



МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ПЕДАГОГИЧЕСКИЙ
УНИВЕРСИТЕТ ИМ. А. И. ГЕРЦЕНА»

**ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
И ТЕХНОЛОГИЧЕСКОГО ОБРАЗОВАНИЯ**
Кафедра информационных технологий и электронного обучения

Основная профессиональная образовательная программа
Направление подготовки 09.03.01 Информатика и вычислительная техника
Направленность (профиль) «Технологии разработки программного
обеспечения»
форма обучения – очная

Анализ источников

«Компьютерное зрение (Computer Vision)»

Обучающегося 4 курса
Лотуги Данилы Сергеевича

Санкт-Петербург
2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 ИСТОРИЧЕСКОЕ РАЗВИТИЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ	4
2 ОСНОВНЫЕ ЗАДАЧИ И МЕТОДЫ КОМПЬЮТЕРНОГО ЗРЕНИЯ	6
3 СОВРЕМЕННЫЕ ТЕХНОЛОГИИ И МОДЕЛИ В КОМПЬЮТЕРНОМ ЗРЕНИИ	8
4 МОДЕЛИ И ИНСТРУМЕНТЫ ДЛЯ РЕШЕНИЯ РЕАЛЬНЫХ ПРОБЛЕМ	10
5 ОЦЕНКА КАЧЕСТВА МОДЕЛИ И ТЕСТИРОВАНИЕ	12
6 ЗАДАЧИ И НЕДОЧЕТЫ В ОБЛАСТИ КОМПЬЮТЕРНОГО ЗРЕНИЯ .	13
7 БУДУЩЕЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ	14
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	15

ВВЕДЕНИЕ

Компьютерное зрение представляет собой одну из ключевых областей искусственного интеллекта, которая сосредоточена на извлечении информации из визуальных данных и их интерпретации, аналогично человеческому зрению. Его важность обусловлена растущей потребностью в автоматизированных системах анализа изображений и видео, которые находят применение в самых разных отраслях.

Основными задачами компьютерного зрения являются обнаружение и распознавание объектов, сегментация сцен, идентификация действий и жестов, а также построение описаний изображений и видео. Например, обнаружение объектов используется для определения местоположения заданных сущностей на изображении или в кадре видео, что является основой многих приложений в области безопасности и автономного транспорта. Распознавание объектов и образов позволяет идентифицировать категории или классы элементов, что находит применение в классификации изображений и анализа данных, как показано авторами в статье [1]. Важнейшей задачей также является реконструкция сцен — процесс воссоздания 3D-структуры окружающей среды из набора изображений, что активно используется в робототехнике и дополненной реальности.

Значимость компьютерного зрения также определяется его широким спектром применений. В медицинской отрасли оно обеспечивает анализ изображений с целью диагностики заболеваний, таких как онкологические или офтальмологические патологии. В автомобильной промышленности технологии компьютерного зрения являются ключевыми для реализации автономного управления, включая распознавание дорожных знаков, анализ полос движения и обнаружение пешеходов. Системы безопасности активно используют детекторы аномалий, идентификацию лиц и анализ видеопотоков для предотвращения угроз. Сельское хозяйство применяет данные технологии для мониторинга состояния посевов, классификации культур и оценки урожайности.

Эти примеры демонстрируют исключительное значение области, которая не только способствует автоматизации сложных процессов, но и открывает новые горизонты применения машинного обучения и анализа данных.

1 ИСТОРИЧЕСКОЕ РАЗВИТИЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Историческое развитие компьютерного зрения представляет собой сложный и многогранный процесс технологической эволюции, охватывающий несколько десятилетий научных исследований и практических разработок. Первоначальные методы компьютерного зрения основывались на аналоговой обработке изображений, где применялись базовые алгоритмы фильтрации и сегментации. В этот период основное внимание уделялось разработке математических моделей для описания и анализа визуальной информации, что заложило фундаментальные основы для последующего развития этой области, как отмечается в фундаментальном исследовании [2].

Значительный прогресс был достигнут с появлением цифровых методов обработки изображений, которые позволили реализовать более сложные алгоритмы анализа визуальных данных. В этот период были разработаны первые методы распознавания образов, основанные на статистическом анализе и геометрических преобразованиях. Эти ранние достижения существенно расширили возможности автоматизированного анализа изображений, что подробно описано в работе [2], где авторы детально анализируют эволюцию визуальных функций от базовых концепций к современным решениям.

