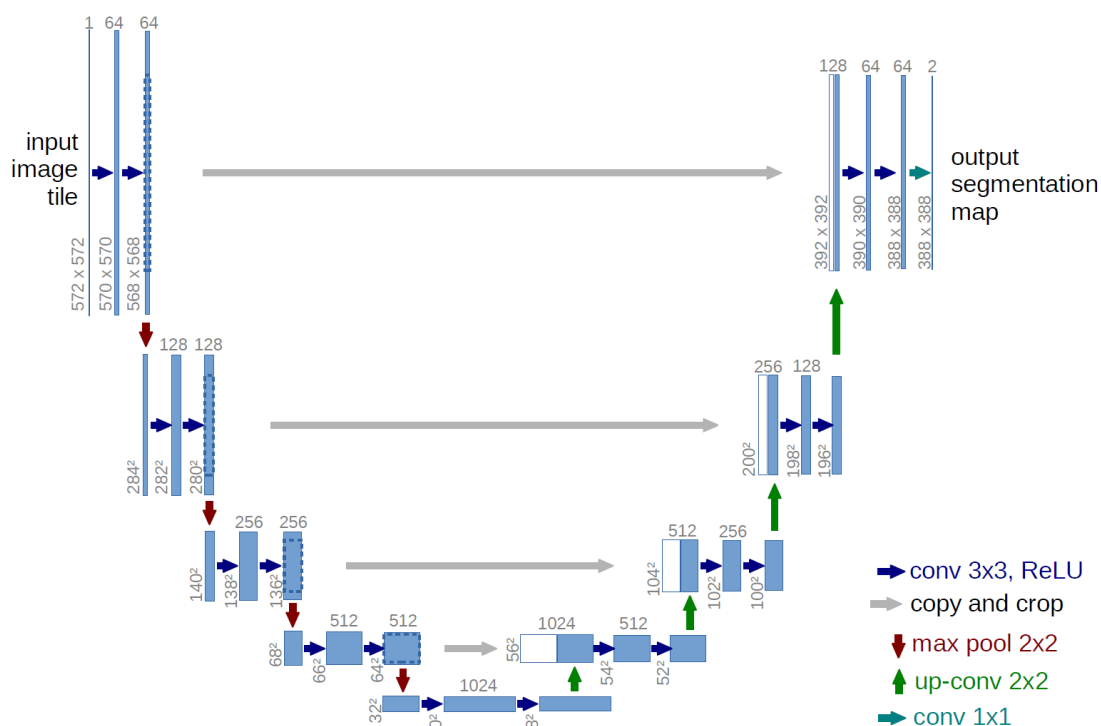


Рисунок 2.6 – Структура архитектуры UNet

Они способны обучаться выполнять обнаружение изменений исключительно на основе наборов данных обнаружения изменений без какой-либо предварительной подготовки или переноса обучения из других наборов данных. Эти архитектуры обучаются сквозным способом с нуля, используя только доступные наборы данных обнаружения изменений. Нейросети можно использовать как в RGB диапазонах изображений, так и для мультиспектральных диапазонов. Данные архитектуры используют полностью сверточные сети кодера-декодера, которые используют концепцию пропусков соединений, введенную в [21]. В [1] впервые предложили две полностью сверточные сиамские архитектуры, использующие пропуски соединений: FC-Siam-conc, FC-Siam-diff.

Эти полностью сверточные сети также способны обрабатывать входные данные любого размера при наличии достаточной памяти. Пропускные соединения представляют собой связи между слоями в одном и том же масштабе подвыборки до и после кодирующей части архитектуры кодера-декодера. Мотивацией для их использования является дополнение нейронной сети более абстрактной и менее локализованной информации пространственными деталями, которые присутствуют в более ранних слоях сети, для получения точного предсказания класса. Структуры рассматриваемых нейронных сетей представлены на рисунке 2.7.

Первая предложенная архитектура напрямую основана на модели UNet и была названа Fully Convolutional Early Fusion. FC-EF содержит только

четыре слоя максимального объединения и четыре слоя повышения дискретизации вместо пяти, присутствующих в модели UNet. Слои в FC-EF также мельче, чем их эквиваленты в U-Net. Как и в модели EF на основе патчей, вход этой сети представляет собой конкатенацию двух изображений в паре, которую необходимо сравнить.

Две другие предложенные архитектуры являются сиамскими расширениями модели FC-EF. Для этого слои кодирования сети разделяются на два потока одинаковой структуры с общими весами, как в традиционной сиамской сети. Каждое изображение передается в один из этих равных потоков. Разница между двумя архитектурами заключается только в том, как выполняются пропуски соединений.

Первый и более интуитивный способ — это объединить два пропускаемых соединения во время шагов декодирования, каждое из которых исходит из одного потока кодирования. Этот подход был назван Fully Convolutional Siamese - Concatenation. Поскольку в задаче обнаружения изменений необходимо обнаружить различия между двумя изображениями, эта же эвристика может быть использована для объединения пропускаемых соединений.

Подход, в котором вместо объединения двух соединений из потоков кодирования, объединяются абсолютное значение их разницы, был назван Fully Convolutional Siamese - Difference.

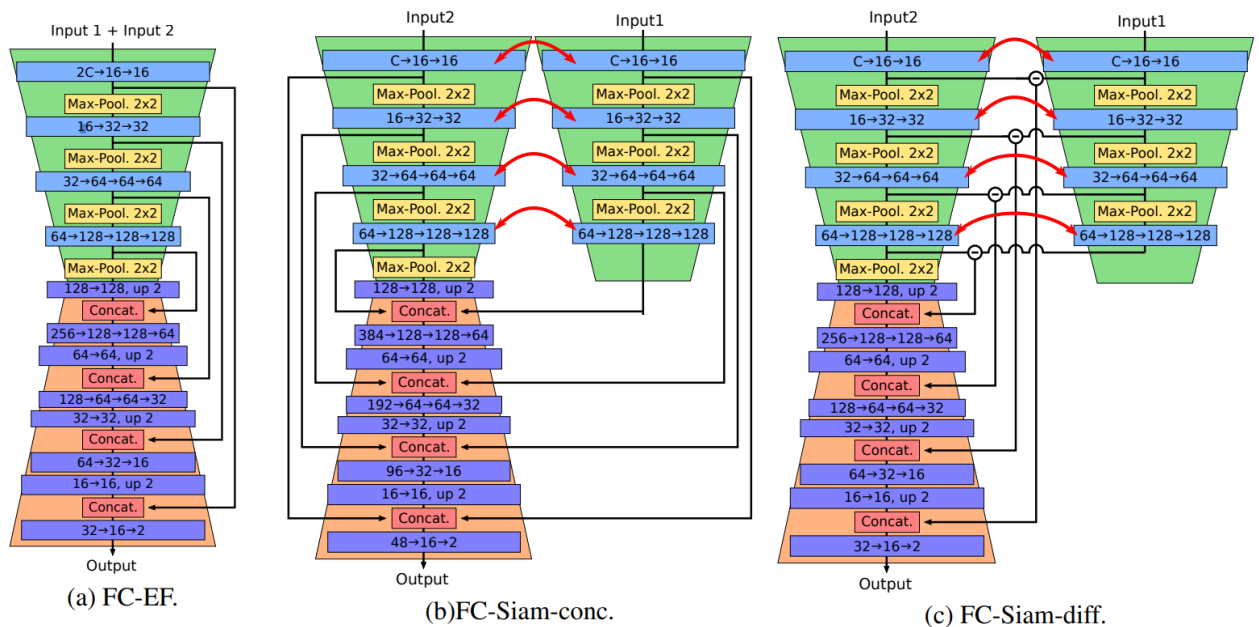


Рисунок 2.7 – Структуры трех архитектур полностью сверточных нейронных сетей

ГЛАВА 3 Введение в теорию обработки видеопоследовательностей

3.1 Общие понятия о видео

Видеопоследовательность представляет собой временную последовательность цифровых изображений, называемых кадрами. При быстром последовательном воспроизведении этих кадров создается иллюзия движения. В контексте цифровой обработки изображений видеопоследовательность зачастую рассматривают именно как набор дискретных образов, каждый из которых является двумерной решеткой пикселей. Каждый пиксель представляет собой выборку изменяющегося в пространстве интенсивного светового потока, зафиксированного на фокальной плоскости. Эта концепция лежит в основе многих современных технологий, включая видеокодеки, сжатие и анализ видео, что делает ее важной для обнаружения изменений на видеопоследовательности.

Для систематического анализа видеопоследовательностей важно выделить их ключевые характеристики, которые влияют на хранение, работу с ними и качество последующей обработки.

Основные характеристики видеопоследовательностей включают:

1. Частота кадров

Частота кадров видеопоследовательности определяется как количество кадров, отображаемых в секунду (fps – frame per second, кадр в секунду). Этот параметр напрямую влияет на плавность воспроизведения движения в видео. Стандартные частоты включают 24 fps для кинематографических фильмов, 30 fps для телевидения. Более высокие частоты, такие как 60 fps или выше, используются в высокоскоростных камерах или для создания эффектов замедленной съемки, так как таким образом запечатлеваются минимальные изменения между кадрами.

2. Разрешение

Разрешение видео является понятием, которое указывает на количество пикселей в каждом кадре. Оно обычно выражается в формате ширина × высота, например, 1920×1080 для Full HD или 3840×2160 для 4K. Более высокое разрешение обеспечивает большую детализацию и ясность изображения, что особенно важно для современных высококачественных дисплеев и профессионального видео. Разрешение напрямую влияет на объем данных и требования к вычислительным ресурсам при обработке, что является важным аспектом для задач, связанных с обнаружением объектов и изменений.

3. Цветовое пространство

Цветовое пространство определяет, как цвета представлены в видео. Наиболее распространенные цветовые пространства включают RGB (красный, зеленый, синий), которое используется для отображения на экранах, и YUV (яркость – люминанса и два цветоразностных компонента – хроминанса), которое часто применяется в видеокодировании для разделения яркости и цветности, что позволяет более эффективно сжимать данные. Другие цветовые пространства, такие как CMYK (голубой, пурпурный, желтый, черный), используются в печати, а HSV (оттенок, насыщенность, яркость) — в определенных задачах обработки изображений, например, для сегментации объектов.

4. Длительность

Длительность видеопоследовательности — это общее время от первого до последнего кадра, измеряемое в секундах или минутах. Этот параметр важен для понимания временного аспекта видео. Также он является одним из ключевых при определении трудозатратности решения поставленных задач анализа видео.

5. Структура видеоданных

Видеоданные могут храниться в различных форматах, которые могут быть сжатыми или несжатыми (например, RAW). Программы, которые кодируют и декодируют данные, называются кодеками.

Имеют место следующие сжатые форматы видеоподанных:

1) MP4 – формат, основанный на стандарте MPEG-4, который поддерживает как видео-, так и аудио- и метаданные (обложки альбомов, субтитры и другое). Является наиболее универсальным и широко поддерживаемым форматом, подходящим для потокового видео и использования на мобильных устройствах.

2) AVI – контейнерный формат, разработанный компанией Microsoft в 1992 году как часть технологии Video for Windows. Он обеспечивает высокое качество, но с большими размерами файлов и менее широкой поддержкой современных технологий.

3) MOV – формат, разработанный компанией Apple для QuickTime и часто используется на устройствах Apple. На других платформах могут возникнуть проблемы с кодеками и программным обеспечением. Обычно файлы MOV также имеют большой размер по сравнению с MP4.

