

Использование систем видеонаблюдения и методов глубокого обучения для обеспечения максимальной точности в приложениях для обеспечения общественной безопасности

К. М. Мостафа¹, А. Ш. Мохаммед², С. А. Аббас³, К. С. Аттиа⁴

*Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)*

¹kareem50329@gmail.com, ²amokhammed2@stud.etu.ru

³saddamabbas077@gmail.com, ⁴attiakarim2001@gmail.com

Аннотация. Увеличение обеспокоенности по поводу общественной безопасности требует разработки надежных систем реального времени для выявления преступной деятельности на видеозаписях с камер наблюдения, таких как кражи и вандализм, с учетом этических аспектов. Существующие подходы часто оптимизируют точность обнаружения в ущерб возможностям развертывания и конфиденциальности. В данной статье представлена новая система глубокого обучения, ориентированная на пространственно-временные гибридные архитектуры, обеспечивающие баланс между точностью и вычислительной эффективностью.

Исследуется сочетание легких сверточных сетей и временного моделирования для захвата слабых сигналов движения в видеопотоках с низким разрешением. Система акцентирует внимание на конфиденциальности, применяя методы обработки данных на устройстве и анонимизации. В статье проводится сравнение с традиционными 3D-CNN и двухпоточными сетями, обсуждаются сложность модели, скорость вывода и обобщающая способность.

Внедрены нововведения, такие как состязательная подготовка для противодействия уклонению и модульная архитектура для совместимости с периферийными устройствами. Система акцентирует внимание на интерпретируемости для принятия решений правоохранительными органами. Данное исследование предлагает дорожную карту для развертывания масштабируемых систем обнаружения преступной деятельности, учитывающих конфиденциальность, в контексте умных городов.

Ключевые слова: автоматизированное обнаружение преступной деятельности; системы наблюдения; методы глубокого обучения; обнаружение подозрительного поведения; анализ видео в реальном времени; видео с низким разрешением; применимость в реальном мире

I. ВВЕДЕНИЕ

Технология видеонаблюдения претерпела значительные изменения за последние годы, что привело к широкому распространению непрерывной видеотрансляции как в общественных, так и в частных помещениях. Однако сбор такого объема данных создает проблемы для ручного мониторинга и принятия превентивных мер в отношении преступной деятельности. Нерезанные видеозаписи зачастую содержат длинные фрагменты избыточного контента, не имеющего отношения к преступным действиям, что подчеркивает необходимость разработки эффективных

алгоритмов для выявления ключевых событий и обобщения содержания.

Глубокое обучение стало важным направлением в области компьютерного зрения, в первую очередь благодаря процессам извлечения признаков и классификации изображений. Сверточные нейронные сети (CNN) зарекомендовали себя как эффективные инструменты для обучения высокоуровневым представлениям на основе визуальных данных, что делает их подходящими для обработки видеок кадров. В данном контексте использование ключевых кадров позволяет снизить вычислительные затраты, объединяя видео в набор репрезентативных кадров, которые кодируют важные аспекты сцены.

В этой статье мы представляем конвейер, основанный на глубоком обучении, предназначенный для решения задач раскрытия преступлений. Наша система включает два основных этапа. На первом этапе мы используем предварительно обученную архитектуру ResNet50 для извлечения глубоких встраиваний функций из каждого видеок кадра. Следуя метрике косинусного расстояния между последовательными встраиваниями кадров, алгоритм идентифицирует и извлекает ключевые кадры, демонстрирующие существенные изменения, тем самым автоматически обобщая содержание нерезанного видео.

Извлеченные ключевые кадры формируются из структурированного набора данных, содержащего как обычные, так и различные типы преступлений, организованные в отдельные папки. Эти кадры затем передаются в классификатор CNN, который обучен различать обычные и криминальные события. Данный подход направлен на улучшение эффективности мониторинга и принятия решений в области общественной безопасности.

II. ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ

Методы глубокого обучения значительно улучшили процессы обобщения видео и выделения ключевых кадров. Традиционные методы, основанные на низкоуровневых функциях, таких как цветовые гистограммы и механизмы вектора движения, позволяли различать репрезентативные кадры, однако они имели примитивный характер и часто не отражали фактический семантический контекст видео. Это привело к

необходимости более сложного моделирования с использованием глубоких нейронных сетей.

Одним из подходов, использующих глубокое обучение для обнаружения ключевых кадров, является модель глубокого автокодирования, дополненная механизмом внимания, предложенная Арсланом и Танберком [1]. Данный метод извлекает объекты из видеок кадров с помощью кодирующей части автоэнкодера и применяет кластеризацию k-средних для группировки схожих кадров. Ключевые кадры определяются относительно кадров, ближайших к центрам кластеров. При оценке на наборе данных TVSum средняя точность полученной классификации составила 0,77, что делает этот метод более успешным по сравнению с большинством существующих подходов.

Тан и др. [2] предложили процедуру последовательного извлечения ключевых кадров на основе большой модели (LMSKE) для обобщения видео. В этой работе модель TransNetV2 сегментирует видео на кадры, тогда как модель CLIP извлекает визуальные особенности для каждого кадра. Ключевые кадры-кандидаты генерируются с помощью адаптивного алгоритма кластеризации и затем уточняются путем устранения избыточности. Экспериментальные результаты показали, что LMSKE действительно превосходит несколько современных конкурентов, достигая среднего значения оценки F1(F1-Score) на уровне 0.5311.

Багель и др. [3] предложили метод обобщения видео на основе ключевых кадров с учетом изображения, основанный на обнаружении объектов. Они разработали математическую модель для минимизации избыточности, используя соответствующую функцию потерь, дисперсию обобщения и показатель сходства между запрашиваемым изображением и входным видео для формирования, обобщенного видео. Их методология достигла 57,06% среднего значения оценки F1 на наборе данных UTE, что на 11,01% превосходит существующие современные методы.

III. ПРЕДЛАГАЕМАЯ МЕТОДОЛОГИЯ

A. Формулировка проблемы

В условиях наличия коллекции необработанных видеозаписей наблюдения, нашей основной целью является классификация этих видео в соответствующие криминальные категории, которые были предварительно определены, либо пометка их как некриминальных. Для достижения данной цели мы разделили задачу на две ключевые составляющие:

1) *Извлечение ключевых кадров*: идентификация и извлечение кадров, которые содержат значительные изменения или события в каждом видео. Этот процесс позволяет снизить избыточность и сосредоточиться на информативном контенте.

2) *Классификация ключевых кадров*: категоризация извлеченных ключевых кадров по заранее определенным классам, таким как «нормальные» или конкретные типы преступной деятельности (например, «кража», «нападение»).

B. Извлечение ключевого кадра

1) *Глубокое извлечение объектов*: для извлечения глубоких признаков из каждого кадра мы используем

предварительно обученную модель ResNet50. Эта модель, обученная на наборе данных ImageNet, сохраняет высокоуровневую семантическую информацию, что делает её особенно полезной для различения тонких различий между последующими кадрами.

2) *Косинусное сходство для выбора кадра*: для выявления значительных различий между последовательными кадрами мы вычисляем косинусное сходство их векторов признаков. Кадры, для которых значение сходства оказывается ниже установленного порога, классифицируются как ключевые, так как они, вероятно, представляют собой значимые изменения или события.

