

ОБНАРУЖЕНИЕ И ПРЕДСКАЗАНИЕ ТРАЕКТОРИИ ДВИЖЕНИЯ ОБЪЕКТОВ

А. Г. Каркоцкий

*Белорусский государственный университет,
Минск, Беларусь, karkotskiy.alexander@yandex.by*

В работе рассматривается комбинация нескольких методов компьютерного зрения и машинного обучения для обнаружения, отслеживания и предсказания траектории движения объектов в видеопотоке. Для этого используются сверточная нейронная сеть YOLOv8 для обнаружения объектов, рекуррентная нейронная сеть LSTM для прогнозирования траектории на основе последовательности координат и CLAHE с Gamma-коррекцией для улучшения качества изображения перед передачей в YOLOv8.

Ключевые слова: обнаружение объектов; предсказание траектории; YOLOv8; LSTM; CLAHE; Gamma-коррекция.

DETECTION AND TRAJECTORY PREDICTION OF OBJECTS

A. G. Karkotskiy

*Belarusian State University,
Minsk, Belarus, karkotskiy.alexander@yandex.by*

The study explores a combination of several computer vision and machine learning methods for detecting, tracking, and predicting object trajectories in a video stream. For this purpose, the YOLOv8 convolutional neural network is used for object detection, an LSTM recurrent neural network is employed for trajectory prediction based on a sequence of coordinates and CLAHE with Gamma correction is applied to enhance image quality before processing by YOLOv8.

Keywords: object detection; trajectory prediction; YOLOv8; LSTM; CLAHE; Gamma correction.

1. Введение

Актуальной задачей современных систем компьютерного зрения является не только обнаружение движущихся объектов, но и точное предсказание их траектории. Эта способность является важной для таких сфер, как автономное вождение и видеонаблюдение, где от точного прогноза по-

ведения пешехода или автомобиля зависит безопасность. Также определенные проблемы могут возникать в условиях недостаточной освещенности, однако, в рамках данной работы мы сосредоточимся, преимущественно, только на обнаружении движущихся объектов и предсказания траектории их движения.

Нейронные сети предоставляют мощные инструменты для анализа изображений в сложных условиях. Благодаря сверточным архитектурам (CNN) они могут эффективно выделять значимые признаки даже в шумных данных. Такие технологии, как YOLO и Faster R-CNN, обеспечивают обнаружение объектов с высокой точностью и минимальной задержкой. Рекуррентные нейронные сети, такие как сеть долгой краткосрочной памяти (LSTM) могут использоваться для анализа последовательности движения и предсказания траектории движения, учитывая, при этом, временную связь между кадрами.

Эффективная обработка данных также является важной для успешного обучения нейронной сети. Методы шумоподавления (Gaussian Blur, Median Filtering) и улучшения контрастности (CLAHE, адаптивная нормализация) помогают улучшить качество изображения, что важно при анализе изображений с низким освещением.

2. Используемые технологии

В работе используется комбинация методов компьютерного зрения и машинного обучения для обнаружения, отслеживания и предсказания траектории движения объектов в видеопотоке. Основу системы составляют три ключевых компонента: YOLOv8 для детекции объектов, LSTM для прогнозирования траекторий и CLAHE с Gamma-коррекцией для предобработки изображений.

YOLOv8 (сверточная нейронная сеть) применяется для обнаружения объектов в реальном времени [1]. Эта архитектура отличается высокой скоростью работы благодаря однопроходному принципу детекции. В отличие от предыдущих версий, YOLOv8 использует anchor-free подход, что упрощает обучение и повышает точность. При реализации модель загружается с предобученными весами (yolov8n.pt) и применяется для обнаружения объектов с сохранением идентификаторов между кадрами. Обнаруженные объекты (люди, автомобили и другие элементы из 80 классов COCO) выделяются ограничивающими рамками, а их координаты сохраняются для дальнейшего анализа.

