

Разработка и исследование программного обеспечения для обнаружения и отслеживания быстро движущихся объектов

Е.Н. Антонянц, Д.А. Гордиенко, Д.В. Савочка, А.М. Хусаинов

Аннотация — Статья посвящена разработке и исследованию программного обеспечения для обнаружения и отслеживания быстро движущихся объектов в видеопоследовательности. В работе рассматриваются основные задачи обнаружения и отслеживания, представлен обзор существующих архитектур и алгоритмов, применяемых для повышения точности и скорости обработки данных. Особое внимание уделено применению сиамской нейронной сети для оценки визуального сходства объектов и валидации детекций в процессе отслеживания.

В рамках исследования разработано программное обеспечение, реализующее модули обнаружения, валидации детекций с помощью сиамской нейронной сети и отслеживания. Проведено экспериментальное тестирование и сравнительный анализ предложенного подхода с традиционными методами, основанными на ручных признаках. Оценка качества распознавания выполнялась с использованием метрик точности и полноты, а производительность многообъектного отслеживания оценивалась по метрикам MOTA и IDF1.

Результаты экспериментов подтверждают превосходство нейросетевого подхода: точность обнаружения объектов достигла 0,84 при полноте 0,87. Полученные результаты показывают более высокую устойчивость к вариациям внешнего вида объектов по сравнению с классическими методами. Применение сиамской сети позволило снизить количество потерь объектов и ошибок переключения идентификаторов, что выражается в значениях MOTA (0,64) и IDF1 (0,73).

Разработанный подход демонстрирует высокую эффективность при решении задач обнаружения и отслеживания быстро движущихся объектов и подтверждает перспективность использования сиамских нейронных сетей для дальнейшего развития интеллектуальных систем компьютерного зрения.

Ключевые слова — обнаружение объектов, отслеживание объектов, многообъектное отслеживание, быстро движущиеся объекты, сиамская нейронная сеть.

I. ВВЕДЕНИЕ

Обнаружение и отслеживание объектов в видеопотоке

Статья получена 23 февраля 2026 г.

Антонянц Егор Николаевич, Новосибирский государственный технический университет, Россия (e-mail: bax201438@gmail.com).

Гордиенко Дмитрий Андреевич, Новосибирский государственный технический университет, Россия (e-mail: dima2323fr3@gmail.com).

Савочка Денис Владимирович, Новосибирский государственный технический университет, Россия (e-mail: densavochka@mail.ru).

Хусаинов Артур Мансурович, Новосибирский государственный технический университет, Россия (e-mail: xusainow.artur2018@yandex.ru).

является ключевой задачей для систем мониторинга, безопасности промышленных объектов, аппаратно-программных комплексов фотовидеофиксации и других прикладных систем, требующих своевременное распознавание быстро движущихся целей. Эффективное решение этой задачи осложняется быстрыми изменениями положения и масштаба объектов, частичными перекрытиями объектов другими предметами (окклюзиями), размытием в движении и пропусками детекции, что приводит к потере информации и ошибкам при интерпретации данных.

Своевременное и точное определение положения объектов важно для обеспечения безопасности, принятия решений в реальном времени и автоматизации процессов наблюдения. Дополнительную сложность создаёт необходимость в обработке больших объёмов видеоданных с минимальной задержкой, то есть сочетать высокую вычислительную эффективность с точными алгоритмами анализа.

Обнаружение и отслеживание объектов в видеопотоке представляет собой комплексную задачу, включающую выявление объектов на отдельных кадрах (детекция) и построение их непрерывных траекторий во времени (трекинг). Эти этапы тесно связаны: качество детекции напрямую влияет на корректность трекинга, а ошибки трекинга могут приводить к фрагментации траекторий и ложным идентификациям.

Обычно различают две постановки задачи отслеживания: однообъектное отслеживание (single object tracking, SOT), где требуется сопровождать заранее заданную цель [1, 2], и многообъектное отслеживание (multi-object tracking, MOT), где необходимо одновременно сопровождать несколько объектов, поддерживая их идентичности во времени [3, 4].

II. ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ ОБНАРУЖЕНИЯ И ОТСЛЕЖИВАНИЯ ОБЪЕКТОВ

Развитие методов машинного обучения и глубоких нейронных сетей существенно расширило возможности автоматизации: современные подходы позволяют не только выделять объекты на отдельных кадрах, но и анализировать их поведение во времени [5]. Модели, такие как Faster R-CNN, SSD и YOLO, обеспечивают высокую точность и скорость обнаружения объектов, даже в условиях сложного фона и частичных окклюзий [6]. Для задачи отслеживания широко применяются подходы SORT, DeepSORT и ByteTrack, которые

обеспечивают устойчивое сопровождение объектов во времени за счёт ассоциации детекций между кадрами с использованием моделей движения и признаков внешнего вида, что повышает устойчивость трекинга в условиях перекрытия объектов и нестабильных детекций [7].

Однако большинство MOT-систем критически зависят от стабильности детекции и корректности межкадровой ассоциации [8, 9]. В условиях длительных окклюзий, резких манёвров объектов, изменений освещения или частоты кадров такие методы склонны к частой потери идентичности и фрагментации траекторий, что существенно снижает надёжность анализа поведения объектов во времени [5, 8]. Сквозные трансформерные архитектуры, хотя и уменьшают зависимость от ручных эвристик ассоциации, требуют значительных вычислительных ресурсов и больших объёмов размеченных данных, что затрудняет их использование в системах реального времени [3, 4].

Для снижения влияния нестабильных детекций на многообъектное отслеживание требуется предварительная оценка корректности обнаруженных объектов. Перспективным подходом является использование моделей сравнения признаков объектов, позволяющих оценить соответствие объекта целевым классам независимо от уверенности детектора.

Сиамские нейронные сети (Siamese Neural Network, SNN), обучаемые на парах изображений и предназначенные для измерения сходства между ними, обеспечивают вычисление оценки сходства между обнаруженным объектом и набором эталонных представлений. Использование этой оценки для фильтрации ложных детекций позволяет уменьшить влияние ошибок детектора, вызванных размытием в движении, изменениями освещения и частичными окклюзиями, и тем самым повысить устойчивость последующего этапа отслеживания без усложнения процедуры межкадровой ассоциации [10, 11].

III. РАЗРАБОТКА СИАМСКОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ОБНАРУЖЕНИЯ БЫСТРОДВИЖУЩИХСЯ ОБЪЕКТОВ

Сиамская нейронная сеть представляет собой архитектуру глубокого обучения, предназначенную для обучения функции сходства между двумя входами, реализуемая через идентичные ветви сети, разделяющие параметры. Каждая ветвь извлекает признаки из входного изображения и проецирует их в общее пространство признаков [12]. На рисунке 1 представлена схема структуры разработанной сиамской нейронной сети.

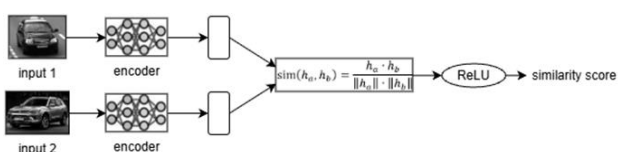


Рисунок 1 – Архитектура сиамской нейронной сети

Формально каждая ветвь сиамской нейронной сети реализует отображение по формуле (1):

$$f_{\theta} : X \rightarrow \mathbb{R}^d, \quad (1)$$

где X – пространство входных изображений, d – размерность признакового пространства, а θ – совокупность обучаемых параметров, общих для обеих ветвей. Для пары входных изображений $x_1, x_2 \in X$ сеть формирует два признаковых вектора $f_{\theta}(x_1)$ и $f_{\theta}(x_2)$ (h_1 и h_2 соответственно), которые затем используются для вычисления меры сходства.

В разработанной сиамской нейронной сети векторы признаков подвергаются L2-нормализации по формуле (2):

$$\hat{h}_i = \frac{h_i}{\|h_i\|_2}, \quad (2)$$

где $\|h_i\|_2 = \sqrt{\sum_{k=1}^d h_i[k]^2}$ – L2-норма вектора h_i , $i=1,2$.

Нормализация гарантирует, что все векторы лежат на единичной гиперсфере, а косинусная метрика, по которой оценивается сходство между входами, не зависит от масштаба признаков.

Косинусное сходство вычисляется по формуле (3):

$$s = \cos(\hat{h}_1, \hat{h}_2) = \hat{h}_1^T \hat{h}_2, \quad (3)$$

где $s \in [-1, 1]$ – значение сходства между двумя нормализованными векторами, $\hat{h}_1^T \hat{h}_2 = \sum_{k=1}^d \hat{h}_1[k] \cdot \hat{h}_2[k]$ – скалярное произведение двух L2-нормализованных векторов.