Начало 2000-х годов ознаменовалось революционным переходом к методам машинного обучения в области компьютерного зрения. Этот период характеризовался появлением первых эффективных моделей глубокого обучения, которые продемонстрировали существенное превосходство над традиционными подходами в задачах распознавания образов и анализа изображений. Особенно важным этапом стало появление сверточных нейронных сетей, которые значительно улучшили точность и эффективность алгоритмов компьютерного зрения.

Последующие годы были отмечены серией технологических прорывов, среди которых особое место занимает создание архитектуры AlexNet в 2012 году, которая произвела революцию в области глубокого обучения для задач компьютерного зрения. За ней последовали другие значимые архитектуры, такие как VGG и ResNet, которые последовательно улучшали показатели точности распознавания и установили новые стандарты производительности в этой области. Как подчеркивается в исследовании [2], эти достижения существенно

расширили возможности практического применения технологий компьютерного зрения в различных областях, от медицинской диагностики до систем автономного управления.

Современный этап развития компьютерного зрения характеризуется интеграцией различных подходов и методов, включая глубокое обучение, статистический анализ и классические алгоритмы обработки изображений. Это позволяет создавать более совершенные системы, способные решать сложные задачи визуального анализа в реальном времени, что открывает новые перспективы для практического применения технологий компьютерного зрения в различных областях человеческой деятельности.

2 ОСНОВНЫЕ ЗАДАЧИ И МЕТОДЫ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Компьютерное зрение как область искусственного интеллекта охватывает широкий спектр задач и методов их решения. Современные подходы к обработке визуальной информации основаны на глубоком машинном обучении и включают различные специализированные алгоритмы для конкретных задач.

Фундаментальной задачей компьютерного зрения является обнаружение и распознавание объектов. В основе современных решений лежат сверточные нейронные сети (CNN), которые позволяют эффективно извлекать признаки из изображений. Развитие данного направления привело к появлению высокопроизводительных архитектур, таких как YOLO (You Only Look Once), SSD (Single Shot Detector) и Faster R-CNN, обеспечивающих как высокую точность, так и скорость обработки. Важно отметить, что эффективность алгоритмов машинного обучения для локализации и классификации объектов существенно зависит от методов описания контуров и их сравнения [3].

Сегментация изображений представляет собой более сложную задачу, требующую попиксельного анализа изображения. В данной области выделяют семантическую сегментацию, где каждому пикселю присваивается определенный класс, и инстанс-сегментацию, позволяющую различать отдельные экземпляры объектов одного класса. Архитектуры U-Net и DeepLab стали стандартом де-факто для решения задач сегментации, демонстрируя высокую точность на различных наборах данных. При этом применение методов медианной обработки существенно повышает эффективность структурных описаний при обработке изображений [4].

Особое место в компьютерном зрении занимают задачи распознавания лиц и жестов. Современные решения в этой области базируются на специализированных библиотеках, таких как OpenCV, Dlib и MTCNN, которые предоставляют готовые инструменты для детектирования и анализа лицевых характеристик. Эти библиотеки используют комбинацию классических алгоритмов компьютерного зрения и глубокого обучения для достижения высокой точности распознавания.

Отслеживание объектов в видеопотоке представляет собой динамическую задачу, требующую учета временной составляющей. Для её решения применяются различные подходы, включая фильтр Калмана, алгоритм Mean-Shift

и нейросетевой трекер GOTURN. Эти методы позволяют осуществлять надежное отслеживание объектов в реальном времени, учитывая различные факторы, такие как изменение освещения, перекрытие объектов и движение камеры.

3 СОВРЕМЕННЫЕ ТЕХНОЛОГИИ И МОДЕЛИ В КОМПЬЮТЕРНОМ ЗРЕНИИ

Современные технологии и модели в компьютерном зрении претерпели значительную эволюцию за последние годы, что привело к существенному прогрессу в области искусственного интеллекта и машинного обучения. Фундаментальную роль в этом развитии сыграли сверточные нейронные сети (CNN), которые стали основой для решения широкого спектра задач компьютерного зрения [5].

