

УДК 381.111

СРАВНЕНИЕ И КРИТИЧЕСКИЙ АНАЛИЗ ПРИНЦИПИАЛЬНОЙ ЛОГИКИ И ПРИМЕНЕНИЯ ПОПУЛЯРНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ВИЗУАЛЬНОЙ ДЕТЕКЦИИ И СЕГМЕНТАЦИИ ОБЪЕКТОВ В ВИДЕОПОТОКЕ

Валадов Антон Сергеевич,

студент, магистрант

1 курс, факультет "Отдел магистратуры", кафедра «Медиатехнологии»

Донской государственной технической университет

Россия, г. Ростов-на-Дону

Кудинов Никита Георгиевич,

руководитель Центра R&D, Технопарк Магика,

генеральный директор, ООО «ЗРЕНИЕ 2.0» (резидент Фонда «Сколково»)

аспирант по направлению «Медиакоммуникация и журналистика»

Донской государственной технической университет

Россия, г. Ростов-на-Дону

Аннотация

Исследование представляет анализ и сравнительный обзор методов визуальной детекции и сегментации объектов в видеопотоке, с акцентом на популярные решения. Работа охватывает методологию и эксперименты, подробно описывающие процесс подготовки данных, выборку и критерии оценки производительности. Результаты экспериментов предоставляют полный обзор сравнительного анализа методов, что позволяет выделить наиболее эффективные подходы в визуальной детекции и сегментации объектов в видеопотоке. Также рассматриваются практические аспекты применения обсуждаемых методов, предлагаются рекомендации и области их оптимального использования. Заключение статьи подводит итоги проведенного исследования, выделяя перспективы будущего развития этой области и возможные направления дальнейших исследований.

Ключевые слова: компьютерное зрение, визуальная детекция, сегментация изображения, нейронные сети

COMPARISON AND CRITICAL ANALYSIS OF THE FUNDAMENTAL LOGIC AND APPLICATION OF POPULAR NEURAL NETWORKS FOR VISUAL DETECTION AND SEGMENTATION OF OBJECTS IN A VIDEO STREAM

Anton S. Valadov,

student, master's student

1st year, Faculty "Master's Degree Division", Department "Media Technologies"
Don State Technical University
Russia, Rostov-on-Don

Nikita G. Kudinov,

Head of R&D Center, Technopark Magika,
General Director, LLC "Sight 2.0" (Skolkovo Foundation resident)
Postgraduate student in the field of Media Communication and Journalism
Don State Technical University
Russia, Rostov-on-Don

ABSTRACT

The study presents an analysis and comparative review of methods for visual detection and segmentation of objects in a video stream, with an emphasis on popular solutions. The work covers methodology and experiments detailing the data preparation process, sampling, and performance evaluation criteria. The experimental results provide a complete overview of the comparative analysis of the methods, which allows us to identify the most effective approaches in visual detection and segmentation of objects in the video stream. The practical aspects of the application of the discussed methods are also considered, recommendations and areas of their optimal use are proposed. The conclusion of the article summarizes the results of the conducted research, highlighting the prospects for the future development of this area and possible directions for further research.

Keywords: computer vision, visual detection, image segmentation, neural networks

Компьютерное зрение – это отрасль компьютерных технологий, в которой разрабатываются научные основы распознавания объектов и создаются прикладные системы, способные обнаруживать и распознавать объекты по аналогии со зрительным восприятием реальности человеком.[1]

Актуальность темы определяется важностью визуального анализа в различных сферах, включая видеонаблюдение, автономную навигацию, медицинскую диагностику и многие другие области. Обеспечение точной и быстрой детекции, а также сегментации объектов в видеопотоках является критическим элементом для создания интеллектуальных систем, способных адаптироваться к разнообразным сценариям и обеспечивать высокий уровень безопасности и эффективности.

Принципиальная логика в контексте визуальной детекции и сегментации

Принципиальная логика (ПЛ) в области визуальной детекции и сегментации объектов в видеопотоке представляет собой формальную систему, основанную на принципах формализации и логической структурированности объектов. Математически, ПЛ может быть выражена через логические операции и предикаты.