Сжатые форматы используют алгоритмы, такие как MPEG, H.264, для уменьшения размера файла при сохранении приемлемого качества. MPEG обозначает более широкий набор стандартов и технологий для сжатия видео и аудио, среди которых H.264 является одним из более современных и

эффективных. H.264 известен также как MPEG-4 Part 10 или AVC (Advanced Video Coding).

Несжатые форматы сохраняют всю информацию с датчика камеры, но требуют значительно больше места для хранения. Выбор формата влияет на совместимость, качество и требования к хранению, что критично для задач, связанных с обработкой больших объемов данных, таких как обнаружение изменений в длинных видеопоследовательностях.

Понимание всех характеристик видеоданных является важным для работы в области их обработки. Они служат основой для принятия решений о дальнейшем анализе видео, что будет рассмотрено в последующих главах. Например, частота кадров и их разрешение напрямую влияют на точность алгоритмов обнаружения движения. Структура данных, в свою очередь, определяет возможности применения специфических методов, что важно для практического применения в реальных системах.

Развитие технологий обработки видео началось еще в 1960-х годах с использованием цифровых компьютеров для улучшения качества изображений. В 1964 году были получены снимки Луны, сделанные зондом Ranger 7, которые подстегнули развитие данной области науки. Сегодня видеопоследовательности широко применяются в таких областях, как видеонаблюдение, медицина, кинематограф и автоматическое распознавание объектов. Современные стандарты, такие как H.264 и H.265, подчеркивают важность эффективного сжатия, а высокое разрешение, такое как 8K, требует новых подходов к обработке.

3.2 Режимы обработки видео и методы, применимые для них: офлайн и онлайн подходы

Задачи непосредственной обработки видео для обнаружения изменений сцен и объектов видеопоследовательности делятся на две основные группы.

Первая группа задач связана с обработкой видео после его полного сохранения – офлайн обработкой. Данный подход удобен и более выгоден ресурсно при ведении пассивного наблюдения, где вмешательство человека в процесс, за которым ведется видеонаблюдение, не требуется. Примером может служить видеосъемка дикой природы, где задача заключается в выделении фрагментов, на которых появляются животные, без необходимости взаимодействия самим животным.

Для офлайн обработки видео могут быть предложены различные методы, включая традиционные подходы, такие как анализ разности кадров, вычисление оптического потока и сравнение гистограмм.

Ниже представлено детальное описание каждого метода с учетом ограничений.

1) Анализ разности кадров

Анализ разности кадров является базовым методом обработки видео, который выявляет изменения между двумя последовательными кадрами. Идея заключается в том, чтобы вычесть один кадр из другого попиксельно, чтобы выделить области, где произошли изменения, такие как движение объектов или изменения в сцене.

Для работы с кадрами опционально производится предварительное преобразование изображения в оттенки серого для упрощения вычислений и уменьшения влияния цветовых изменений. Алгоритм состоит в получении двух последовательных кадров, после чего каждый пиксель одного кадра вычитается из соответствующего пикселя другого кадра. Результатом является изображение, называемое картой разности, где области с изменениями имеют ненулевые значения, а неизменные области — близки к нулю. Для отделения значимых изменений от шума применяется пороговое значение: пиксели со значением выше значения порога считаются измененными. Для уточнения результатов также используют морфологические операции эрозии и дилатации, чтобы удалить шум и заполнить разрывы в обнаруженных областях. Эрозия и дилатация часто используются вместе в обработке изображений. Сначала может быть использована эрозия для удаления шумов, а затем дилатация, чтобы восстановить размеры объектов. Этот подход называется «морфологическим закрытием» или «морфологическим открытием», в зависимости от порядка применения операций. Полученное бинарное изображение используется для выявления движущихся объектов или изменений в сцене.

Данный метод прост в реализации и эффективен для задач, где требуется быстрое обнаружение движения или изменений. К его ограничениям относят чувствительность к шуму, так как любое незначительное изменение освещения зачастую приводит к ложным срабатываниям из-за попиксельного сравнения. Этот недостаток делает метод менее предпочтительным для сложных сцен с множеством движущихся объектов по причине сильно неточных результатов.

2) Оптический поток

Оптический поток является более сложным методом, который оценивает движение объектов между двумя последовательными кадрами видео. Он позволяет не только обнаружить изменения, но и понять направление и скорость движения объектов.

Данный метод вычисляет поле скорости, которое описывает, как каждый пиксель перемещается от одного кадра к другому. Таким образом оптический поток определим, как векторное поле, где каждому пикселю присваивается вектор, показывающий направление и скорость его перемещения.

Если оптический поток применяется к каждому пикселю изображения, то он называется плотным. С помощью плотного потока получают полное поле движения, не ограниченное только ключевыми точками или областями, что характерно для разреженного оптического потока.

Для использования оптического потока вводится аксиоматическое предположение постоянства яркости о том, что интенсивность пикселя остается постоянной при его перемещении между кадрами, иначе точность метода снижается. Другим минусом является то, что вычисление оптического потока, особенно плотного, требует значительных ресурсов, что может быть проблемой для реального времени.

К основным алгоритмам вычисления оптического потока относят следующие:

а. Метод Лукаса-Канаде [15]

Он используется для вычисления оптического потока в разреженных точках (sparse optical flow). Вводится предположение, что поток в локальной окрестности пикселя (например, в окне 3x3) является постоянным. Данный метод решает базовые уравнения оптического потока для всех пикселей в окне с использованием критерия наименьших квадратов. Преимущества метода в том, что он устойчив к шуму, благодаря использованию информации от нескольких пикселей. Также он может быть вычислен эффективно, что делает его подходящим для реального времени при небольших движениях. Однако в однородных областях, где нет градиентов, метод выдает менее точные результаты, так как там отсутствует информация для оценки. Уравнение для нахождения вектора скорости для метода Лукаса-Канаде представлено на формуле ниже.

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial t} = 0,$$

где V_x, V_y – компоненты скорости оптического потока $I(x, y, t)$;
 $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t}$ – производные изображения в (x, y, t) в соответствующем направлениях.

б. Метод Хорна-Шунка

Метод Хорна-Шунка является глобальным методом оценки оптического потока, который вычисляет плотный оптический поток (dense optical flow). Он

минимизирует глобальную ошибку по всему изображению. Для использования метода вводится глобальное ограничение гладкости, то есть полагается, что соседние пиксели имеют похожие векторы потока. Это необходимо для решения проблемы апертуры, которая возникает, когда движение не может быть определено только на основе локальной информации (например, для краев). Таким образом метод может заполнять информацию о потоке в однородных областях, используя ограничение гладкости. К минусам относят то, что этот метод чувствителен к шуму из-за глобального подхода и вычислительно более интенсивен, чем локальные методы, что может быть проблемой для реального времени.

с. Алгоритм Фарнебека

Данный алгоритм считается эффективным методом для плотного оптического потока и известен своей точностью, особенно для больших движений. Он вычисляет поток для каждого пикселя в изображении и основан на полиномиальном расширении. Метод использует многомасштабный подход (пирамиду изображений), где на каждом уровне пирамиды окрестность каждого пикселя аппроксимируется квадратичным полиномом, в отличие от предыдущих методов, где строятся полиномы первой степени. Оценка смещения минимизирует разницу между полиномами в двух кадрах и уточняется итеративно от более грубого к более тонкому масштабу, комбинируя результаты с разных уровней пирамиды.

Подводя итог, оптический поток позволяет следить за движением конкретных точек или объектов на протяжении нескольких кадров. С помощью анализа поля скорости, распознают сложные движения, такие как жесты или действия.

Другими полезными приложениями является то, что посредством оптического потока можно корректировать движение камеры для стабилизации видеозаписи и оценивать глубину и восстанавливать трехмерные сцены из двухмерных видеопоследовательностей.

3) Сравнение гистограмм интенсивности пикселей

Данный метод используется для измерения сходства между двумя изображениями или кадрами видео путем анализа их гистограмм. Гистограмма представляет собой распределение интенсивностей пикселей в изображении. Сравнение гистограмм может помочь выявить значительные изменения в сцене, такие как смена кадра или эффект затухания.

Алгоритм состоит в том, что для каждого кадра вычисляется гистограмма, это может быть гистограмма интенсивности для черно-белых изображений или гистограммы для каждого цветового канала (например, R, G, B) для цветных изображений. Далее выбирается метрика, по которой будут

сравниваться гистограммы – вычисляют степень сходства между гистограммами двух кадров.

Для сравнения гистограмм применяются разнообразные метрики, к числу наиболее распространенных относятся следующие:

а. Корреляция

Она определяет степень линейной зависимости между двумя гистограммами, отражая, насколько их формы согласуются в рамках линейной модели.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}},$$

где r – коэффициент корреляции;

x_i, y_i – значения двух выборок;

$\bar{x}, \bar{y}$ – средние значения из выборок.

б. Хи-квадрат

Данная метрика вычисляет расхождение между наблюдаемыми и ожидаемыми частотами, предоставляя количественную оценку различий между гистограммами. Основная формула для расчета хи-квадрата по Пирсону:

$$\chi_n^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i},$$

где E_i – ожидаемое значение;

O_i – наблюдаемое, фактическое значение, полученное в эксперименте.

с. Пересечение

Пересечение – это мера, оценивающая область перекрытия двух гистограмм, что позволяет измерить их общую часть и степень схожести.

д. Расстояние Бхаттачарья

Метрика сходства, основанная на сравнении вероятностных распределений, которая эффективно выявляет близость двух гистограмм в терминах их статистической природы.

Последним этапом алгоритма также устанавливается порог: если мера сходства ниже определенного порога, это указывает на значительное изменение, такое как смена сцены.

Данный метод эффективен для выявления переходов между разными сценами в видео, где присутствуют резкие смены (cuts) или постепенные переходы (fades). К недостаткам относят то, что изменения освещения могут существенно повлиять на гистограммы, что снижает точность метода. Кроме того, гистограммы отражают глобальное распределение интенсивностей,

поэтому метод не подходит для обнаружения локальных изменений, таких как движение небольших объектов.

Современные исследования показывают, что глубокое обучение показывает более точные результаты. Например, сверточные нейронные сети (CNN) используются для автоматического извлечения признаков и классификации изменений сцен. Эффективно применять для задач, связанных с видеонаблюдением, именно двухпоточковые нейронные сети, например, Two-Stream CNN [23], где один поток анализирует пространственные характеристики кадров, а второй — временные изменения между кадрами.

Вторая группа задач связана с обработкой видео в реальном времени, синхронно с его записью. Этот подход необходим в системах, где требуется оперативная реакция — действия человека или автоматическое оповещение. В таких случаях выделяются две подзадачи, которые могут решаться:

1. Так же разбиение видеопотока на клипы или кадры с последующим сравнением для обнаружения изменений
2. предсказание наличия движения в будущем и определение его характеристик.

Для первой подзадачи подходят некоторые уже рассмотренные методы на основании оптического потока. Большой интерес для исследования в современное время представляется задача прогнозирования того, произойдет ли движение в видеопотоке и, в случае положительного ответа, дополнительного определения характера движения. Для решения задачи предсказания широко применяются подходы на основе глубокого обучения.

Рассмотрим некоторые алгоритмы, которые были предложены в последние годы и показали значительные результаты.