Для интерпретации результата в вероятностной форме значение сходства линейно нормализуется по формуле (4):

$$p = \frac{s+1}{2}, \quad (4)$$

в результате чего выход сети принимает значения в диапазоне $[0; 1]$ и может трактоваться как вероятность визуального соответствия пары изображений, где 1 – полное сходство, 0 – отсутствие сходства.

Обучение сети осуществляется на размеченных парах изображений с использованием бинарной кросс-энтропийной функции потерь по формуле (5):

$$L = -y \log(p) - (1-y) \log(1-p), \quad (5)$$

где $y \in \{0, 1\}$ – целевая метка, отражающая принадлежность пары к одному объекту или классу: 1 – пара соответствует одному объекту, 0 – разные объекты.

Разработанная архитектура сиамской нейронной сети обеспечивает интерпретируемость результата и делает возможным её применение в качестве модуля валидации результатов детекции: появляется возможность оценивать достоверность обнаруженных объектов путём сравнения их визуальных признаков с набором эталонных изображений, снижая влияние ложных срабатываний детектора и повышая устойчивость следующего этапа обработки – многообъектного отслеживания [11, 13].

IV. РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ И ОТСЛЕЖИВАНИЯ БЫСТРОДВИЖУЩИХСЯ ОБЪЕКТОВ

Сиамская нейронная сеть была интегрирована в разработанный пайплайн обработки видеопотока сразу после этапа детекции. Для каждого обнаруженного объекта вычисляется мера сходства относительно эталонного набора. Детекции с низкой вероятностью сходства отбрасываются перед передачей в алгоритм многообъектного отслеживания. Такая интеграция позволяет повысить качество трекинга за счёт фильтрации некорректных детекций. Последовательность этапов работы программного обеспечения представлена на рисунке 2.

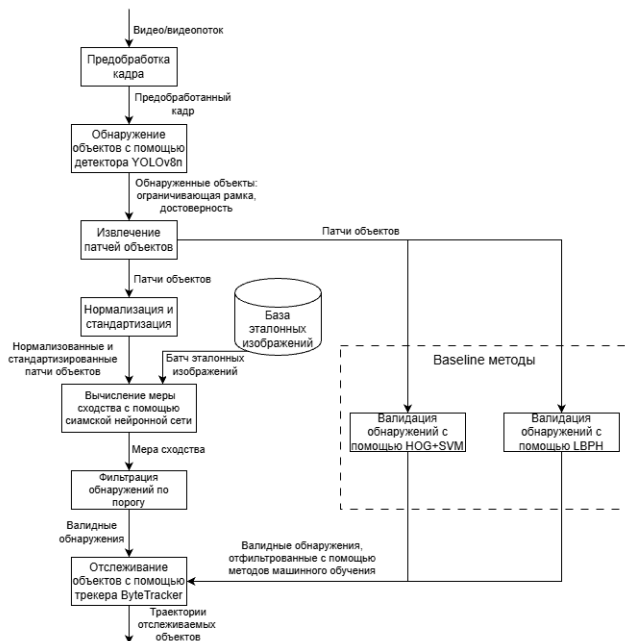


Рисунок 2 – Пайплайн разработанного программного обеспечения

Обработка видеопотока начинается с предварительно обработки кадров и последующего обнаружения объектов с использованием нейросетового детектора YOLOv8n. На этом этапе формируются наборы обнаружений, описываемые координатами ограничивающих прямоугольников и значениями уверенности детектора. Для каждого обнаружения из кадра извлекается соответствующий патч изображения, который затем подвергается нормализации и стандартизации для снижения влияния изменений освещенности и контраста.

Нормализованные патчи передаются в сиамскую нейронную сеть, где выполняется вычисление меры сходства путём сравнения обнаруженного объекта с набором эталонных изображений. Детекции с низким значением сходства отфильтровываются, а валидированные обнаружения используются в качестве входных данных для алгоритма многообъектного отслеживания ByteTrack, обеспечивающего построение траекторий объектов во времени.

V. ИССЛЕДОВАНИЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ И ОТСЛЕЖИВАНИЯ БЫСТРОДВИЖУЩИХСЯ ОБЪЕКТОВ

Для сравнительного анализа в качестве альтернативных модулей валидации детекций используются классические методы на основе ручных признаков, включая HOG+SVM и алгоритм LBPН, работающие с теми же входными детекциями и позволяющие оценить эффективность предложенного подхода в сопоставимых условиях.

Метод HOG+SVM сочетает гистограммы направленных градиентов (Histogram of Oriented Gradients, HOG) – дескриптор, который делит изображение на небольшие ячейки, вычисляет ориентации локальных градиентов интенсивности внутри ячеек и нормализует их в блоках для устойчивости – с классификатором опорных векторов (Support Vector Machine, SVM). HOG обеспечивает представление визуальных контуров и формы объекта в виде вектора признаков, а SVM обучается на этих векторах для отделения классов, например, валидных объектов от шумных детекций [14, 15]. Классификация изображения выполняется по формуле (6).

$$f(x) = w^T \phi_{\text{HOG}}(x) + b, \quad (6)$$

где x – входное изображение, ϕ_{HOG} – вектор признаков HOG, w – вектор весов классификатора SVM, b – смещение.

Алгоритм гистограммы локальных бинарных шаблонов (Local Binary Pattern Histogram, LBPН) извлекает локальную текстурную информацию из изображений путём пороговой бинаризации окружения каждого пикселя относительно центрального значения и последующего построения гистограмм этих бинарных паттернов по всему изображению. Итоговые гистограммы представляют текстурные характеристики объекта и используются в классификационных процедурах для отделения валидных объектов от неверных выявлений [16]. Формирование локального бинарного паттерна осуществляется по формуле (7).

$$\text{LBP}(x_c, y_c) = \sum_{y=0}^{P-1} s(I_p - I_c) 2^p, \quad (7)$$

где I_c – интенсивность центрального пикселя, I_p – интенсивность p -го соседнего пикселя, P – число соседних пикселей, $s(\cdot)$ – пороговая функция:

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}.$$

Для обучения и оценки моделей, используемых в качестве валидационных модулей, был выбран датасет из открытого доступа, содержащий изображения транспортных средств трёх категорий: автомобили, самолёты и корабли. Для каждого класса было отобрано 180 изображений, общий объём выборки составил 540 изображений. Датасет был разделён на обучающую и тестовую выборки в соотношении 80/20 (432 и 108 изображений соответственно). Перед обучением каждое изображение приводилось к единому размеру и подвергалось нормализации и стандартизации с целью снижения влияния различий в освещённости и контрасте.

Для всех моделей использовались фиксированные наборы параметров. Для сиамской нейронной сети

применялся энкодер на базе ResNet-50 с размерностью внутреннего представления 128, веса энкодера были заморожены. Сравнение пар изображений выполнялось по косинусному сходству с порогом 0,7. Обучение проводилось с размером батча 32, числом эпох 50, скоростью обучения 0,001, валидационной долей 0,2. В методе HOG+SVM использовались стандартные параметры HOG, а классификация выполнялась SVM с RBF-ядром и включённой оценкой вероятностей. Для LBPН применялся оператор LBP с радиусом 1 и 8 точками, гистограммы строились на сетке 8 на 8, а сходство оценивалось по корреляции гистограмм с порогом 0,7.

Модели обучались в одинаковых условиях на одних и тех же данных. Качество методов оценивалось с использованием стандартных метрик классификации: точность (precision), полнота (recall) и F1-мера. Сравнительный анализ результатов обучения моделей представлен в таблице 1.

Таблица 1 – Сопоставительная характеристика показателей обучения моделей

Метрики	SNN	HOG+SVM	LBPН
Precision	0,84	0,65	0,58
Recall	0,87	0,75	0,69
F1-мера	0,85	0,70	0,63

Полученные результаты показывают преимущество сиамской нейронной сети по всем основным метрикам. Высокое значение точности (0,84) указывает на способность SNN эффективно отфильтровывать ложноположительные совпадения, а значение полноты (0,87) свидетельствует о том, что сеть корректно выявляет большинство истинно схожих пар изображений, обеспечивая сбалансированное качество распознавания. Это подтверждается наивысшим значением F1-меры (0,85) среди всех рассмотренных методов.