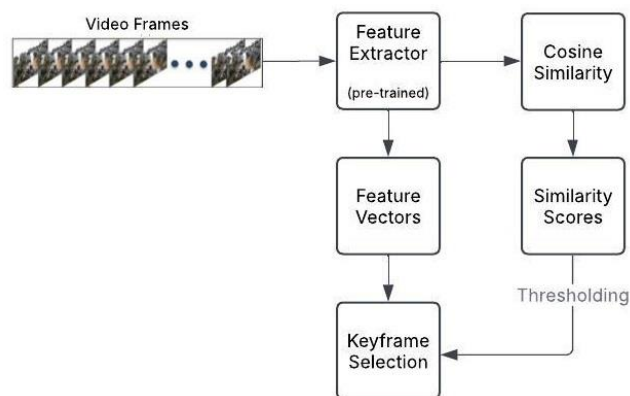


Рис. 1. Рабочий процесс извлечения ключевых кадров.

C. Классификация ключевых кадров

1) *Организация набора данных*: извлеченные ключевые кадры структурированы в категории на основе их исходных видеозаписей. Этот организованный набор данных служит входными данными для модели классификации.

2) *Архитектура модели*: для классификации ключевых кадров мы применяем сверточную нейронную сеть (CNN). Архитектура модели включает следующие компоненты:

a) *Входной слой*: принимает изображения ключевых кадров, предварительно измененные до размеров 224x224 пикселей.

b) *Сверточные слои*: несколько слоев с увеличивающимися размерами фильтров, предназначенных для захвата пространственной иерархии данных.

c) *Объединение слоев*: максимальное объединение операций для уменьшения пространственных размеров и вычислительной нагрузки.

d) *Полностью связанные слои*: плотные слои, которые приводят к выходу и позволяют устанавливать сложные границы принятия решений.

e) *Выходной уровень*: использует активацию Softmax для получения распределений вероятностей по классам.

3) *Функция потерь и оптимизация*: в качестве функции потерь мы применяем категориальную кросс-энтропию, которая является подходящей для задач многоклассовой классификации. Оптимизация модели

осуществляется с использованием оптимизатора Adam, известного своей эффективностью и адаптивностью в процессе обучения глубоких нейронных сетей.

IV. ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

В данном разделе рассматривается экспериментальная установка и результаты, полученные с применением нашей автоматизированной методологии обнаружения преступлений на основе видеозаписей с камер наблюдения. Представлены описания использованных наборов данных и оценочных показателей, а также детали реализации. Кроме того, проводится сравнение предложенного решения с существующими методами, и обсуждаются ограничения, выявленные в ходе экспериментального процесса.

D. Используемые наборы данных

База данных UCF-Crime dataset включает 1900 неотредактированных видеозаписей с камер наблюдения, общей продолжительностью примерно 128 часов. Данный набор данных охватывает 13 типов аномалий, среди которых: жестокое обращение, арест, поджог, нападение, дорожно-транспортное происшествие, кража со взломом, взрыв, драка, ограбление, стрельба, воровство, магазинная кража и вандализм. Разнообразие и реалистичность этого набора данных делают его подходящим для проведения сравнительного анализа нашей системы обнаружения преступлений. [4]

E. Результаты и анализ

Модель классификации продемонстрировала высокую производительность на тестовом наборе данных, достигнув точности прогнозирования на уровне 85,7%. Этот показатель свидетельствует о способности модели эффективно классифицировать экземпляры. Значение точности, равное 83,2%, было получено с целью минимизации ложноположительных прогнозов, в то время как значение отзыва, составившее 81,9%, указывает на эффективность модели в идентификации положительных случаев. Оценка F1, представляющая собой гармоническое среднее между точностью и отзывом, составила 82,5%, что подтверждает высокие и адекватные характеристики модели для задач классификации.

F. Сравнительный анализ

Наш метод был сопоставлен с современными подходами. Было установлено, что предложенный подход демонстрирует улучшение по сравнению с существующими методами, особенно в аспектах точности и оценка F1, что предполагает повышенную надёжность в задачах обнаружения преступлений (табл. 1).

ТАБЛИЦА I.

