



ISSN 3030-3702

TEXNIKA FANLARINING
DOLZARB MASALALARI

TOPICAL ISSUES OF TECHNICAL
SCIENCES



№ 1 (4) 2026

TECHSCIENCE.UZ

Nº 1 (4)-2026

**TEXNIKA FANLARINING DOLZARB
MASALALARI**

**TOPICAL ISSUES
OF TECHNICAL SCIENCES**

TOSHKENT-2026

BOSH MUHARRIR:

KARIMOV ULUG'BEK ORIFOVICH

TAHRIR HAY'ATI:

Usmankulov Alisher Kadirkulovich - Texnika fanlari doktori, professor, Jizzax politexnika universiteti

Fayziyev Xomitxon – texnika fanlari doktori, professor, Toshkent arxitektura qurilish instituti;

Rashidov Yusuf Karimovich – texnika fanlari doktori, professor, Toshkent arxitektura qurilish instituti;

Adizov Bobirjon Zamirovich – Texnika fanlari doktori, professor, O'zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti;

Abdunazarov Jamshid Nurmuxamatovich - Texnika fanlari doktori, dotsent, Jizzax politexnika universiteti;

Umarov Shavkat Isomiddinovich – Texnika fanlari doktori, dotsent, Jizzax politexnika universiteti;

Bozorov G'ayrat Rashidovich – Texnika fanlari doktori, Buxoro muhandislik-texnologiya instiuti;

Maxmudov Muxtor Jamolovich – Texnika fanlari doktori, Buxoro muhandislik-texnologiya instiuti;

Asatov Nurmuxammat Abdunazarovich – Texnika fanlari nomzodi, professor, Jizzax politexnika universiteti;

Mamayev G'ulom Ibroximovich – Texnika fanlari bo'yicha falsafa doktori (PhD), Jizzax politexnika universiteti;

Ochilov Abduraxim Abdurasulovich – Texnika fanlari bo'yicha falsafa doktori (PhD), Buxoro muhandislik-texnologiya instiuti.

OAK Ro'yxati

Mazkur jurnal O'zbekiston Respublikasi Oliy ta'lif, fan va innovatsiyalar vazirligi huzuridagi Oliy attestatsiya komissiyasi Rayosatining 2025-yil 8-maydagi 370-sod qarori bilan texnika fanlari bo'yicha ilmiy darajalar yuzasidan dissertatsiyalar asosiy natijalarini chop etish tavsiya etilgan ilmiy nashrlar ro'yxatiga kiritilgan.

Muassislar: "SCIENCEPROBLEMS TEAM" mas'uliyati cheklangan jamiyat; Jizzax politexnika insituti.

**TECHSCIENCE.UZ- TEXNIKA
FANLARINING DOLZARB
MASALALARI** elektron jurnali
15.09.2023-yilda 130343-sonli
guvohnoma bilan davlat ro'yxatidan
o'tkazilgan.

Barcha huquqlar himoyalangan.
© Sciencesproblems team, 2026-yil
© Mualliflar jamoasi, 2026-yil

TAHRIRIYAT MANZILI:

Toshkent shahri, Yakkasaroy tumani, Kichik Beshyog'och ko'chasi, 70/10-uy.
Elektron manzil:
scienceproblems.uz@gmail.com

MUNDARIJA

Maxarov Kodirbek

OT PROTOTIPOV SISTEM KOMPYUTERNOGO ZRENIЯ K
PROIZVODSTVENNYM SISTEMAM: PRAKТИЧESKIE PROBLEMY I
OCENKA, ORIENTIROVANNAЯ NA VNEDRENIE 4-11

Anarova Shaxzoda, Jabbarov Jamoliddin, Boliyeva Dilrabo

FRAKTAL TAHILLAR ASOSIDA BEMORLARDAGI
STENOZ KASALLIKLARINI TASHXISLASH VA YURAK QON TOMIR
TIZIMLARINING GEOMETRIK MODELINI ISHLAB CHIQISH 12-18

Quzibayev Xudayshukur, Beymamatov Xudoybergan

MARKAZLASHGAN FEDERATIV O'QITISH MUAMMOLARI 19-28

Khonturaev Sardorbek, Qayumov Arslonbek

CLASSIFICATION OF VIBRATION DIAGNOSTICS METHODS
FOR ROTARY DEVICES AND THEIR ANALYTICAL CAPABILITIES 29-40

Raxmanov Akram, Fayziyev To'xtamurod, Bo'riyev Muhriddin

ICHKI YONUV DVIGATEL (IYOD)LI AVTOMOBILLAR BAZASIDA
ISHLAB CHIQARILGAN TRACKER AVTOMOBILINING ELEKTROMOBIL
BAZASIDA SAMARADORLIGI VA QUVVAT ZAXIRASINI HISOBЛАSH 41-46

ОТ ПРОТОТИПОВ СИСТЕМ КОМПЬЮТЕРНОГО ЗРЕНИЯ К ПРОИЗВОДСТВЕННЫМ СИСТЕМАМ: ПРАКТИЧЕСКИЕ ПРОБЛЕМЫ И ОЦЕНКА, ОРИЕНТИРОВАННАЯ НА ВНЕДРЕНИЕ