LSTM (долгая краткосрочная память) — это рекуррентная нейронная сеть, предназначенная для прогнозирования траекторий на основе последовательностей координат [2]. При реализации LSTM анализирует исто-

рию перемещений объектов (последние 5 точек) и предсказывает их будущее положение. Модель обучается в режиме онлайн, т.е. в режиме реального времени, корректируя свои предсказания на основе новых данных. Входные данные включают не только координаты (x , y), но и их производные (dx , dy), что позволяет также учитывать скорость движения. На выходе сеть предсказывает смещение объекта, которое масштабируется и визуализируется в виде «зоны предсказания» и стрелки направления.

CLANE и **Gamma-коррекция** используются для улучшения качества изображения перед подачей в YOLOv8 [3]. Метод CLANE (Contrast Limited Adaptive Histogram Equalization) адаптивно выравнивает контраст в локальных областях кадра, что позволяет исправить затемненные или засвеченные элементы сцены. Gamma-коррекция дополнительно регулирует яркость, делая объекты более различимыми. При реализации эти методы применяются к каждому кадру перед детекцией, что повышает надежность работы системы в сложных условиях освещения.

Визуализация включает ограничивающие рамки, историю перемещений, предсказанные траектории и направления движения, что делает анализ наглядным и удобным для интерпретации.

3. Сеть YOLOv8

Опишем архитектуру нейронной сети YOLOv8 [1].

В программной реализации используется небольшая и оптимизированная версия YOLOv8 – YOLOv8n.

Нейронная сеть YOLOv8 состоит из трех основных составляющих: backbone, neck и head.

В основе сети лежит сверточный нейронный сетевой экстрактор признаков (backbone), который извлекает многоуровневые признаки из входного изображения и использует специальные архитектурные блоки (Cross Stage Partial) для уменьшения количества параметров и повышения скорости.

Задача neck заключается в объединении признаков из разных уровней backbone, чтобы сеть могла работать с объектами разного размера. В реализации YOLOv8n используется PAN-FPN (Path Aggregation Network с Feature Pyramid Network) — гибридная архитектура для объединения информации из разных уровней, которая позволяет повысить чувствительность к мелким и крупным объектам.

Head отвечает за предсказание координат ограничивающих рамок (bounding box), классов и степени уверенности распознанного объекта (confidence score).

В слоях используется функция активации SiLu (1) для стабилизации обучения и повышения нелинейности [4].

$$\text{SiLu}(x) = x \cdot \sigma(x), \quad (1)$$

где x – это входной сигнал; $\sigma(x) = \frac{1}{1 + e^{-x}}$ – это сигмоида.

Функция SiLu позволяет добиться лучшей производительности, чем традиционная ReLu (2) [4].

$$\text{ReLu}(x) = \max(0, x), \quad (2)$$

где x – это входной сигнал.

4. Сеть LSTM

Опишем архитектуру используемой нейронной сети LSTM. Ее изображение приведено на рис. 1.