Метод HOG+SVM демонстрирует средние показатели: при умеренной точности (0,65) и полноте (0,75) наблюдаются пропуски схожих объектов, особенно при вариациях внешнего вида. LBPН показывает наименьшие значения среди рассматриваемых методов: точность составляет 0,58, а полнота – 0,69. Это означает, что модель чаще допускает ложные срабатывания и пропускает часть действительно схожих объектов. В результате F1-мера обоих базовых методов оказалась ниже, чем у сиамской нейронной сети, что подтверждает более высокую устойчивость и обобщающую способность SNN при оценке визуального сходства объектов.

Также было произведено исследование влияния модулей валидации на качество многообъектного отслеживания с целью оценки воздействия предварительной фильтрации детекций на качество работы трекера. Параметры трекера были зафиксированы для всех сценариев эксперимента: порог инициализации трека 0,5, размер буфера трека 30, порог сопоставления 0,8, порог многообъектного отслеживания 0,8, порог близости 0,5, порог внешнего вида 0,25, частота кадра принята равной 30. Различия в результатах обусловлены исключительно наличием и типом используемого модуля валидации.

Исследовались четыре конфигурации системы: базовый вариант ByteTrack без предварительной валидации обнаружений, а также варианты с использованием валидации на основе сиамской нейронной сети, HOG+SVM и LBPН. Во всех случаях на вход трекера подавались результаты одного и того же детектора, а модули валидации применялись на этапе фильтрации обнаруженных объектов до их передачи в трекер. Процесс отслеживания быстро движущихся объектов в интерфейсе разработанного программного обеспечения представлен на рисунке 3.



Рисунок 3 – Механизм отслеживания быстро движущихся объектов в разработанном программном обеспечении

Качество многообъектного отслеживания оценивалось с помощью стандартных метрик: общая эффективность отслеживания (MOTA), качество сохранения идентичностей объектов (IDF1).

Результаты показывают, что использование модулей валидации положительно влияет на качество многообъектного отслеживания по всем рассмотренным метрикам. Базовая конфигурация ByteTrack без валидации демонстрирует наименьшие показатели, что связано с передачей в трекер ложных и нестабильных детекций, приводящих к фрагментации траекторий и частым ошибкам идентификации.

Наилучшие результаты достигнуты с сиамской нейронной сетью: MOTA составил 0,64, IDF1 – 0,73. HOG+SVM и LBPН также улучшают метрики относительно базового варианта, но уступают SNN из-за ограниченной выразительности ручных признаков.

VI. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования было разработано программное обеспечение для обнаружения и отслеживания быстро движущихся объектов. В предложенном решении был реализован модуль валидации результатов детекции на основе сиамской нейронной сети, посредством которого было снижено влияние ошибок детектора за счёт предварительной фильтрации обнаруженных объектов перед этапом трекинга. Проведённые эксперименты показали, что использование SNN в роли валидационного модуля позволяет повысить устойчивость отслеживания по сравнению как с базовой конфигурацией трекера без валидации, так и с классическими методами на основе классических методов машинного обучения HOG+SVM и LBPН.

Разработанное программное обеспечение может

применяться в задачах анализа видеопотоков с динамичными сценами, включая системы мониторинга дорожного движения, наблюдение за воздушными и морскими объектами, а также другие прикладные сценарии, требующие надёжного сопровождения быстро движущихся целей.

В качестве направлений дальнейшего развития можно выделить расширение поддерживаемых классов объектов, адаптацию алгоритмов к работе в реальном времени на ограниченных вычислительных ресурсах и исследование влияния совместной оптимизации детекции, валидации и трекинга на качество всей системы.