В контексте глубоких нейросетей особое внимание следует уделить архитектурам, специализированным для анализа изображений. ResNet, DenseNet и Inception представляют собой высокоэффективные архитектурные решения, каждое из которых обладает уникальными преимуществами в обработке визуальной информации. Практическая реализация данных архитектур демонстрирует значительное улучшение точности распознавания образов и эффективности обработки изображений в реальных условиях, что подтверждается результатами внедрения автоматизированных систем компьютерного зрения [6].

Генеративные состязательные сети (GANs) открыли новые горизонты в области синтеза и модификации изображений. Данный класс алгоритмов позволяет не только создавать высококачественные синтетические изображения, но и выполнять сложные операции по трансформации существующих визуальных данных. Особую значимость GANs приобретают в контексте аугментации данных и создания синтетических датасетов для обучения других моделей компьютерного зрения.

Значительным прорывом в области компьютерного зрения стало появление архитектуры Vision Transformers (ViT). Данный подход, заимствованный из области обработки естественного языка, продемонстрировал исключительную эффективность в задачах анализа изображений. ViT обладает способностью к более глубокому пониманию глобальных зависимостей в изображениях благодаря механизму самовнимания, что особенно важно при работе с комплексными визуальными данными.

В условиях ограниченности доступных данных особую актуальность приобретают методы Transfer Learning и Fine-tuning. Данные подходы позволяют эффективно адаптировать предварительно обученные модели к специфич-

ческим задачам, существенно сокращая требования к объему обучающей выборки и вычислительным ресурсам. Это особенно актуально в промышленных приложениях, где сбор большого объема размеченных данных может быть затруднен или экономически нецелесообразен.

4 МОДЕЛИ И ИНСТРУМЕНТЫ ДЛЯ РЕШЕНИЯ РЕАЛЬНЫХ ПРОБЛЕМ

В современном мире машинного обучения и компьютерного зрения существует множество инструментов и библиотек, позволяющих эффективно решать широкий спектр практических задач. Ведущие позиции среди фреймворков занимают TensorFlow и PyTorch, предоставляющие обширные возможности для разработки и обучения нейронных сетей. TensorFlow, разработанный компанией Google, обеспечивает гибкую экосистему инструментов, библиотек и ресурсов сообщества, что делает его особенно привлекательным для промышленной разработки. PyTorch, поддерживаемый Facebook, отличается более интуитивным интерфейсом и динамическим построением вычислительных графов, что особенно ценится в исследовательской среде.

Для задач компьютерного зрения библиотека OpenCV предоставляет богатый набор оптимизированных алгоритмов обработки изображений и видео. В сочетании с Keras, высокоуровневым API для создания нейронных сетей, эти инструменты позволяют эффективно решать широкий спектр прикладных задач. Практическое применение этих технологий в различных отраслях демонстрирует значительный потенциал, что подтверждается успешными внедрениями в сельскохозяйственной сфере [7].

Особое внимание уделяется оптимизации моделей для работы на устройствах с ограниченными вычислительными ресурсами. TensorFlow Lite представляет собой облегченную версию фреймворка, специально разработанную для мобильных и встраиваемых устройств. Данное решение позволяет эффективно развертывать модели машинного обучения на периферийных устройствах, обеспечивая обработку данных в реальном времени при минимальных затратах вычислительных ресурсов.

Обработка видеопотока в реальном времени представляет собой особый класс задач, требующий оптимального баланса между производительностью и точностью. Современные решения, основанные на комбинации OpenCV и оптимизированных нейросетевых моделей, позволяют достигать высокой производительности при сохранении приемлемой точности распознавания объектов и анализа сцен. Это открывает широкие возможности для применения в системах

видеонаблюдения, промышленной автоматизации и других областях, требующих оперативной обработки визуальной информации.

5 ОЦЕНКА КАЧЕСТВА МОДЕЛИ И ТЕСТИРОВАНИЕ

Оценка качества моделей и тестирование представляет собой критически важный этап в разработке систем компьютерного зрения. Процесс оценки требует систематического подхода к анализу различных метрик производительности, что подтверждается исследованиями в области практического применения моделей компьютерного зрения [8].

Основополагающими метриками для оценки эффективности моделей являются accuracy (точность), precision (прецизионность), recall (полнота) и F1-score (F1-мера), представляющая собой гармоническое среднее между precision и recall. Особое место занимает метрика mean Average Precision (mAP), которая широко применяется в задачах детекции объектов и позволяет оценить качество работы модели при различных порогах уверенности. Современные исследования демонстрируют возможности оптимизации данных метрик даже в условиях работы с моделями по принципу черного ящика [9].