Пусть X - пространство входных визуальных данных, а Y - множество классов объектов. ПЛ формулируется как функция $F: X \rightarrow Y$, где каждый элемент, где $x \in X$ преобразуется в соответствующий класс объекта $y \in Y$ на основе логических условий.

Методы, основанные на PL, оценивают характеристики объектов, такие как цвет, текстура, форма, через систему формальных правил. Однако, ограниченность проявляется в сложности формализации абстрактных понятий и статичности правил, что может затруднить адаптацию к изменчивым сценам.

Основные концепции и принципы работы популярных нейронных сетей для компьютерного зрения

Ключевые концепции и принципы работы современных нейронных сетей, предназначенных для компьютерного зрения:

Свёрточные нейронные сети (CNN):

- Определение: CNN являются одними из наиболее распространенных архитектур для задач классификации и детекции изображений.
- Принцип работы: они используют операции свертки для выделения иерархии признаков из изображений, начиная с низкоуровневых деталей и постепенно переходя к более абстрактным характеристикам.

Рекуррентные нейронные сети (RNN):

- Определение: RNN предназначены для обработки последовательных данных, что делает их подходящими для анализа временных зависимостей в видеопотоках.
- Принцип работы: Эти сети обладают памятью, позволяющей учесть контекст предыдущих кадров, что особенно полезно в задачах сегментации движущихся объектов.

3D-сверточные нейронные сети (3D CNN):

- Определение: 3D CNN расширяют концепцию свёрточных сетей на трехмерное пространство, что учитывает не только пространственные, но и временные характеристики видеопотоков.
- Принцип работы: они способны захватывать изменения в видеопоследовательностях, что делает их мощным инструментом для визуальной детекции и сегментации в динамичных сценах.[2]

Процессы компьютерного зрения

Внедрение технологий компьютерного зрения ориентировано на эффективное использование ресурсов исполняющего устройства, необходимой производительности и конкретной области применения. Процессы компьютерного зрения включают несколько ключевых шагов:

- Захват объектов: Объекты могут быть представлены в формате изображений 2D или 3D, аудиозаписей или видеоматериалов.
- Предварительная обработка: перед применением алгоритмов производится предварительная обработка объектов. Этот этап включает в себя такие шаги, как удаление шумов, коррекция яркости и контрастности для обеспечения успешного выявления объектов. Также проводится проверка корректности загрузки объекта и установка оптимального уровня громкости для аудиофайлов. Кроме того, изображения масштабируются для улучшения их качества.

- Выделение деталей: Разнообразные детали с различной степенью сложности выделяются из входного файла. Например, для изображений подразумевается выделение геометрических примитивов, таких как линии, границы, кромки, а также выделение локализованных точек интереса, таких как капли, точки или углы. Детали, характеризующиеся более сложной структурой, относятся к структуре, форме или движению.
- Детектирование и сегментация: на этапе обработки анализируются выбранные детали, и принимается решение, какие из них следует выделить для дальнейшей обработки. Это может включать выделение групп точек или сегментацию объекта на изображении.
- Высокоуровневая обработка: на завершающем этапе формируется список, содержащий информацию о найденных объектах. Например, для изображений этот список может включать координаты точек или области изображения, где, вероятно, расположен объект. [3]

Сравнение основных подходов сегментации и детекции видеопотока

В таблице 1 описаны популярные алгоритмы, такие как YOLO (You Only Look Once), SSD (Single Shot Multibox Detector), Faster R-CNN (Region-based Convolutional Neural Network), DeepLab [4]. Каждый из этих подходов имеет свои особенности, предназначенные для различных типов задач, от обнаружения объектов в реальном времени до точной сегментации изображений.

Таблица 1 – Описание популярных подходов сегментации и детекции изображений.