БИБЛИОГРАФИЯ

- [1] Cui Y. et al. Mixformer: End-to-end tracking with iterative mixed attention // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2022. – P. 13608–13618.
- [2] Ye B. et al. Joint feature learning and relation modeling for tracking: A one-stream framework // European conference on computer vision. – Cham : Springer Nature Switzerland, 2022. – P. 341–357.
- [3] Meinhardt T. et al. Trackformer: Multi-object tracking with transformers // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2022. – P. 8844–8854.
- [4] Zeng F. et al. Motr: End-to-end multiple-object tracking with transformer // European conference on computer vision. – Cham : Springer Nature Switzerland, 2022. – P. 659–675.
- [5] Kadam P., Fang G., Zou J. J. Object tracking using computer vision: A review // Computers. – 2024. – Т. 13. – № 6. – P. 136.
- [6] Zaidi S. S. A. et al. A survey of modern deep learning based object detection models // Digital Signal Processing. – 2022. – Т. 126. – P. 103514.
- [7] Abouelyazid M. Comparative evaluation of sort, deepsort, and bytetrack for multiple object tracking in highway videos // International Journal of Sustainable Infrastructure for Cities and Societies. – 2023. – Т. 8. – № 11. – P. 42–52.
- [8] Adžemović M. Deep Learning-Based Multi-Object Tracking: A Comprehensive Survey from Foundations to State-of-the-Art // arXiv preprint arXiv:2506.13457. – 2025.
- [9] Cheng Y. et al. Improved particle filter algorithm for multi-target detection and tracking // Sensors (Basel, Switzerland). – 2024. – Т. 24. – № 14. – P. 4708.
- [10] Javed S. et al. Visual object tracking with discriminative filters and siamese networks: a survey and outlook // IEEE transactions on pattern analysis and machine intelligence. – 2022. – Т. 45. – № 5. – P. 6552–6574.
- [11] Lim S. C., Huh J. H., Kim J. C. Siamese Trackers Based on Deep Features for Visual Tracking // Electronics. – 2023. – Т. 12. – № 19. – P. 4140.
- [12] Chicco D. Siamese neural networks: An overview // Artificial neural networks. – 2021. – P. 73–94.
- [13] Xie J., Guo L. Siamese Network for Similar Data Recognition Filtering // Proceedings of the 2024 7th International Conference on Big Data Technologies. – 2024. – P. 81–85.
- [14] Vieira L. Comparing Performance of Preprocessing Techniques for Traffic Sign Recognition Using a HOG-SVM // arXiv preprint arXiv:2504.09424. – 2025.
- [15] Siamese neural network for few-shot visual inspection of steel surface defects in predictive maintenance systems / E. Antonyants, A. Aletdinova, V. Zaozernov, A. Leontiev, A. Vasiliev. – DOI 10.1109/APEIE66761.2025.11289355. – Text : electronic // 17 International scientific and technical conference actual problems of electronic instrument engineering (APEIE–2025) : proc., Novosibirsk, 14–16 Nov. 2025. – Novosibirsk : Publ. NSTU, 2025. – 6 p. – URL: <https://ieeexplore.ieee.org/document/11289355> (access date: 21.02.2026).
- [16] Sharma D., Singh B. K. Smart CCTV detection using local binary pattern histogram (LBPH) // Int Res J Adv Sci Hub. – 2023. – Т. 5. – № 08. – P. 290–298.

Антонянц Егор Николаевич. Новосибирский государственный технический университет, г. Новосибирск, Россия. Аспирант факультета автоматики и вычислительной техники. Количество печатных работ: 59. Область научных интересов: машинное обучение,

системы компьютерного зрения, информационные сети. e-mail: bax201438@gmail.com (ответственный за переписку).

Гордиенко Дмитрий Андреевич. Новосибирский государственный технический университет, г. Новосибирск, Россия. Студент бакалавриата факультета автоматики и вычислительной техники. Количество печатных работ: 1. Область научных интересов: глубокое обучение, нейронные сети, системы компьютерного зрения.

Савочка Денис Владимирович. Новосибирский государственный технический университет, г. Новосибирск, Россия. Студент бакалавриата факультета автоматики и вычислительной техники. Количество печатных работ: 1. Область научных интересов: машинное обучение, системы искусственного интеллекта, системы компьютерного зрения.

Хусанов Артур Мансурович. Новосибирский государственный технический университет, г. Новосибирск, Россия. Студент бакалавриата факультета автоматики и вычислительной техники. Количество печатных работ: 1. Область научных интересов: глубокое обучение, синтез речи, системы компьютерного зрения.

Development and research of software for detection and tracking of fast-moving objects

E.N. Antonyants, D.A. Gordienko, D.V. Savochka, A.M. Khusainov

Аннотация — The article is devoted to the development and research of software for detecting and tracking fast-moving objects in a video sequence. The paper considers the main tasks of detection and tracking and provides an overview of existing architectures and algorithms used to improve the accuracy and speed of data processing. Special attention is paid to the application of a Siamese neural network for assessing the visual similarity of objects and validating detections during the tracking process.