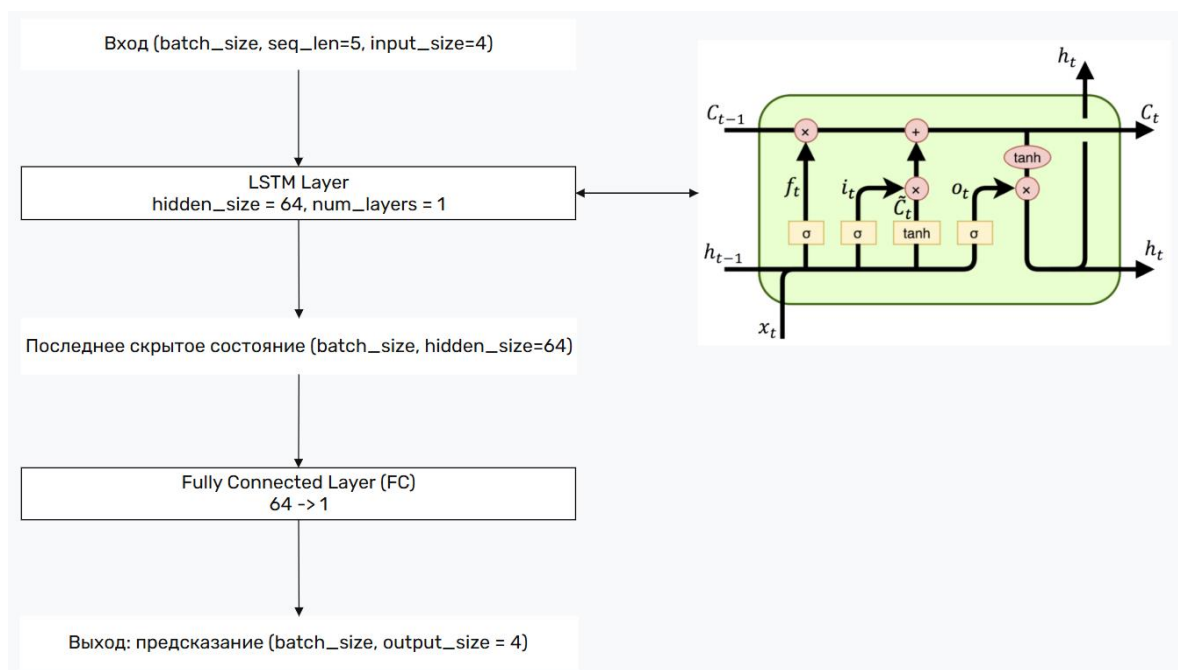


Рис. 1. Архитектура используемой нейронной сети LSTM

На вход поступает временной шаг, который содержит 4 признака: координаты центра объекта на кадре и приращение координат (разница с предыдущим шагом). Далее используется один LSTM-слой [2]. Один LSTM-слой позволяет бороться с переобучением, обеспечивает быстроту работы, что является важным при работе в реальном времени. В то же время

один слой LSTM способен предсказывать линейные и немного криволинейные траектории, однако в случае предполагаемого сложного движения объектов (скачки, повороты, ускорения) предполагается увеличение количества слоев, что, однако, может сказаться на производительности. После обработки последовательности из 5 шагов берется скрытое состояние последнего шага, которое содержит обобщенную информацию о движении объекта. Это состояние проходит через полносвязный слой, который дает прогноз: то вероятное изменение координат для следующего шага.

5. Общая схема реализации

Система реализует следующую сквозную схему обработки видео:

- 1) улучшение кадра с помощью CLAHE и Gamma-коррекции → Обнаружение объектов с использованием YOLOv8 → Отслеживание движения → Прогнозирование траектории движения с использованием LSTM;
- 2) онлайн-обучение адаптирует модель под конкретную сцену;
- 3) визуализация делает анализ наглядным и интуитивно понятным.

На рис. 3 приведен пример работы разработанной системы на одном кадре видео, который представлен на рис 2.

Код структурирован так, что каждый компонент можно модифицировать независимо. Это делает систему гибкой для применений в автономных роботах, видеонаблюдении и анализе трафика.