Подход	YOLO	SSD	Faster R-CNN	DeepLab
Алгоритм	YOLO (You Only Look Once)	SSD (Single Shot Multibox Detector)	Faster R-CNN (Region-based CNN)	DeepLab (Deep Semantic Segmentation)
Тип задачи	Объектная детекция	Объектная детекция, сегментация	Объектная детекция	Сегментация изображений
Принцип работы	Однократное просмотр всего изображения	Однократный просмотр для обнаружения и классификации объектов.	Многokrатный просмотр для обнаружения и сегментации объектов.	Использование ASPP для сегментации
Архитектура	Сверточные нейронные сети	Сверточные нейронные сети	Комбинация сверточных и рекуррентных нейронных сетей.	Специализированная архитектура сети с энкодером и декодером.
Скорость обработки	Высокая, благодаря однократному просмотру	Высокая, особенно для реального времени	Средняя, зависит от сложности модели и размера изображения	Коррелирует со сложностью модели и размером изображения.

Применимость	Широкий спектр, включая реальное время. Эффективность в реальном времени	Области, где важна высокая скорость и эффективность обнаружения объектов. Быстрая и точная детекция объектов с высокой эффективностью.	Задачи, где важны точность и сложные объекты с множественным и деталями. Точная детекция, поддержка множественных классов объектов.	Задачи сегментации, где важна высокая детализация изображений. Точная сегментация, учет контекста изображения.
Недостатки	Ограничения в точности по сравнению с более тщательными методами	Возможные ограничения в точности и скорости при работе с большими объемами данных.	Значительные вычислительные затраты при обработке больших объемов данных.	Значительные вычислительные затраты при обработке высоко разрешенных

Вывод по сравнению подходов

Из проведенного сравнительного анализа различных подходов к компьютерному зрению и сегментации объектов видно, что каждый метод обладает своими сильными и слабыми сторонами, что делает их более или менее подходящими для конкретных сценариев применения.

- YOLO выделяется своей высокой скоростью обработки и способностью обнаруживать объекты в реальном времени.
- SSD обеспечивает быструю и точную детекцию объектов с однократным просмотром изображения.
- Faster R-CNN обеспечивает точную детекцию и сегментацию, поддерживая множество классов объектов.
- DeepLab особенно эффективен в задачах сегментации изображений, учитывая контекст и структуру объектов.

Сравнение различных моделей YOLO

YOLO на данный момент наиболее популярное решение в «компьютерном зрении». Самым последним дополнением к семейству YOLO было подтверждено в январе 2023 года с выпуском YOLO-v8 от Ultralytics (также выпущен YOLO-v5).

На рисунке 1 показано, что при сравнении YOLO-v8 с YOLO-v5 и YOLO-v6, обученными на изображениях разрешением 640, все варианты YOLO-v8 выдают лучшую производительность при сходном количестве параметров, указывая на эффективность аппаратных средств и архитектурные реформы. [5]