Для улучшения качества моделей применяются различные методы, среди которых ключевую роль играют методы расширения данных (data augmentation) и нормализация. Data augmentation позволяет искусственно увеличить разнообразие обучающей выборки путем применения различных преобразований к исходным изображениям, что способствует повышению обобщающей способности модели. Нормализация данных, в свою очередь, обеспечивает стабильность процесса обучения и улучшает сходимость модели. Как показано авторами в работе [8], качество входных изображений и их предварительная обработка существенно влияют на надежность результатов работы модели.

В контексте практического применения моделей компьютерного зрения особую значимость приобретает процесс тестирования, который должен включать в себя не только оценку стандартных метрик, но и анализ работы модели в различных условиях эксплуатации. Это становится особенно актуальным при внедрении систем компьютерного зрения в промышленные приложения, где требуется обеспечить стабильную работу модели при различных условиях освещения, ракурсах съемки и других вариациях входных данных.

6 ЗАДАЧИ И НЕДОЧЕТЫ В ОБЛАСТИ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Компьютерное зрение как область искусственного интеллекта сталкивается с рядом существенных проблем, требующих комплексного решения. Фундаментальным вызовом является зависимость производительности моделей от качества тренировочных данных. Недостаточное разнообразие или ограниченный объем обучающей выборки может привести к существенному снижению точности распознавания и классификации изображений в реальных условиях.

Проблема интерпретируемости моделей компьютерного зрения представляет собой отдельный существенный вызов. Современные архитектуры нейронных сетей, несмотря на высокую эффективность, зачастую функционируют как "черный ящик" что затрудняет понимание процесса принятия решений и выявление потенциальных ошибок в их работе.

В контексте этических проблем особую актуальность приобретают вопросы приватности и защиты персональных данных при использовании систем видеонаблюдения и распознавания лиц. Отдельного внимания заслуживает проблема манипуляций с изображениями, в частности, создание и распространение deepfakes, что может иметь серьезные социальные и политические последствия.

Значительным техническим вызовом остается разработка устойчивых моделей для работы в условиях различной освещенности, изменения ракурса съемки и наличия шумов. Как показано авторами в статье [10], распространенные повреждения изображений, такие как изменение освещенности и размытость, могут существенно влиять на производительность моделей компьютерного зрения в реальных приложениях. Исследователи подчеркивают необходимость разработки более надежных методов, способных эффективно работать в условиях подобных искажений.

Отдельную сложность представляет проблема частичной видимости объектов на изображениях, возникающая вследствие их перекрытия или обрезки. Данная проблема требует разработки специализированных алгоритмов, способных корректно идентифицировать и классифицировать объекты даже при наличии неполной информации о их структуре и форме.

7 БУДУЩЕЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Будущее компьютерного зрения характеризуется стремительным развитием технологий и появлением инновационных подходов к решению сложных задач визуального восприятия. Современные исследования демонстрируют значительный потенциал в использовании мульти-агентных систем, где несколько алгоритмических единиц совместно работают над анализом и интерпретацией визуальных данных. Эта тенденция открывает новые возможности для повышения точности и надежности систем компьютерного зрения.

Ключевым аспектом развития области является её тесная интеграция с другими направлениями искусственного интеллекта. Особенно важным становится взаимодействие с автоматическим обучением и робототехникой, что создает синергетический эффект в решении комплексных задач. В частности, как показано авторами в работе [11], диффузионные модели демонстрируют значительный потенциал для увеличения объема обучающих данных, что существенно повышает производительность и надежность систем машинного обучения в задачах компьютерного зрения.

Современные тенденции в развитии компьютерного зрения включают растущий интерес к объяснимому искусственному интеллекту, что особенно важно для практического применения технологий в критически важных областях. Разработка улучшенных архитектур нейронных сетей направлена на достижение более высокой точности и надежности решений. Значительный прогресс наблюдается в области оценки качества систем компьютерного зрения, что подтверждается исследованием [12], где представлены инновационные методологии тестирования и оценки производительности систем распознавания текста на изображениях.