- 1) Задаче вычитания фона (BGS) — фундаментальная задача обработки видео, которая является ключевым компонентом многих приложений, в том числе для распознавания активности на видео. BGS можно определить как бинарную сегментацию видео, направленную на выделение областей переднего плана и фона в каждом видеокadre. Один из простых подходов к решению этой задачи — создание модели фона, и каждый новый кадр сравнивается с ней для выявления объектов на переднем плане. Одним из наиболее успешных подходов к повышению обобщающей способности алгоритмов компьютерного зрения, обучаемых на основе ограниченных данных, является использование расширения данных. Дополнения к пространственным данным, такие как случайные обрезки, повороты, изменение цвета и так далее, оказались очень успешными в задачах, связанных с изображениями. Одним из алгоритмов, который основывается на этих идеях, является алгоритм BSUV-Net [16]. Этот алгоритм может работать на

произвольных видео без аннотаций и хорошо обобщается на различные наборы данных для вычитания фона. Он обучается на аугментированных фонах из различных видео с симуляцией эффектов панорамирования, наклона, зума, а также различных условий освещения и погоды. Благодаря этому алгоритм способен работать с высокой скоростью, что делает его пригодным для реального времени применения.

- 2) Рекуррентные нейронные сети (RNN), такие как LSTM и GRU, которые способны моделировать временные зависимости в последовательностях данных. Например, в исследовании [2], LSTM используется для предсказания траекторий движения объектов в видеопотоке, что позволяет обнаруживать аномалии, такие как резкие изменения скорости или направления. Более того, недавние достижения в области самообучения привели к разработке методов, таких как TSCP2 [7], который использует контрастное предсказательное кодирование для обнаружения точек изменения во временных рядах. Этот метод особенно перспективен для задач реального времени, поскольку он не требует предварительной разметки данных, что часто является узким местом в анализе видео. Простыми словами, контрастное обучение – это подход, при котором обучение происходит не только по принципу близости, но и по принципу различия. То есть основываясь на позитивных и негативных примерах. Методы определения точки изменения (CPD) определяют моменты времени, связанные с изменениями тенденций и свойств данных временных рядов, чтобы описать базовое поведение системы.

Очень сложным сценарием является проблема обнаружения изменений с использованием видеозаписей, полученных с движущейся камеры. В таких случаях обнаружение изменений может быть выполнено путем сравнения входного видео с другим эталонным видео, представляющим нормальные условия окружающей среды. Несмотря на различия в методах, все методы для решения этой проблемы основаны на схожем конвейере обработки:

- (i) исходное и целевое видео выравниваются по времени, чтобы можно было легко сравнить соответствующие кадры
 - (ii) разница между исходным и целевым видео вычисляется в некоторой конкретной области
 - (iii) пиксели или кадры с достаточно высокой оценкой несходства связаны непосредственно с изменениями в объектах
- 3) TCF-LMO [19] – Twin-CNN Features and Learnable Morphological Operations – метод, который использует двойные сверточные сети ResNet-50 и обучаемые морфологические операции для обнаружения

изменений в видео с движущейся камерой. Этот метод имеет всего 1028 обучаемых параметров, что делает его подходящим для реального времени, особенно с учетом его 18% превосходства над текущими методами.

3.3 Форматы видеопоследовательностей: подходы к обработке сжатых и несжатых данных

Для применения любого метода или алгоритма обнаружения изменений сцен и объектов видеопоследовательности необходимо исходить из характеристик видео. Также важным этапом является предварительная обработка видео.

Видео может храниться в несжатом (растровом) или сжатом виде, рассмотрим подходы для работы с видеопоследовательностью для каждого из вариантов.

- I. Если видео не сжато, то оно может декодироваться в последовательность отдельных кадров, каждый из которых является статичным изображением. Эти кадры могут быть далее преобразованы для упрощения анализа. Вторым подходом, который может быть использован для преобразования видеопоследовательности в формат, который используется на вход в методах обработки видеопоследовательностей может быть нарезка на клипы. Клип – это последовательность кадров, непродолжительная по времени. В данном случае будем считать, что в клипе происходит только какое-то одно логическое действие или одна сцена. Таким образом клип будет ограничиваться смысловой нагрузкой в нем, а не конкретным временным промежутком. Кадр, в свою очередь, фрагмент видеоряда, статичное изображение, на котором находит отображение часть движения (при его наличии). При последовательном воспроизведении кадров создается полное движение, которое человек видит на видео. Для разделения видеопоследовательности на кадры необходимо задать частоту, с которой независимые кадровые изображения должны появляться на экране. Для человеческого глаза от 16 кадров в секунду воспринимаются как движение. Стандартной для кинематографии и видеосъемки считается частота 24 кадра в секунду, так как видео, составленные с такой частотой кадры, отображают запись с похожей быстротой на то, как человек видит окружающий мир. Для плавных движений при съемке предметов, которые изменяют свое положение с очень быстрой скоростью, используют от 30 кадров в секунду.

Деление видеопоследовательности на клипы используется в методах обнаружения смен сцен, для индексации видео. Анализ клипов также производится для ряда задач, таких как классификация сцен и других, связанных с пониманием структуры видео, например, автоматического монтажа или анализа фильмов.

В свою очередь деление на кадры применяются для обнаружения движения объектов. Методы, использующие кадры, анализируют отдельные кадры или пары последовательных кадров для выявления изменений на низком уровне. Ключевая разница в том, что клиповые методы фокусируются на структуре видео, а кадровые — на деталях внутри сцены.

Клиповые методы могут быть более эффективными для длинных видео, так как они сначала сегментируют видео на крупные единицы, а затем анализируют их. Кадровые методы, напротив, могут быть вычислительно более затратными, так как обрабатывают каждый кадр.

Методы, использующие клипы, обычно начинаются с обнаружения смен сцен (scene change detection), чтобы разделить видео на отдельные клипы, а затем анализируют каждую сцену.

Примером таких методов является PySceneDetect, который использует пороговые методы или адаптивные детекторы для выявления резких переходов или фейдов – переходов, при которых элемент постепенно исчезает и снова появляется. PySceneDetect — это приложение командной строки и библиотека Python для обнаружения изменений сцен в видео и автоматического разделения видео на отдельные клипы. Это бесплатное программное обеспечение с открытым исходным кодом, в котором доступно несколько методов обнаружения: от простого обнаружения появления и исчезновения на основе порогового значения до расширенного обнаружения быстрого перехода с учетом содержимого для каждого кадра.

PySceneDetect можно использовать самостоятельно как автономный исполняемый файл, с другими приложениями как часть конвейера обработки видео или интегрировать напрямую в другие программы или скрипты через API Python, требуется лишь наличие библиотек программного обеспечения OpenCV и Numpy.

II. Видеопоследовательности в сжатом формате, таких как MPEG или H.264, содержат информацию о кадрах в виде коэффициентов дискретного косинусного преобразования (DCT), векторов движения и других метаданных, которые доступны в сжатом потоке, что делает их идеальными для анализа в реальном времени. Коэффициенты DCT помогают анализировать пространственные изменения, векторы движения — временные, а метаданные обеспечивают контекст. Методы

обнаружения изменений могут работать напрямую в сжатом домене, что позволяет избежать затрат на полную декодировку.

Другими метаданными, которые могут храниться для видео в сжатом формате, могут быть типы кадров. MPEG формат поддерживает три типа кадров (фреймов):

а. I-фреймы (Intra-frames):

Это самостоятельные кадры, которые не зависят от других кадров. Они содержат полную информацию о кадре, как статичное изображение, и используются для начала декодирования или быстрого доступа (например, при перемотке). Имеют наибольший размер, так как не используют предсказание.

б. P-фреймы (Predictive frames):

Зависят от предыдущего I- или P-фрейма. Хранят только изменения относительно эталона, что делает их меньше по размеру. Используются для повышения сжатия, но требуют наличия эталона для декодирования.

в. B-фреймы (Bi-directional predictive frames):

Зависят от предыдущего и следующего I- или P-фрейма. Обеспечивают наибольшее сжатие, храня различия между двумя направлениями, но увеличивают задержку из-за необходимости перестановки кадров. В некоторых стандартах, таких как H.264, могут быть эталонами, но обычно не используются как таковые.

Коэффициенты дискретного косинусного преобразования (DCT) — это числовые значения, полученные в результате преобразования блоков изображения (обычно 8x8 пикселей) из пространственного домена в частотный. DCT разбивает блок изображения на сумму косинусных функций с разными частотами. Низкочастотные коэффициенты (например, DC-коэффициент) описывают общую яркость или цвет, а высокочастотные — мелкие детали. После DCT коэффициенты квантуются, что уменьшает размер данных за счет потери незначительных деталей. DCT позволяет компактно представлять визуальную информацию, удаляя избыточные высокочастотные компоненты, которые менее заметны для человеческого глаза.

Векторы движения — это данные, описывающие перемещение макроблоков (обычно 16x16 пикселей) между кадрами в сжатом видео. Они используются в P- и B-фреймах для межкадрового предсказания, указывая, как блоки из эталонного фрейма (I- или P-фрейма) перемещаются в текущем фрейме. Алгоритмы оценки движения (motion estimation) находят совпадения между макроблоками в соседних кадрах, вычисляя вектор (dx, dy), который указывает направление и величину смещения. Эти векторы кодируются в сжатом потоке, уменьшая объем данных, так как хранятся только изменения.

Векторы движения устраняют временную избыточность, позволяя хранить только различия между кадрами вместо полной информации.

Рассмотрим некоторые методы, которые могут быть использованы для обнаружения изменений в сжатых видеопоследовательностях.

- 1) Анализ коэффициентов дискретного косинусного преобразования (DCT-коэффициентов) [13]. Этот подход основан на статистическом тестировании различий между блоками DCT, при котором из значимых изменений исключаются шумовые помехи и вариации освещенности. В большинстве методов сжатия изображения разбиваются на блоки, содержащие $N \times N$ пикселей, которые затем преобразуются в коэффициенты DCT. Для анализа рассматривается пара соответствующих блоков в двух изображениях, считаются значения их DCT-коэффициентов и сравниваются с выбранным порогом.
- 2) Резкие смены сцен часто сопровождаются низкой корреляцией между кадрами, что проявляется в хаотичных или нулевых векторах движения. В статье [10] описывается использование интерполированных макроблоков в В-фреймах, которые зависят от векторов движения, для оценки корреляции и обнаружения смен сцен. Векторы движения также позволяют отслеживать перемещение объектов.
- 3) В работе «An Effective and Fast Scene Change Detection Algorithm for MPEG Compressed Videos» [12] рассматривается быстрый алгоритм обнаружения смены сцены, который работает с видеопоследовательностью в области сжатия MPEG. Предлагаемый алгоритм использует схему оценки и компенсации движения в формате MPEG, проверяя статус прогнозирования для каждого макроблока внутри В-кадров и Р-кадров. В результате определение местоположения как резких, так и размытых изменений сцены осуществляется с помощью последовательности сравнительных тестов, и не требуется выделения признаков или дифференциации гистограмм. Эксперименты показали, что данный алгоритм обеспечивает точность до 94% при обнаружении резкой смены сцены и 100% при обнаружении постепенной смены сцены. Предложенный алгоритм подходит для реализации в режиме реального времени.

Важно отметить, что существуют споры о том, считать ли работу в сжатом домене "прямой" обработкой видео, так как даже такие методы опираются на структуру, связанную с кадрами.

ГЛАВА 4 Задача обнаружения изменений сцен и объектов видеопоследовательности

4.1 Постановка задачи обнаружения изменений на видеопоследовательностях

В данной работе поставлена задача обнаружения изменений на наборе видео. Она разделена на три логические части.