Within the framework of the study, software was developed that implements modules for detection, validation of detections using a Siamese neural network, and tracking. Experimental testing and a comparative analysis of the proposed approach with traditional methods based on handcrafted features were conducted. The quality of recognition was evaluated using precision and recall metrics, and the performance of multi-object tracking was assessed using the MOTA and IDF1 metrics.

The results of the experiments confirm the superiority of the neural network approach: the object detection accuracy reached 0,84 with a recall of 0,87. The obtained results show higher robustness to variations in the appearance of objects compared to classical methods. The use of a Siamese network made it possible to reduce the number of object losses and identity switch errors, which is reflected in the MOTA (0,64) and IDF1 (0,73) values.

The developed approach demonstrates high efficiency in solving the problems of detecting and tracking fast-moving objects and confirms the promise of using Siamese neural networks for the further development of intelligent computer vision systems.

Keywords — *object detection, object tracking, multi-object tracking, fast-moving objects, Siamese neural network*

REFERECES

- [1] Cui Y. et al. Mixformer: End-to-end tracking with iterative mixed attention // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2022. – P. 13608–13618.
- [2] Ye B. et al. Joint feature learning and relation modeling for tracking: A one-stream framework // European conference on computer vision. – Cham : Springer Nature Switzerland, 2022. – P. 341–357.
- [3] Meinhardt T. et al. Trackformer: Multi-object tracking with transformers // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2022. – P. 8844–8854.
- [4] Zeng F. et al. Motr: End-to-end multiple-object tracking with transformer // European conference on computer vision. – Cham : Springer Nature Switzerland, 2022. – P. 659–675.
- [5] Kadam P., Fang G., Zou J. J. Object tracking using computer vision: A review // Computers. – 2024. – T. 13. – № 6. – P. 136.
- [6] Zaidi S. S. A. et al. A survey of modern deep learning based object detection models // Digital Signal Processing. – 2022. – T. 126. – P. 103514.
- [7] Abouelyazid M. Comparative evaluation of sort, deepsort, and bytetrack for multiple object tracking in highway videos // International Journal of Sustainable Infrastructure for Cities and Societies. – 2023. – T. 8. – № 11. – P. 42–52.
- [8] Adžemović M. Deep Learning-Based Multi-Object Tracking: A Comprehensive Survey from Foundations to State-of-the-Art // arXiv preprint arXiv:2506.13457. – 2025.
- [9] Cheng Y. et al. Improved particle filter algorithm for multi-target detection and tracking // Sensors (Basel, Switzerland). – 2024. – T. 24. – № 14. – P. 4708.
- [10] Javed S. et al. Visual object tracking with discriminative filters and siamese networks: a survey and outlook // IEEE transactions on pattern analysis and machine intelligence. – 2022. – T. 45. – № 5. – P. 6552–6574.
- [11] Lim S. C., Huh J. H., Kim J. C. Siamese Trackers Based on Deep Features for Visual Tracking // Electronics. – 2023. – T. 12. – № 19. – P. 4140.
- [12] Chicco D. Siamese neural networks: An overview // Artificial neural networks. – 2021. – P. 73–94.
- [13] Xie J., Guo L. Siamese Network for Similar Data Recognition Filtering // Proceedings of the 2024 7th International Conference on Big Data Technologies. – 2024. – P. 81–85.
- [14] Vieira L. Comparing Performance of Preprocessing Techniques for Traffic Sign Recognition Using a HOG-SVM // arXiv preprint arXiv:2504.09424. – 2025.
- [15] Siamese neural network for few-shot visual inspection of steel surface defects in predictive maintenance systems / E. Antonyants, A. Aletdinova, V. Zaozernov, A. Leontiev, A. Vasiliev. – DOI 10.1109/APEIE66761.2025.11289355. – Text : electronic // 17 International scientific and technical conference actual problems of electronic instrument engineering (APEIE–2025) : proc., Novosibirsk, 14–16 Nov. 2025. – Novosibirsk : Publ. NSTU, 2025. – 6 p. – URL: <https://ieeexplore.ieee.org/document/11289355> (access date: 21.02.2026).
- [16] Sharma D., Singh B. K. Smart CCTV detection using local binary pattern histogram (LBPH) // Int Res J Adv Sci Hub. – 2023. – T. 5. – № 08. – P. 290–298.