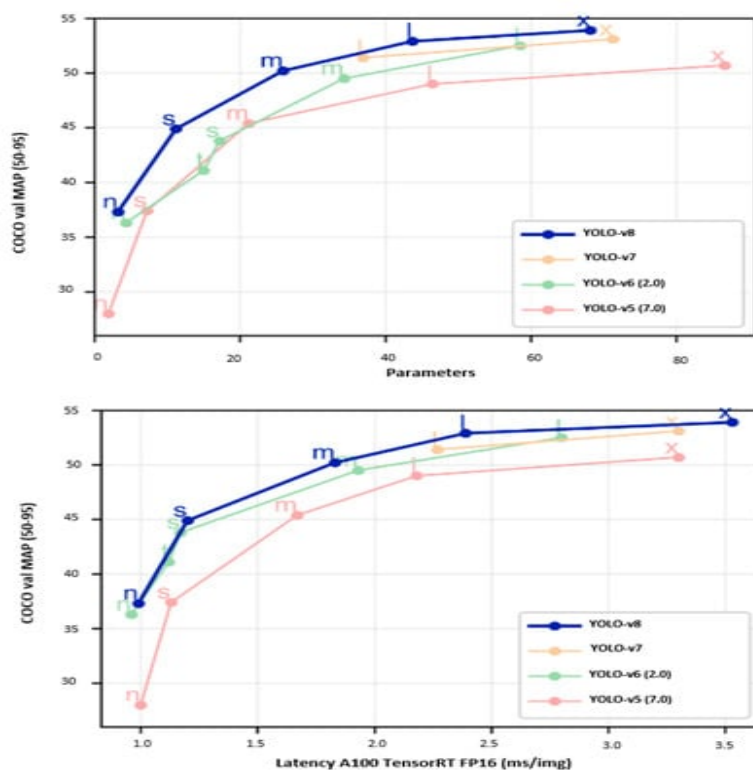


Рисунок 1 – Сравнение версий YOLO

Вывод по сравнению версий YOLO

В результате вычислительного эксперимента, выявлены следующие характеристики моделей: YOLO-v8 обеспечивает впечатляющую производительность в реальном времени. С увеличением размера модели – она либо равна, либо превосходит остальные модели. Использование YOLO-v8 в задачах сегментации – наиболее оптимальное решение, исходя из статистических показателей производительности и точности.

Заключение

В результате проведенного исследования дана развернутая характеристика популярным алгоритмам детекции и сегментации изображений. Описаны сферы и принципы их применения, преимущества и недостатки. Приведены сравнения моделей в виде графика популярнейшей архитектуры для данного типа задач – YOLO. Результаты подробно проанализированы и описаны.

Список литературы:

1. Бахтеев Д. В. Технология компьютерного зрения и возможности её применения в криминалистической деятельности. Сибирские правовые чтения: сборник научных статей / [отв. ред. В. И. Осейчук] ; Министерство науки и высшего образования Российской Федерации, Тюменский государственный университет, Институт государства и права. – Тюмень: Издательство Тюменского государственного университета, 2019.
2. Шапович, Е. Г. Виртуальный собеседник на базе искусственного интеллекта / Е. Г. Шапович // Второй Республиканский форум молодых ученых учреждений высшего образования Республики Беларусь: сб. науч. тр. / М-во образования Респ. Беларусь, М-во науки и высшего образования Рос. Федерации, Белорус.-Рос. ун-т; редкол.: В. М. Пашкевич (общ. ред.) [и др.]. – Могилев: Белорус.-Рос. ун-т, 2023. – С. 204-205
3. Ульянов Р.Д., Озерова М.И. Компьютерное зрение: Методы распознавания лиц

4. Горячкин Б.С., Китов М.А. КОМПЬЮТЕРНОЕ ЗРЕНИЕ // E-Scio. 2020. №9 (48). URL: <https://cyberleninka.ru/article/n/kompyuternoe-zrenie-1> (дата обращения: 15.12.2023).
5. Hussain, Muhammad. "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection." *Machines* 11, no. 7 (2023): 677. DOI: 10.3390/machines11070677

References:

1. Bakhteev D. V. Computer vision technology and the possibilities of its application in forensic activities. *Siberian Legal Readings : a collection of scientific articles* / [ed. V. I. Oseychuk] ; Ministry of Science and Higher Education of the Russian Federation, Tyumen State University, Institute of State and Law. – Tyumen : Publishing House of Tyumen State University, 2019.
2. Shapovich, E. G. Virtual interlocutor based on artificial intelligence / E. G. Shapovich // *The Second Republican Forum of young scientists of institutions of higher education of the Republic of Belarus: collection of scientific tr.* / Ministry of Education of the Republic of Belarus. Belarus, Ministry of Science and Higher Education of the Russian Federation, Belarus.-Russian University; editorial board: V. M. Pashkevich (general ed.)[et al.]. – Mogilev: Belarus.-Russian University, 2023. – pp. 204-205
3. Ulyanov R.D., Ozerova M.I. Computer vision: Methods of face recognition
4. Goryachkin B.S., Kitov M.A. COMPUTER VISION // E-Scio. 2020. №9 (48). URL: <https://cyberleninka.ru/article/n/kompyutern>
5. Hussain, Muhammad. "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection." *Machines* 11, no. 7 (2023): 677. DOI: 10.3390/machines11070677