Рис. 2. Один кадр видео, использованного для тестирования системы

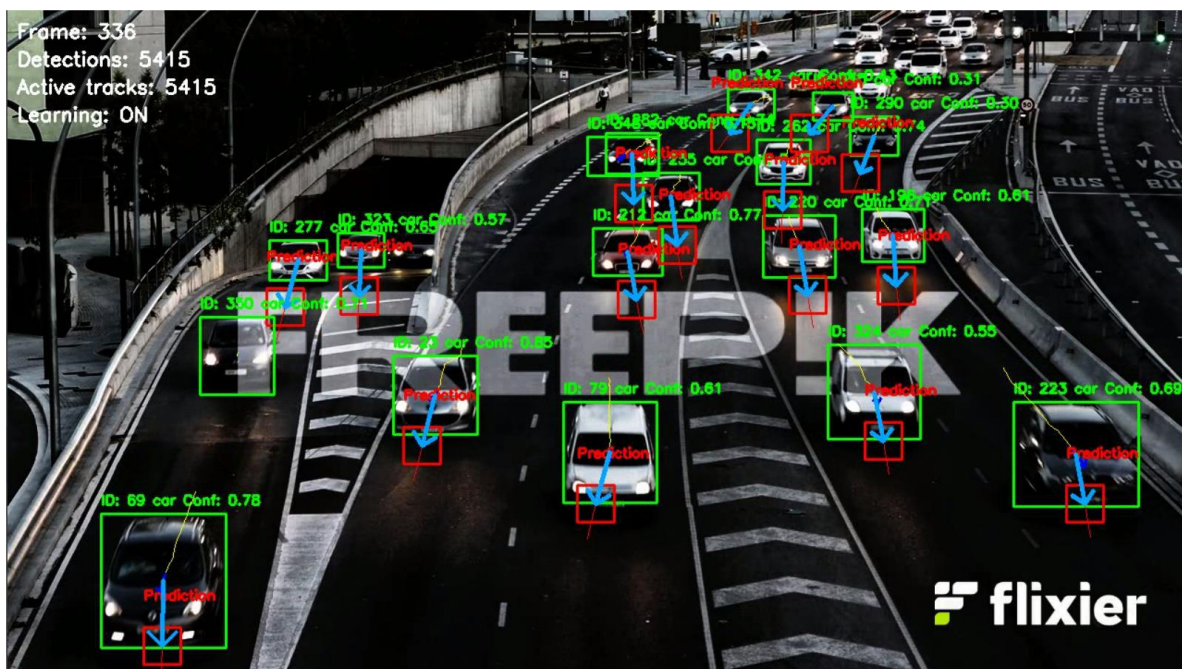


Рис. 3. Пример работы разработанной системы на одном кадре видео

Заключение

В рамках данной работы была успешно разработана и реализована комплексная система для обнаружения, отслеживания и предсказания траектории движения объектов в видеопотоке в реальном времени. Ключевым достижением является эффективная комбинация методов компьютерного зрения и машинного обучения, каждый из которых решает специфическую задачу в рамках общей последовательности обработки данных.

В качестве детектора объектов была использована сверточная нейронная сеть YOLOv8, которая продемонстрировала высокую скорость работы благодаря своей однопроходной архитектуре и anchor-free подходу. Для повышения надежности детекции в сложных условиях освещения был применен этап предобработки изображения с использованием методов CLAHE и Gamma-коррекции, что позволяет улучшить контрастность и качество входных данных для нейронной сети.

Для решения задачи прогнозирования траектории была задействована рекуррентная нейронная сеть LSTM для анализа временной последовательности координат. Модель, обученная на истории перемещений объектов и их производных (скорости), показала способность предсказывать будущее положение объектов с учетом их динамики. Важной особенностью реализации является возможность онлайн-обучения, что позволяет системе адаптироваться к конкретной сцене и поведению объектов в реальном времени.

Предложенное решение является модульным и гибким, что позволяет независимо модифицировать и улучшать каждый из его компонентов. Визуализация результатов в виде ограничивающих рамок, истории перемещений и предсказанных траекторий обеспечивает наглядность и удобство интерпретации.

Проведенное исследование подтверждает перспективность использования связки YOLOv8 и LSTM для задач прогнозирования движения. Разработанная система представляет практическую ценность и может быть применена в таких областях, как автономное вождение, системы видеонаблюдения, анализ трафика и робототехника. В качестве направлений для дальнейших исследований можно выделить увеличение сложности модели LSTM для предсказания нелинейных траекторий. В том числе необходимо уделить отдельное внимание более глубокой предобработке кадров для улучшения видимости в сложных условиях.

Библиографические ссылки

1. YOLOv8 Architecture: A Deep Dive into its Architecture. URL: <https://yolov8.org/yolov8-architecture/> (date of access: 10.09.2025).
2. Introduction to Long Short-Term Memory (LSTM). URL: <https://medium.com/analytics-vidhya/introduction-to-long-short-term-memory-lstm-a8052cd0d4cd> (date of access: 11.09.2025).
3. Automatic Contrast-Limited Adaptive Histogram Equalization With Dual Gamma Correction / Y. Chang [et al.] // Proceedings of the IEEE. 2018. Vol. 6. P. 11782–11792.
4. Activation Functions in Focus: Understanding ReLU, GELU, and SiLU. URL: https://medium.com/@varun_mishra/activation-functions-in-focus-understanding-relu-gelu-and-silu-841ed1c6df0c (date of access: 13.09.2025).