1. Разделение видео на кадры
2. Выбор ключевых кадров со сменой сцен
3. Обнаружение изменений между полученными кадрами

Выбран набор из 17 коротких видео из библиотеки IoT примеров от Intel, он представлен на рисунке 4.2. В каждом из них происходит некоторый характер движения, видео подобраны так, чтобы они имели разные характеристиками. Каждое из них различной длительности и с различной частотой и размерами.

Далее приведена краткая сравнительная характеристика трех видео из набора данных по критериям, которые уже были рассмотрены в предыдущей главе:

1) Видео с камеры наблюдения, где проезжают дважды по одной машины и один раз две одновременно. Продолжительность видео 30 секунд, частота кадров 12.5 fps, ширина кадра 768 пикселей, высота кадра 432 пикселей.

2) Видеосъемка конвейера, на котором вертикально катятся винтики разных размеров. Длительность видео 1 минута 38 секунд, частота 59.94 кадров/с, разрешение 1920 x 1080.

3) Видеосъемка конвейера, на котором катятся винтики одинаковых размеров с горизонтальным и вертикальным размещением. Продолжительность 30 секунд, частота кадров 29.937 fps, разрешение видео 1024 x 576.

Библиотека PySceneDetect позволяет автоматически определять разрешение видео и также частоту кадров, что показывает более точные результаты, которые показаны на рисунке 4.1., чем встроенный в Windows 10 медиаплеер, результаты которого на рисунке 4.3.

```
INFO:pyscenedetect:Loaded 1 video, framerate: 29.937 FPS, resolution: 1024 x 576
INFO:pyscenedetect:Detecting scenes...
```

Рисунок 4.1. – Лог исполнения кода библиотеки PySceneDetect

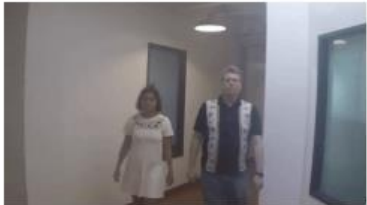
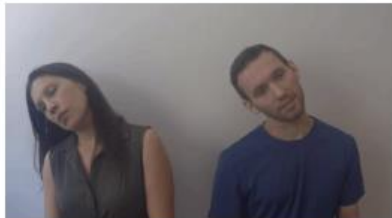
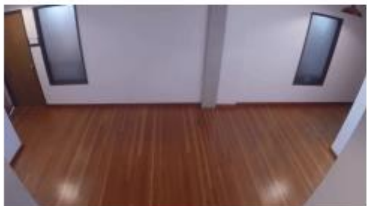
		
bolt-detection.mp4	bolt-multi-size-detection.mp4	bottle-detection.mp4
		
car-detection.mp4	classroom.mp4	face-demographics-walking-and-pause.mp4
		
face-demographics-walking.mp4	fruit-and-vegetable-detection.mp4	head-pose-face-detection-female-and-male.mp4
		
head-pose-face-detection-female.mp4	head-pose-face-detection-male.mp4	one-by-one-person-detection.mp4
		
people-detection.mp4	person-bicycle-car-detection.mp4	store-aisle-detection.mp4
		
worker-zone-detection.mp4	driver-action-recognition.mp4	

Рисунок 4.2. – Видео из выбранного набора данных

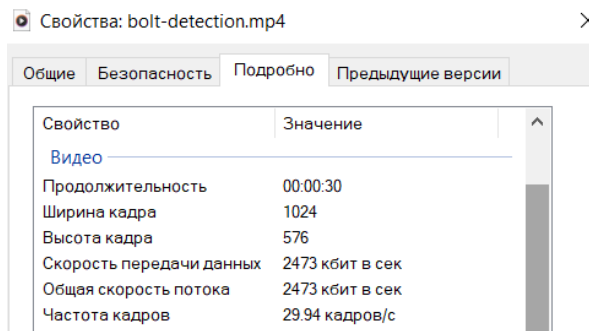


Рисунок 4.3 – Свойства видео через встроенный медиаплеер

4.2 Выбор методов для решения задачи разделения на кадры

Для нарезания видеопоследовательностей на кадры были выбраны два метода.

Первый метод извлекает кадры с заданным временным интервалом.

Второй – извлекает кадры основываясь на методе библиотеки PySceneDetect – ContentDetector, который сравнивает разницу в содержании между соседними кадрами с установленным порогом, который при превышении сокращает сцены. Он обнаруживает быстрые переходы, используя изменения в цвете и интенсивности между кадрами, разница между которыми рассчитывается в цветовом пространстве HSL. Для каждого кадра рассчитывается оценка от 0 до 255,0, которая отражает разницу в содержании между текущим и предыдущим кадром (чем выше, тем больше различий). Деление производится, когда оценка кадра превышает пороговое значение

Также существуют другие детекторы в данной библиотеке:

- ThresholdDetector обнаруживает медленные переходы (Fade In/Fade Out) с использованием средней интенсивности пикселей в цветовом пространстве RGB. Результирующие разрезы помещаются между соседними событиями затухания и увеличения яркости.
- AdaptiveDetector представляет собой двухпроходной алгоритм, который сначала вычисляет оценку кадра с помощью ContentDetector, а затем применяет скользящее среднее при обработке результата. То есть высчитывает среднее значение в различиях в цветовом пространстве HSL. Такой метод может помочь избежать ложных срабатываний в таких ситуациях, как движение камеры, и может улучшить обращение с быстрым движением.
- Histogramdetector находит быстрые переходы, сравнивая YUV-гистограммы. После преобразования каждого кадра в YUV-

формат используется Y-канал для создания гистограммы каждого кадра. Гистограммы между кадрами сравниваются, чтобы определить степень их сходства.

- Hashdetector находит быстрые переходы, используя перцептивное хэширование. Перцептивный хэш берется из соседних кадров и используется для вычисления расстояния Хэмминга между ними. Затем расстояние нормализуется по квадрату размера хэша и сравнивается с пороговым значением.

Для проверки качества разделения видео на кадры производился анализ различий между кадрами попарно по соседству с использованием трех метрик:

- процентного изменения пикселей – выявляет количественные изменения на уровне пикселей

$$a = \frac{X - Y}{Y} * 100\%,$$

где X, Y – пиксели из первого и второго кадра

- индекса структурного сходства (SSIM) – оценивает воспринимаемое сходство между кадрами по яркости, контрасту и структуре на основе следующей формулы:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

μ_x, μ_y – средние значения из двух выборок;

σ_x^2, σ_y^2 – дисперсия значений из двух выборок;

σ_{xy} – ковариация выборок;

c_1, c_2 – переменные, зависящие от динамического диапазона пикселей.

- количества совпадений ключевых точек ORB (Oriented FAST and Rotated BRIEF) – фокусируется на ключевых особенностях, устойчивых к изменениям освещения или небольшим сдвигам. Реализация взята из библиотеки OpenCV представляющей широкий спектр алгоритмов для обработки изображений и видео, обнаружения функций, распознавания объектов и других задач компьютерного зрения. Библиотека разработана Intel, поддерживает различные языки программирования, включая C++, Python и Java и широко используется в робототехнике, наблюдении, для решения задач медицинской визуализации и дополненной реальности.

4.3 Выбор методов для решения задачи обнаружения ключевых кадров

В данном случае задачей было получение кадров, где происходят смены сцен. Мной были выбраны четыре метода для обнаружения смен сцен в видеофайлах.

1. Метод библиотеки PySceneDetect с использованием детектора ContentDetector
2. Метод сравнения гистограмм яркости в оттенках серого
3. Метод разности кадров, основанный на разнице пикселей между последовательными кадрами
4. Метод сравнения гистограммы цветов по каналам RGB.

Каждый метод анализирует видео, чтобы определить кадры, в которых происходит значительное изменение содержимого, в том числе переход между сценами. После выполнения каждого метода получаются списки номеров кадров, где эти смены происходят.

Рассмотрим работу каждого метода подробнее.

1. Метод использовал библиотеку PySceneDetect с детектором ContentDetector

Детектор ContentDetector анализирует изменения в содержимом кадров по интенсивности пикселей в цветовом пространстве HSV и определяет смены сцен, если изменения превышают заданный порог. Таким образом, метод зависит от выбора порога, при сравнении со значением которого будет считаться, что произошло изменение в сценах.

2. Метод сравнения гистограмм яркости в оттенках серого

Данный метод был подробно рассмотрен в главе 3.2. Выбранная метрика Хи-квадрат: она рассчитывает количественную оценку различий между гистограммами соседних кадров. Так же необходимо задание соответствующего порога: если разница между гистограммами по метрике хи-квадрат превышает порог, текущий кадр считается началом новой сцены. Для реализации данного и последующих методов была использована библиотека OpenCV

3. Метод разности кадров, основанный на разнице пикселей между последовательными кадрами

Обнаруживает смены сцен, вычисляя среднюю разницу в интенсивности пикселей между последовательными кадрами в оттенках серого. Если средняя разница превышает выбранный порог, то текущий кадр считается началом новой сцены. Для получения средней разницы необходимо вычислять абсолютную разницу пикселей между текущим и предыдущим кадром. Далее

просуммировать значения по всем пикселям и разделить на общее количество пикселей, чтобы получить среднюю разницу. В данном методе меньшие значения порога увеличивают чувствительность метода.

4. Метод сравнения гистограмм по каналам RGB

Данный метод обнаруживает смены сцен, сравнивая гистограммы цветов по каналам RGB последовательных кадров. В случае, если суммарная разница гистограмм по метрике хи-квадрат превышает порог, то текущий кадр считается началом новой сцены. Отличие от метода гистограмм интенсивности пикселей в том, что там работа происходит в оттенках серого, за счет чего происходит одно сравнение между пикселями двух кадров. При сравнении гистограмм для каждого цветового канала: синего, зеленого, красного, необходимо сравнить каждую пару канала по выбранной метрике и, например, просуммировать разницу, которую в дальнейшем сравнить с порогом чувствительности.

4.4 Выбор методов для решения задачи обнаружения изменений на кадрах

Были выбраны архитектуры универсального UNet и SiameseNet для сравнения их реализаций на данном наборе данных. Архитектура UNet представлена в главе 2 на рисунке 2.6.

Выбраны архитектуры типа UNet, так как в настоящее время они имеют наибольшую популярность, ввиду универсальности использования; кроме того, архитектуры просты в реализации и эффективно обучаются. Их код был полностью реализован с нуля, то есть нейронные сети не были предобучены.

На вход нейронным сетям для обучения подавались части изображений. Был выбран размер кусочков изображения (патчей) 96x96 пикселей.

Количество эпох — это количество раз, когда все данные обучения отображаются в сети во время обучения. На практике нейросетей обычно придерживаются такого правила: увеличивать количество эпох до тех пор, пока точность на валидационной выборке не начнет снижаться, даже если точность на обучаемой выборке увеличивается (то есть даже если начинает возникать переобучение). При решении данной задачи было избрано количество эпох 25, размер подвыборок — 16.

Размер подвыборок — это количество «пакетов», передаваемых в нейронную сеть, после чего происходит обновление параметров. Хорошим значением по умолчанию для размера подвыборок может быть 16, 32, однако пробуют и 64, 128 и т. д. NVIDIA рекомендует использовать числа, кратные 8, для тензорных ядер [18]. Гудфеллоу, Бенджио и Курвилльв своей книге

«Глубокое обучение» пишут, что при использовании графических процессоров, размеры подвыборок степени двойки обычно обеспечивает лучшее время выполнения [1].