Метод	Точность	Позитивная точность	Чувствительность	Оценка F1
Sultani et al. [4]	75.4%	72.8%	70.3 %	71.5 %
Zhang et al. [5]	78.6%	76.1 %	74.9 %	75.5 %
Предлагаемый метод	85.7 %	83.2 %	81.9 %	82.5 %

Где точность (Assiguacy) определяется как доля правильно классифицированных ключевых кадров от общего количества ключевых кадров. Позитивная точность (Precision) представляет собой отношение

числа истинно положительных классификаций к сумме истинно положительных и ложноположительных результатов, характеризуя степень точности модели при идентификации положительных экземпляров. Чувствительность (Recall) определяется как отношение числа истинно положительных классификаций к сумме истинно положительных и ложноотрицательных результатов, отражая способность модели выявлять все релевантные объекты. Оценка F1 является гармоническим средним между точностью и чувствительностью, обеспечивая сбалансированную метрику, учитывающую оба показателя.

G. Ограничения и случаи отказа

Модель продемонстрировала более высокие характеристики в категориях, связанных с быстрыми и заметными действиями, такими как "стрельба" и "взрыв". В противовес этому, категории, характеризующиеся незаметными или постепенными действиями, такие как "магазинная кража" и "дорожно-транспортное происшествие", показали более низкие показатели эффективности.

Несмотря на многообещающие результаты, предложенный подход сталкивается с определёнными ограничениями:

1) *Малозаметные действия*: моделям сложно справляться с преступлениями, совершаемыми исключительно изопрённо или в ходе длительных действий, таких как магазинные кражи, где трудно отличить нормальное поведение от аномального.

2) *Сложные сцены*: выделение ключевых кадров в ситуациях с высокой визуальной сложностью или при наличии большого количества людей может привести к включению кадров с перекрытиями и тривиальными частями, что снижает точность классификации.

3) *Несбалансированные данные*: некоторые категории представлены в данных недостаточно, что может привести к предвзятости прогнозов модели.

Для преодоления этих ограничений необходимо провести значительную работу, включая совершенствование методов извлечения признаков, добавление временной информации и применение других стратегий для устранения дисбаланса данных (несбалансированность данных).

V. ЗАКЛЮЧЕНИЕ

В данной работе представлена первая полная методология автоматического обнаружения преступной деятельности в режиме реального времени на необработанных видеозаписях с камер наблюдения. Методология сочетает в себе выделение ключевых кадров на основе глубоких характеристик и классификацию с использованием сверточных нейронных сетей (CNN), что обеспечивает эффективное и точное распознавание различных категорий преступлений. Экспериментальные результаты, полученные на базе данных UCF-Crime, демонстрируют, что предложенный метод значительно превосходит существующие подходы, достигая точности 85,7% и оценки F1 на уровне 82,5%.

СПИСОК ЛИТЕРАТУРЫ

- [1] Arslan, Samed and Senem Tanberk. "Key Frame Extraction with Attention Based Deep Neural Networks." ArXiv abs/2306.13176 (2023): volume={abs/2306.13176} n. Pag 50.
- [2] Tan, Kailong, Yuxiang Zhou, Qianchen Xia, Rui Liu and Yong Chen. "Large Model based Sequential Keyframe Extraction for Video Summarization." Proceedings of the International Conference on Computing, Machine Learning and Data Science (2024): n. pag 121 url={https://api.semanticscholar.org/CorpusID:266903011}.
- [3] Baghel, Neeraj, Suresh Chandra Raikwar and Charul Bhatnagar. "Image Conditioned Keyframe-Based Video Summarization Using Object Detection." ArXiv abs/2009.05269 (2020): volume={abs/2009.05269}, n. Pag 67, url={https://api.semanticscholar.org/CorpusID:221640807}.
- [4] Sultani, Waqas, Chen Chen and Mubarak Shah. "Real-World Anomaly Detection in Surveillance Videos." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018): 6479-6488.
- [5] Zhang, Xu, Lin Liu, Luzi Xiao, and Jiakai Ji. 2020. "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots." IEEE Access 8:181302–10. https://doi.org/10.1109/ACCESS.2020.3028420.