Будущее компьютерного зрения также тесно связано с развитием методов анализа данных. Интеграция современных подходов к обработке больших данных с алгоритмами компьютерного зрения позволяет создавать более эффективные системы, способные обрабатывать и анализировать визуальную информацию в масштабах, ранее считавшихся недостижимыми. Это открывает новые перспективы для применения технологий компьютерного зрения в различных областях, от медицинской диагностики до систем автономного управления транспортными средствами.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Дмитриевна Ч.А., Эрфановна Б.К., Александровна К.М., Сергеевна С.А., Михайловна Ю.В. Основные задачи компьютерного зрения // — 2019. — URL: <https://api.semanticscholar.org/CorpusID:213038111>.
2. Weinmann M. Visual Features - From Early Concepts to Modern Computer Vision // Advanced Topics in Computer Vision. — 2013. — URL: <https://api.semanticscholar.org/CorpusID:41270009>.
3. Макаров М.А. ИССЛЕДОВАНИЕ КОМПОЗИЦИИ АЛГОРИТМОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ, БАЗИРУЮЩИХСЯ НА МАШИННОМ ОБУЧЕНИИ, ДЛЯ РЕШЕНИЯ ЗАДАЧИ ЛОКАЛИЗАЦИИ И КЛАССИФИКАЦИИ ОБЪЕКТОВ // Труды Института системного анализа Российской академии наук. — 2019. — URL: <https://api.semanticscholar.org/CorpusID:126276426>.
4. Gorokhovatsky V. METHODS OF RELEVANT IMAGES SEARCH BASED ON THE MEDIAN PROCESSING OF STRUCTURAL DESCRIPTIONS // Radio Electronics, Computer Science, Control. — 2017. — URL: <https://api.semanticscholar.org/CorpusID:66975400>.
5. Michalski P., Ruszczak B., Tomaszewski M. Convolutional Neural Networks Implementations for Computer Vision // IC BCI. — 2018. — URL: <https://api.semanticscholar.org/CorpusID:64870529>.
6. Клемышев И.М., Колчин С.М., Лебедев С.С., Старков С.О. Автономная многоракурсная нейросетевая система компьютерного зрения для автоматизации контроля ручных операций // Успехи кибернетики / Russian Journal of Cybernetics. — 2024. — URL: <https://api.semanticscholar.org/CorpusID:273816753>.
7. Губашева Х., Абдулаева С. ОПЫТ ИСПОЛЬЗОВАНИЯ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ ОБЛЕГЧЕНИЯ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ // Тенденции развития естественных наук в современном информационном пространстве и их применение в агробιοтехнологиях. — 2022. — URL: <https://api.semanticscholar.org/CorpusID:255077871>.

8. *Finardi F.A.R., Holanda G.M. de, Adorni C.Y., Nader M.V.P.D., Pinheiro K.R.* Evaluating the application of a computer vision model in the customer service chatbot of an electric utility // 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). — 2021. — P. 1–5. — URL: <https://api.semanticscholar.org/CorpusID:243946239>.
9. *Huang C., Zhai S., Guo P., Susskind J.* MetricOpt: Learning to Optimize Black-Box Evaluation Metrics. — 2021. — arXiv: [2104.10631 \[cs.LG\]](https://arxiv.org/abs/2104.10631). — URL: <https://arxiv.org/abs/2104.10631>.
10. *Wang S., Veldhuis R.N.J., Strisciuglio N.* The Robustness of Computer Vision Models against Common Corruptions: a Survey // ArXiv. — 2023. — Vol. abs/2305.06024. — URL: <https://api.semanticscholar.org/CorpusID:263871668>.
11. *Alimisis P., Mademlis I., Radoglou-Grammatikis P., Sarigiannidis P., Papadopoulos G.T.* Advances in Diffusion Models for Image Data Augmentation: A Review of Methods, Models, Evaluation Metrics and Future Research Directions. — 2024. — arXiv: [2407.04103 \[cs.CV\]](https://arxiv.org/abs/2407.04103). — URL: <https://arxiv.org/abs/2407.04103>.
12. *Shu J., Miu B.-J., Chang E., Gao J., Liu J.* Computer Vision Intelligence Test Modeling and Generation: A Case Study on Smart OCR // 2024 IEEE International Conference on Artificial Intelligence Testing (AITest). — IEEE, 07/2024. — P. 21–28. — URL: <http://dx.doi.org/10.1109/AITest62860.2024.00011>.