Функция потерь – «NLLLoss». Функция потерь Negative Log Likelihood Loss (NLLLoss) часто используется в задачах, связанных с классификацией и сегментацией, включая применение UNet. NLLLoss эффективно справляется с задачей, так как она измеряет, насколько хорошо предсказанные вероятности изменения и не изменения пикселя соответствуют истинным меткам. Функция потерь вычисляется по следующей формуле:

$$NLL\ Loss = -\frac{1}{N} \sum_{i=1}^N \log(p_i)$$

Метрика качества для нейронной сети – это числовая мера, которая используется для оценки производительности нейронной сети в решении конкретной задачи машинного обучения. Метрики качества помогают качественно и количественно измерить эффективность работы модели и сравнить ее с другими моделями или подходами. В данном случае метрикой качества избрана «Ассигасу» – «Точность», она высчитывает процент правильных предсказаний модели от общего числа примеров.

В качестве алгоритма оптимизации функции потерь используется метод градиентного спуска, суть которого сводится к поэтапному изменению значений параметров функции на значение, пропорциональное градиенту в точке.

Была взята вариация градиентного спуска – оптимизация Адама — это метод стохастического градиентного спуска, основанный на адаптивной оценке моментов первого и второго порядка. Согласно Кингме и Лей Ба [8], этот метод «вычислительно эффективен, требует мало памяти, инвариантен к диагональному масштабированию градиентов и хорошо подходит для задач, больших по объему данных/параметров».

Шаг (скорость) обучения — это гиперпараметр, определяющий размер шага, с которым меняется значение градиента, который обновляет параметры сети во время обучения. Большая скорость обучения может значительно ускорить сходимости, но также может привести и к нестабильному и колеблющемуся обучению, если градиент не может попасть в минимум из-за большого шага. Небольшая скорость обучения может обеспечить стабильное и плавное обучение, одновременно с более медленной сходимостью. Необходимо выбирать то значение шага, которое обеспечивает наилучший компромисс между скоростью обучения и стабильностью. Был выбран стандартный шаг обучения, равный 0.001.

ГЛАВА 5 Сравнение результатов выбранных методов для решения поставленной задачи

5.1 Результаты решения задачи разделения на кадры

Для эксперимента взяты видео из набора данных с неразмеченными данными – кадрами, на которых произошли смены сцен. В данном случае невозможно сравнить по метрикам выбранные методы с истинным значением кадров, поэтому сравнение производилось отдельно для каждого видео.

Видеопоследовательности делились на кадры двумя способами. После деления соседние кадры попарно сравнивались между собой по выбранным метрикам качества. При интерпретации результатов необходимо понимать, что значат значения выбранных метрик.

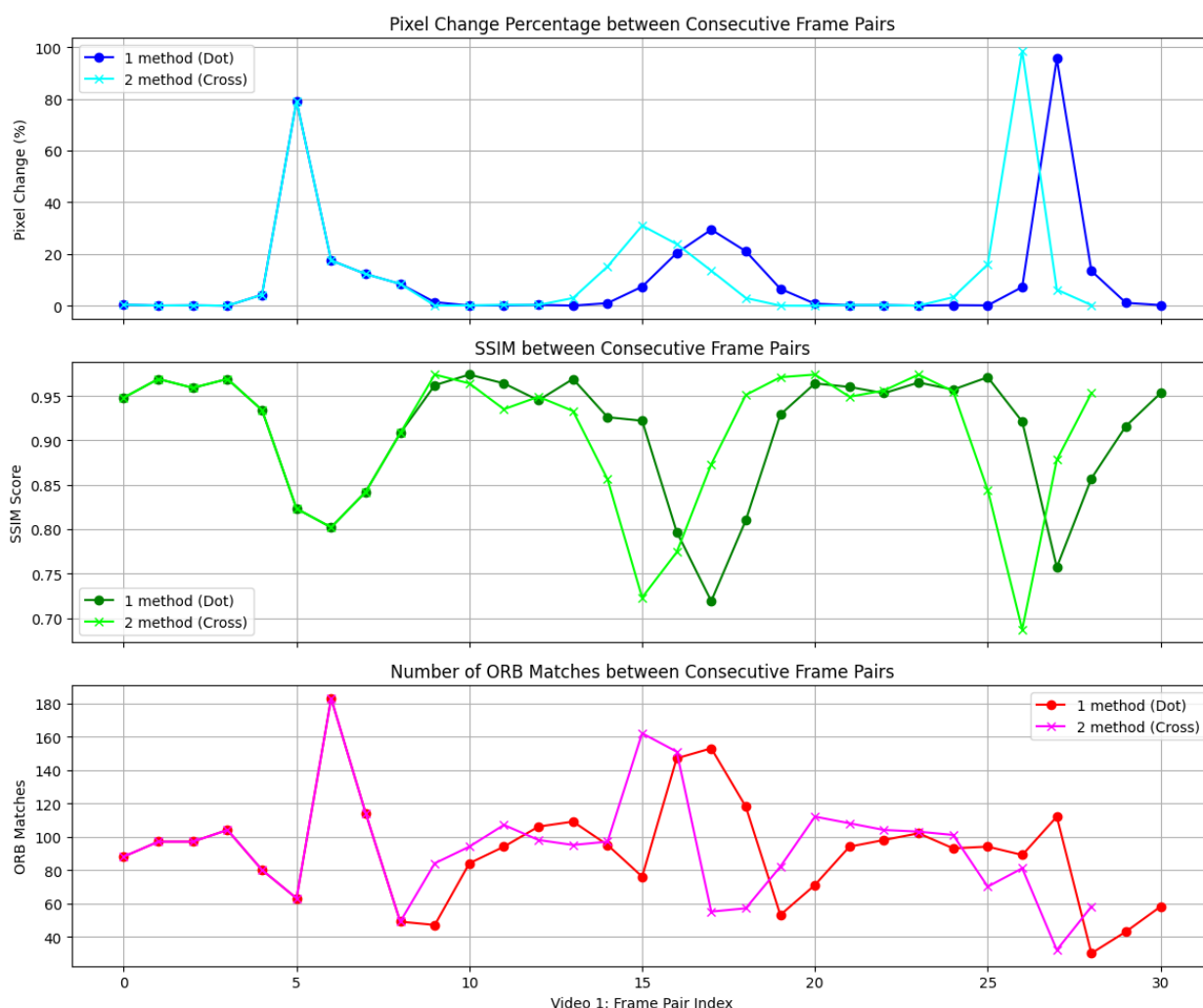


Рисунок 5.1 – Результаты исполнения методов для первого видео

На рисунках 5.1–5.3 показаны сравнения методов извлечения кадров: первый был делением видеопоследовательности на кадры с временным интервалом в 1 секунду, второй основывался на библиотеке PySceneDetect, для которой необходимо было задавать порог чувствительности. Порог подбирался на основании результатов, полученных ранее для соответствующих видео.

Метрика процент изменения пикселя выдает число от 0 до 100 %, которые отображает процент пикселей, которые значительно изменились между двумя последовательными кадрами. Более высокое значение означает более существенные визуальные различия, такие как движение, движение объекта или сцена.

SSIM (индекс структурного сходства) оценивает, как структурно схожи два кадра, его значение может быть от -1, то есть совершенно разнородные кадры, до 1 – идентичные кадры. Более высокие значения означают, что кадры более похожи.

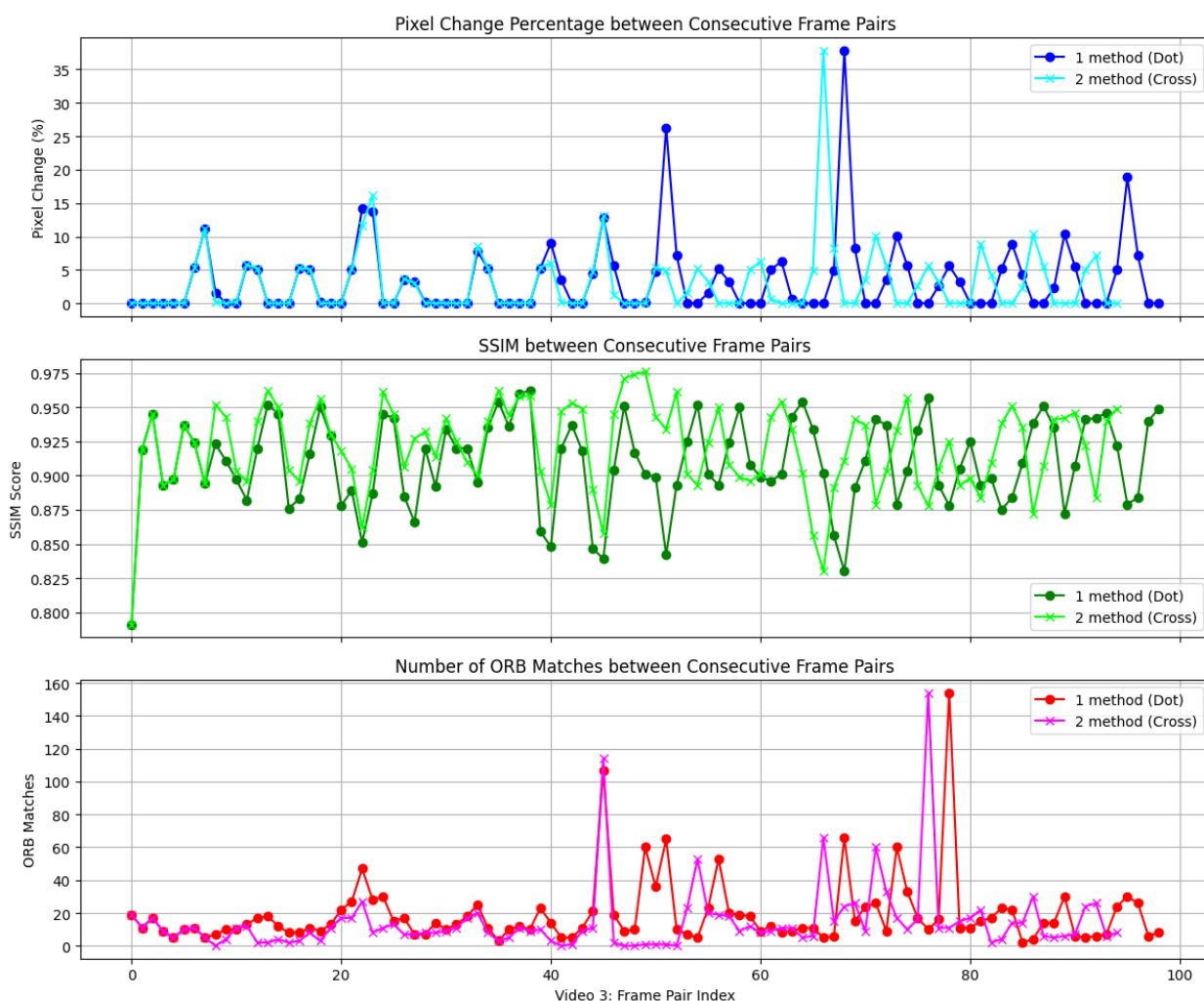


Рисунок 5.2 – Результаты исполнения методов для второго видео

Метрика количества совпадений ключевых точек ORB подсчитывает количество соответствующих особенностей (ключевых точек) между двумя кадрами. Большое число точек совпадений предполагает, что кадры имеют сходные элементы, даже если значения пикселей различаются.

По графикам, представленным на рисунках 5.1–5.3 заметна закономерность, что при подобранном значении порога чувствительности для PySceneDetect результаты будут очень похожи на простое деление видеопоследовательности на кадры с заданным интервалом, что, в свою очередь, более интуитивно ясно. Однако данная библиотека позволяет значительно уменьшить количество значимых кадров, сохраняя при этом относительное качество результата. На рисунке 5.3 видны более значительные различия, что указывает на большую точность метода, так как он выдает более уверенные результаты.

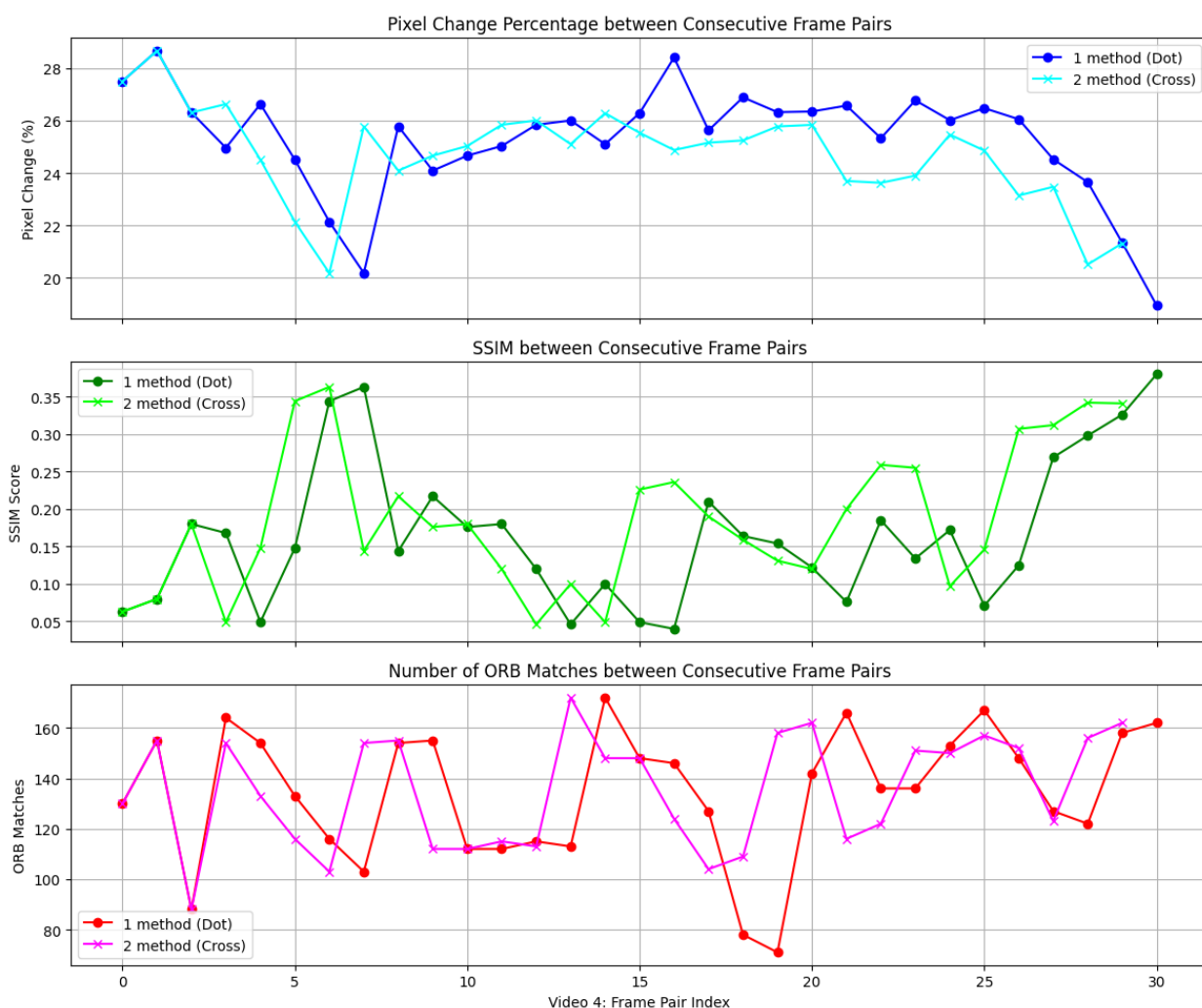


Рисунок 5.3 – Результаты исполнения методов для третьего видео

При сравнении метрик можно сделать вывод, что более интерпретируемые результаты показывают процент изменения пикселя между кадрами и метрика совпадения ключевых точек ORB. В свою очередь индекс структурного сходства не может дать точные результаты, что указывает на большую зависимость от зашумленности и неспособность выделять небольшие изменения на кадрах.

5.2 Результаты решения задачи обнаружения кадров со сменой сцен

Первым важным замечанием является то, что для каждого метода необходимо выбирать значение порога чувствительности. Поэтому для каждого метода были выбраны свои варианты значений порогов. Также они видоизменялись для каждого видео на основании их характеристик для получения более значимых результатов.

Приведем значения порогов, которые были использованы для первого видео с обнаружением машин.

Для метода с использованием PySceneDetect были выбраны значения порогов от 1 до 30 включая. Для метода сравнения гистограмм яркости в оттенках серого были выбраны пороги от 1000 до 70000 с шагом сначала тысяча, а потом десять тысяч, начиная с 10000.

Для метода разности кадров, основанный на разнице пикселей между последовательными кадрами, были выбраны пороги от 1 до 10 включая.

Для метода сравнения гистограмм цвета по каналам RGB были выбраны такие же пороги, как и для метода гистограмм в оттенках серого для более детального сравнения этих двух методов.

Результаты проведения экспериментов приведем в виде графиков, которые помогут наглядно понять разницу между методами и применением их для различных видео.

Также приведем в виде таблицы 5.1. время, которое было затрачено на выполнение методов для трех видео из набора данных, которые были рассмотрены подробно в разделе 4.1.

Таблица 5.1. Сравнение характеристик видео и времени, затраченного на их обработку

Видео	Затраченное время	Длительность видео	Качество видео
1	1:45	0:30	12.5 fps
2	39:16	1:38	59.94 fps
3	4:33	0:30	29.937 fps

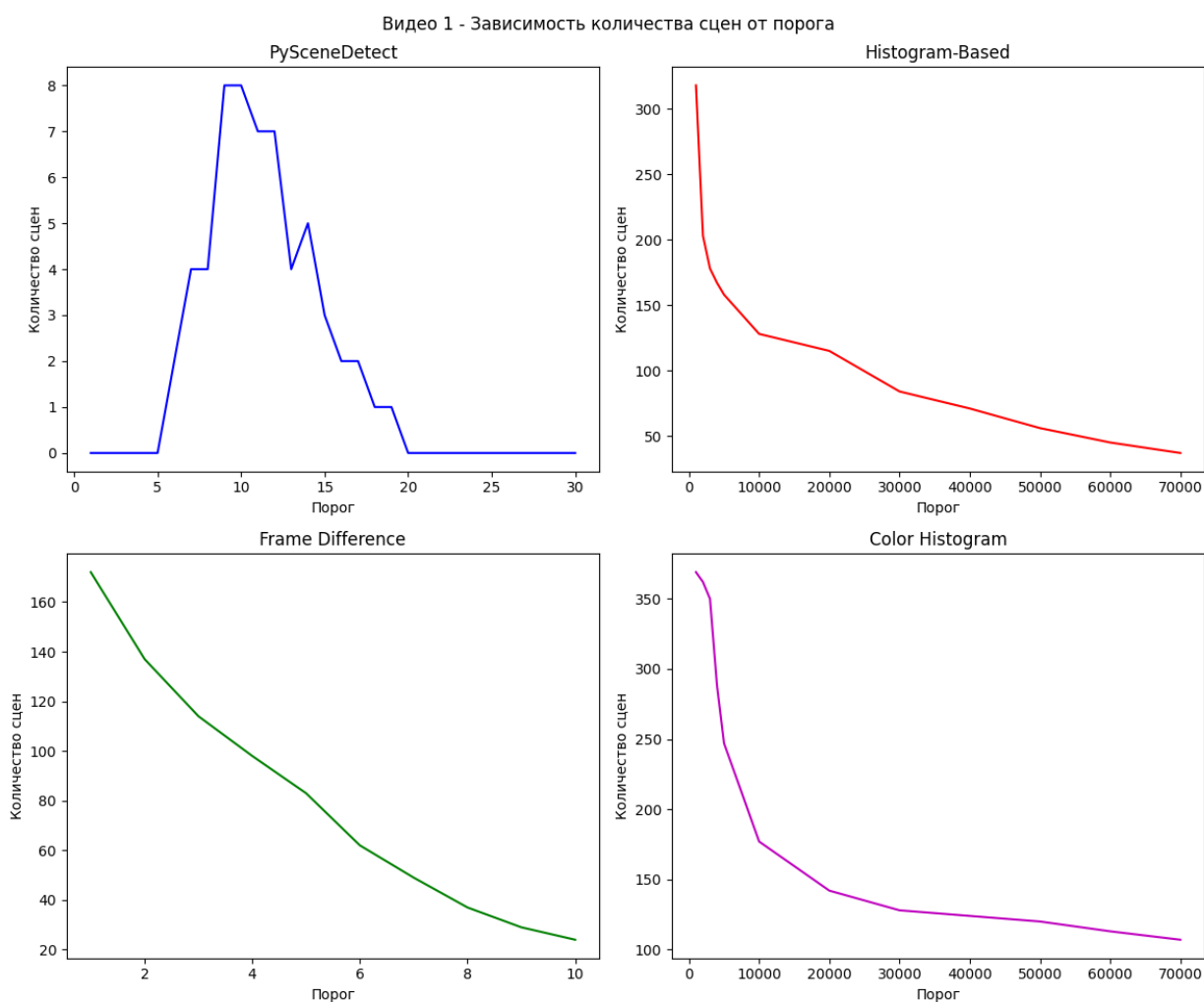


Рисунок 5.4 – Результаты исполнения методов для первого видео

Рассмотрим зависимость результатов выполнения методов от качества видео.

Высокое разрешение и частота кадров, как в видео 2 — разрешение 1920x1080, и частота кадров 59.94 fps, предоставляют больше деталей, что увеличивает количество обнаруженных изменений, особенно для разности кадров и гистограмм, это видно на рисунке 5.5.

Низкое разрешение и частота, как например в видео 1, где частота кадров всего 12.5 кадров в секунду, приводят к меньшему количеству изменений, так как меньше деталей для анализа, это находит отображение на Рисунке 5.4.

Время выполнения также увеличивается для видео с высоким качеством, например, видео 2 заняло 15 минут 16 секунд, в то время как видео 1 — только 45 секунд.

Рассмотрим зависимость результатов выполнения методов от продолжительности видео.

Длинные видео, такие как Видео 3 (1 минута 38 секунд), содержат больше кадров, что увеличивает количество потенциальных изменений, особенно для методов, чувствительных к движению, таких как разность кадров и гистограммные методы. Для библиотеки PySceneDetect, ориентированной на резкие смены сцен, длина видео оказывает меньшее влияние, если в видео мало смен сцен, как в Видео 1 (30 секунд).

Сравним каждый метод и определим их сильные стороны.

Методы библиотеки PySceneDetect лучше подходят для обнаружения резких смен сцен, как в фильмах, но менее чувствителен к движению внутри сцены. Например, в Видео 1 он обнаружил до 8 сцен, а в Видео 2 – до 107 при низких порогах.

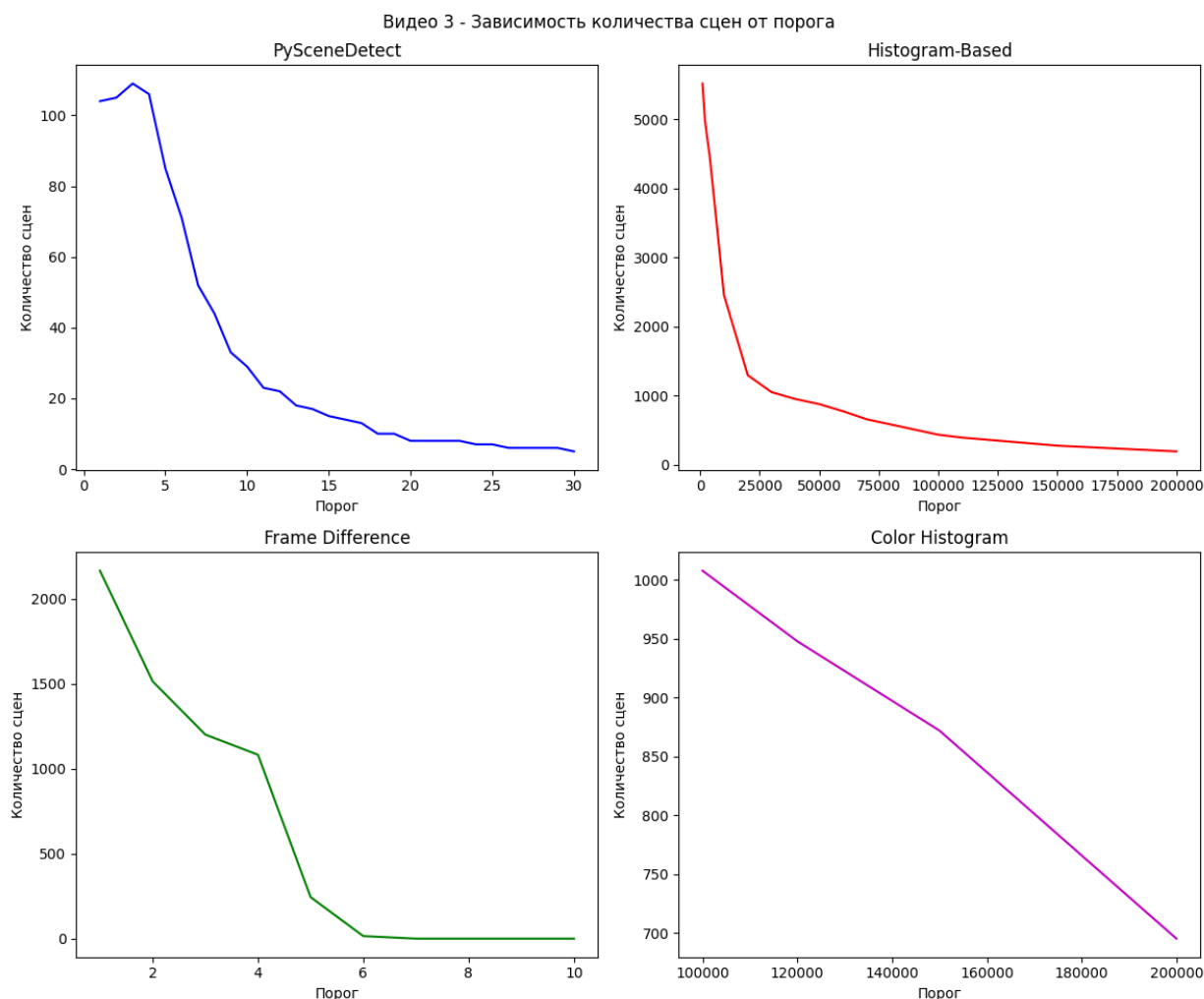


Рисунок 5.5 – Результаты исполнения методов для второго видео

Анализ гистограмм яркости в серых оттенках чувствителен к изменениям яркости, обнаруживая как резкие, так и постепенные изменения. Это находит отображение на рисунке 5.5 – в видео 2 обнаружено от 5821 до 121 кадра сцены при увеличении порога.

Метод разности кадров является самым простым методом, он очень чувствителен к движению, что делает его подходящей для детекции движения, но не для смен сцен. Так на рисунке 5.5 видно, что в видео 2 обнаружено до 2165 сцен при низких порогах, но это число очень быстро падает до 0.

Анализ гистограммы яркости в RGB действительно похож по результатам на анализ гистограммы яркости, но учитывает цвет, что полезно для видео с изменениями цветовой палитры. Особенно разница между методами заметна для видео 2 на рисунке 5.5.

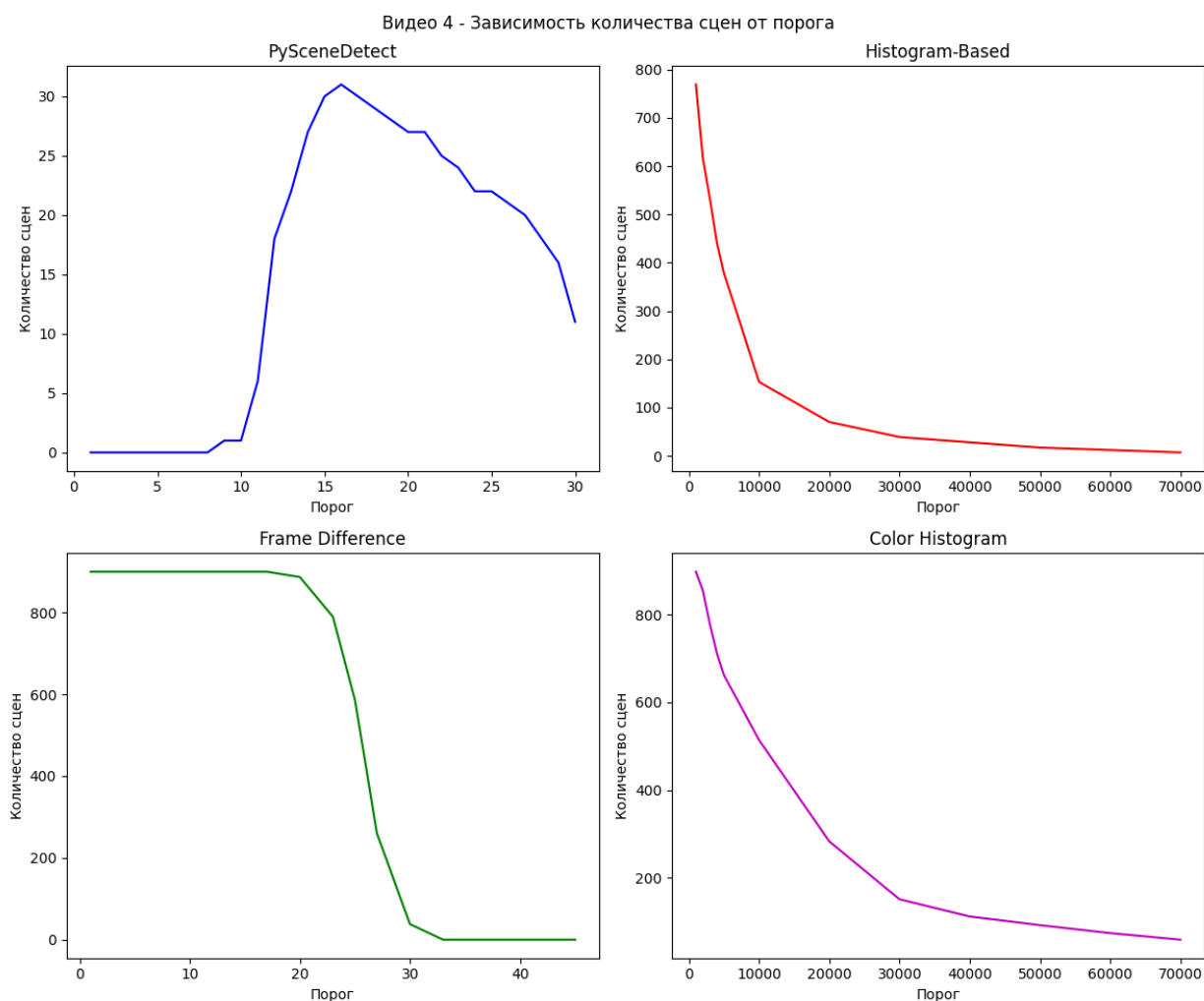


Рисунок 5.6 – Результаты исполнения методов для третьего видео

Ключевые выводы, полученные в ходе сравнения методов обнаружения изменений:

- Исследование показывает, что методы зависят от длины и качества видео, с большей чувствительностью к движению в длинных видео с высокой частотой кадров.
- Библиотека PySceneDetect лучше для обнаружения резких смен сцен, в то время как гистограммные методы подходят для постепенных изменений.
- Метод разности кадров сильно реагирует на движение, что может быть избыточным для задач смены сцен.
- Качество видео, включая разрешение и частоту кадров, влияет на точность и время выполнения методов.

5.3 Результаты решения задачи обнаружения изменений на кадрах

Были реализованы две нейронные сети Unet и SiameseNet для их сравнения. Итоги обучения оцениваются графически. Сравнение качества модели производится на тестовой и обучающей выборках. Для этого набор данных из 17 видео был разделен на две неравные части: 10 видео для обучающей выборки и 7 для тестовой.

Сравним результаты функции потерь для построенных нейронных сетей. На 25 эпохе на обучающей выборке SiameseNet показала ошибку равную 0.131, на тестовой – 0.282. Для UNet – ошибка на обучающей выборке составила 0.094, на тестовой – 0.260. Графические результаты на протяжении всех эпох представлены на рисунке 5.7.

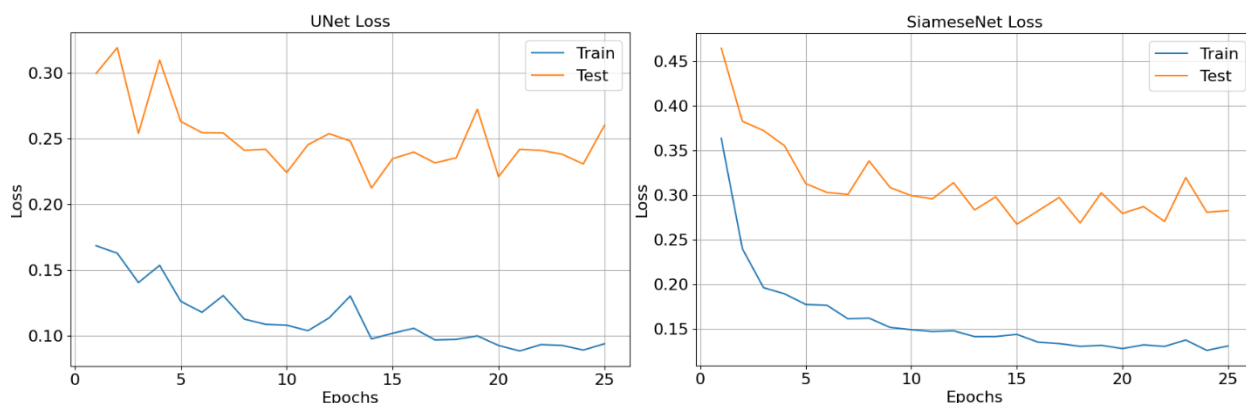


Рисунок 5.7 – Сравнение результатов функции потерь для двух построенных нейронных сетей

Можно сделать вывод, что для UNet в среднем результат лучше, кроме того по графику видна характерная «зубчатость», характеризующая, что при

дальнейшем обучении нейронной сети могут быть получены более эффективные результаты.

Для сравнения точности были использованы три вариации точности: общая точность, точность определения наличия изменений и точность определения отсутствия изменений. Ввиду большого количества характеристик они представлены в таблице 5.2.

Таблица 5.2. Результаты построенных нейронных сетей для метрики Точности

	SiameseNet	UNet
Train Accuracy	96.32328075	97.52564873
Train No Change Accuracy	96.84914398	97.92961121
Train Change Accuracy	71.14205933	78.18170166
Test Accuracy	92.7114362	92.87098536
Test No Change Accuracy	94.32310486	93.97841644
Test Change Accuracy	62.28187561	71.96183014

Заметно, что UNet быстрее выучился на обучающей выборке, на тестовой выборке их результаты похожи, для определения отсутствия изменений немного лучше работает SiameseNet, для обнаружения изменений – UNet. Графическое сравнение результатов метрики качества на двух построенных нейронных сетях представлено на рисунке 5.8.

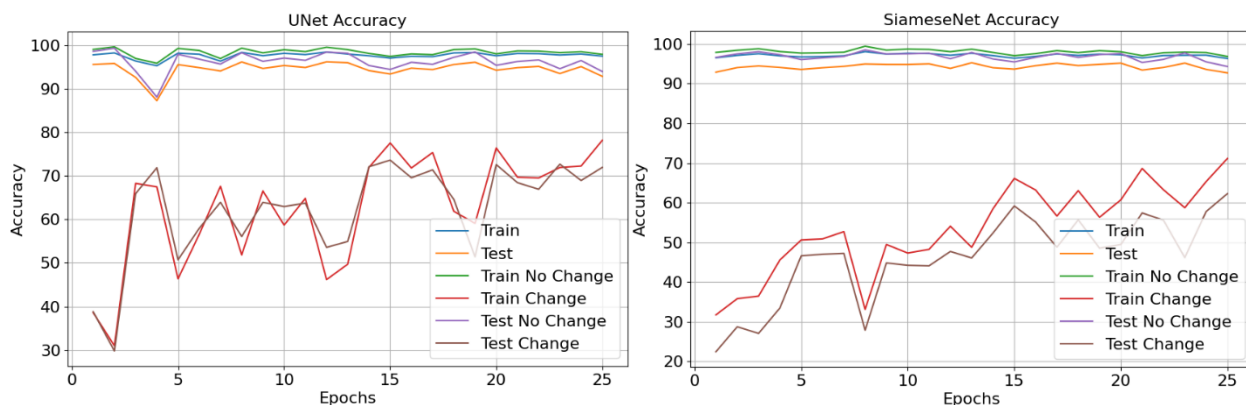


Рисунок 5.8 – Сравнение результатов метрики качества на двух построенных нейронных сетях

Таким образом, обе построенные нейронные сети показали хорошие результаты. За процесс обучения стала видна тенденция, что две нейронные сети ведут себя по-разному: одна больше подходит для обнаружения отсутствия изменений, в то время как другая – для наличия изменений на изображении.

ЗАКЛЮЧЕНИЕ

Чтобы правильно использовать огромный объем данных, получаемых современными видеокамерами, радарами, спутниками дистанционного зондирования, необходимо разрабатывать эффективные неконтролируемые и автоматические методы обнаружения изменений. В связи с быстро растущим интересом к этой теме и большим количеством новых моделей, представленных в последние годы, в данной работе представлены теоретические и практические сравнительные характеристики для различных методов обнаружения изменений сцен и объектов видеопоследовательностей.

Результаты полученного исследования позволяют создать более глубокое понимание процессов обнаружения изменений на видеопоследовательностях. Для различных по структуре видеоданных рассмотрены и описаны соответствующие наиболее подходящие алгоритмы и методы. Проанализированы основные подходы к обработке видео данных, включая методы предобработки, сегментации, сравнения кадров и выявления значимых изменений с учётом шумов и динамических сцен.

Наибольший интерес представляют несжатые видеопоследовательности, работа с ними в офлайн и онлайн режимах. Классические методы проверены годами и не теряют в своей действенности, однако требуют более точной и детальной настройки под ситуации, в которых ведётся видеонаблюдение.

После проведённых экспериментов по делению видеопоследовательностей на кадры на выбранном наборе данных были выявлены на практике главные недостатки методов и предложены варианты их избежания. В том числе была подтверждена работоспособность выбранных алгоритмов и их пригодность для использования в задачах видеонаблюдения, мониторинга окружающей среды и других прикладных областях.

Также в проделанной работе выделены рекомендации по выбору методов для обнаружения изменения в сценах и объектах на кадрах. Это включает в себя рекомендации по использованию глубокого обучения – нейронных сетей для сложных задач, где необходима высокая точность и устойчивость к шуму, а также более простых подходов для случаев, когда ресурсоемкость не имеет первостепенное значение.

Значимость работы заключается в возможности автоматизации процесса анализа видеоданных, что позволяет своевременно выявлять важные события, сокращать объёмы обрабатываемой информации и улучшать качество принимаемых решений.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.
2. A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei and S. Savarese, Social LSTM: Human Trajectory Prediction in Crowded Spaces // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). –2016. – P. 961–971.
3. Bhandari, A.; Kumar, A.; Singh, G. Feature Extraction using Normalized Difference Vegetation Index (NDVI): A Case Study of Jabalpur City / A. Bhandari, A. Kumar, G. Singh // Procedia Technol. – 2012. – Vol. 6. – P. 612–621.
4. Bovolo, F.; Bruzzone, L. A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain / F. Bovolo, L. Bruzzone // IEEE Trans. Geosci. Remote Sens. – 2007. – Vol. 45, No. 1. – P. 218–236.
5. Celik, T. Unsupervised change detection in satellite images using principal component analysis and K-means clustering / T. Celik // IEEE Geosci. Remote Sens. Lett. – 2009. – Vol. 6, No. 4. – P. 772–776.
6. Daudt, C.R. Fully convolutional siamese networks for change detection / C.R. Daudt, B. Le Saux, A. Boulch // Proceedings of the International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018. – P. 4063–4067.
7. Deldari, S. Time Series Change Point Detection with Self-Supervised Contrastive Predictive Coding [Электронный ресурс] / Deldari, S., Smith, D. V., Xue, H., Salim, F. D // Proceedings of The Web Conference 2021. – Association for Computing Machinery, 2021. – Режим доступа: <https://doi.org/10.1145/3442381.3449903>. – Дата доступа: [05.03.2021]
8. Diederik P. Kingma Adam: a method for stochastic optimization / Diederik P. Kingma, Jimmy Lei Ba – arXiv.org e-Print archive, 2015 – 15 p.
9. Du, B. Unsupervised Deep Slow Feature Analysis for Change Detection in Multi-Temporal Remote Sensing Images / B. Du, L. Ru, C. Wu, L. Zhang // IEEE Trans. Geosci. Remote Sens. – 2019. – Vol. 57, No. 12. – P. 9976–9992.
10. Fernando, W.A.C. Sudden scene change detection in compressed video using interpolated macroblocks in B-frames / W. A. C. Fernando // Multimed Tools Appl. – 2006. – No. 28. – P. 301–320. – DOI: 10.1007/s11042-006-7716-7.

11. Kåsen, I. Band Selection for Hyperspectral Target Detection Based on a Multinormal Mixture Anomaly Detection Algorithm / I. Kåsen, A. Rødningsby, T.V. Haavardsholm, T. Skauli // SPIE 2008. – Vol. 6966. – P. 53–58.
12. Li, Z. An Effective and Fast Scene Change Detection Algorithm for MPEG Compressed Videos / Li, Z., Jiang, J., Xiao, G., Fang, H. // Campilho, A., Kamel, M.S. (eds) Image Analysis and Recognition. ICIAR 2006. – Lecture Notes in Computer Science, vol. 4141. – Springer, Berlin, Heidelberg, 2006. – DOI: 10.1007/11867586_20.
13. Liu, Q. A DCT-Domain Approach to Image Change Detection and Its Application to Patient Video Monitoring / Liu, Q., Sclabassi, R. J., Sun, M. // 2005 IEEE 7th Workshop on Multimedia Signal Processing. – Shanghai, China, 2005. – C. 1-4. – DOI: 10.1109/MMSP.2005.248623.
14. Liu, S. Review of Change Detection in Multitemporal Hyperspectral Images: Current Techniques, Applications, and Challenges / S. Liu, D. Marinelli, L. Bruzzone, F. Bovolo // IEEE Geosci. Remote Sens. Mag. – 2019. – Vol. 7. – P. 140–158.
15. Lucas, B. D., Kanade, T. An iterative image registration technique with an application to stereo vision // Proceedings of the International Joint Conference on Artificial Intelligence. – 1981. – C. 674–679.
16. Tezcan M. O. BSUV-Net 2.0: Spatio-Temporal Data Augmentations for Video-Agnostic Supervised Background Subtraction / Tezcan M. O., P. Ishwar, J. Konrad // IEEE Access – 2022. – Vol. 9. – P. 53849–53860.
17. Nielsen, A.A. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data / A.A. Nielsen // IEEE Trans. Image Process. – 2007. – Vol. 16, No. 2. – P. 463–478.
18. NVIDIA cuDNN Developer Guide [Electronic resource] – NVIDIA Corporation. – Mode of access: <https://docs.nvidia.com/deeplearning/cudnn/pdf/cuDNN-Developer-Guide.pdf>. – Date of access: November 2023.
19. Padilla, R. Change detection in moving-camera videos with limited samples using twin-CNN features and learnable morphological operations / Padilla, R., da Silva, A. F., da Silva, E. A. B., Netto, S. L. // Signal Processing: Image Communication. – 2023. – T. 115. – C. 116969. – DOI: 10.1016/j.image.2023.116969.
20. Ronneberger, O. U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // Medical Image Computing and Computer-Assisted Intervention (MICCAI); Springer: Berlin/Heidelberg, Germany, 2015. – Vol. 9351. – P. 234–241.

21. Ronneberger, O. U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2015. – P. 234–241.
22. Su, H. Hyperspectral Anomaly Detection: A survey / H. Su, Z. Wu, H. Zhang, Q. Du // IEEE Geosci. Remote Sens. Mag. – 2022. – Vol. 10. – P. 64–90.
23. Wang, L., Xiong, Y., Wang, Z., & Qiao, Y. Towards Good Practices for Very Deep Two-Stream ConvNets / Wang, L., Xiong, Y., Wang, Z., Qiao, Y. // DOI: 10.48550/arXiv.1507.02159.
24. Wu, C. Slow feature analysis for change detection in multispectral imagery / C. Wu, B. Du, L. Zhang // IEEE Trans. Geosci. Remote Sens. – 2014. – Vol. 52, No. 5. – P. 2858–2874.
25. Wu, C. Kernel Slow Feature Analysis for Scene Change Detection / C. Wu, L. Zhang, B. Du // IEEE Trans. Geosci. Remote Sens. – 2017. – Vol. 55, No. 4. – P. 2367–2384.
26. Zhang, L. Automatic radiometric normalization for multitemporal remote sensing imagery with iterative slow feature analysis / L. Zhang, C. Wu, B. Du // IEEE Trans. Geosci. Remote Sens. – 2014. – Vol. 52, No. 10. – P. 6141–6155.