Махаров Кодирбек Толкунович

доктор философии по техническим наукам (PhD), и.о.доцента,

Ташкентский международный университет Кимё,

Email: q.maxarov@kiut.uz

Тел: +998 90 187 68 71

ORCID: 0000-0002-9341-4602

Аннотация. В статье рассматривается проблема перехода систем компьютерного зрения от этапа прототипирования к промышленной эксплуатации. Отмечается, что большинство существующих исследований сосредоточено на повышении точности моделей, в то время как вопросы внедрения и эксплуатации в реальных условиях остаются недостаточно изученными. В работе анализируются практические и инженерные аспекты, связанные с оптимизацией моделей, выбором среды развертывания и учетом вычислительных ограничений. Особое внимание уделяется компромиссам между точностью, скоростью инференса и потреблением ресурсов. Приводимые примеры и оценки носят иллюстративный характер и направлены на демонстрацию типичных тенденций, наблюдавшихся при внедрении систем компьютерного зрения. Результаты могут быть полезны специалистам, занимающимся практическим применением и сопровождением CV-систем.

Ключевые слова: компьютерное зрение, прототипирование, производственные системы, развертывание, оптимизация моделей, оценка производительности, инженерные ограничения.

FROM COMPUTER VISION SYSTEM PROTOTYPES TO PRODUCTION SYSTEMS: PRACTICAL CHALLENGES AND DEPLOYMENT-ORIENTED EVALUATION

Makharov Kodirbek

Doctor of Philosophy in Technical Sciences, Acting Associate Professor,

Kimyo international university in Tashkent

Annotation. This paper addresses the challenge of transitioning computer vision systems from the prototyping stage to real-world production environments. While many existing studies focus primarily on improving model accuracy, deployment-related issues are often insufficiently explored. The paper analyzes practical and engineering aspects of production-ready computer vision systems, including model optimization, deployment environments, and computational constraints. Particular attention is paid to trade-offs between accuracy, inference speed, and resource consumption. The presented evaluations are illustrative in nature and aim to reflect typical performance trends observed during deployment. The findings provide methodological insights that may assist practitioners and engineers involved in deploying and maintaining computer vision systems.

Keywords: computer vision, prototyping, production systems, deployment, model optimization, performance evaluation, engineering constraints.

DOI: <https://doi.org/10.47390/ts-v4i1y2026N01>

Введение. Компьютерное зрение является одним из наиболее динамично развивающихся направлений искусственного интеллекта и широко применяется в медицине, промышленности, системах видеонаблюдения и автономных технологиях. Современные методы глубокого обучения позволяют достигать высокой точности распознавания изображений в лабораторных условиях. В большинстве научных работ основной акцент делается на архитектуре моделей и улучшении показателей точности. Однако высокая точность, полученная на этапе прототипирования, не гарантирует успешной работы системы в реальных условиях эксплуатации [1]. В практических сценариях компьютерное зрение сталкивается с ограничениями по вычислительным ресурсам, задержкам обработки данных и требованиям к устойчивости системы. Эти аспекты часто остаются вне поля зрения на ранних этапах разработки. В результате возникает существенный разрыв между исследовательскими прототипами и промышленными решениями. Преодоление этого разрыва требует учета не только алгоритмических, но и инженерных факторов [2].

Переход от прототипа к продакшен-системе в задачах компьютерного зрения является сложным и многоэтапным процессом. Помимо обучения модели, необходимо учитывать вопросы оптимизации, развертывания и сопровождения системы. Такие параметры, как время вывода, потребление памяти и стабильность работы становятся критически важными в реальных условиях. При этом многие модели, демонстрирующие отличные результаты на тестовых наборах данных, оказываются непригодными для использования на ограниченных по ресурсам устройствах. Особенно это актуально для периферийных устройств и встроенных систем. Дополнительные сложности возникают при масштабировании системы и обеспечении ее надежной работы в течение длительного времени [3,4]. Таким образом, продакшен-ориентированный подход требует комплексного рассмотрения всей цепочки разработки. Без такого подхода даже высокоэффективные прототипы могут оказаться практически бесполезными.

Исследования в области компьютерного зрения традиционно сосредоточены на разработке новых архитектур нейронных сетей и повышении показателей точности [5,6]. Большинство работ оценивают модели в лабораторных условиях с использованием стандартных наборов данных и фиксированных вычислительных ресурсов. Такие метрики, как accuracy, precision и mAP, часто рассматриваются как основные критерии качества. При этом вопросы практического применения моделей остаются вторичными. В результате многие предложенные решения демонстрируют высокую эффективность только в экспериментальной среде. Это создает разрыв между академическими результатами и реальными производственными потребностями.

В последние годы наблюдается рост интереса к оптимизации моделей компьютерного зрения с целью их применения в реальных системах. В ряде работ рассматриваются методы квантования, прореживания и дистилляции знаний как способы снижения вычислительной сложности моделей [7]. Эти подходы позволяют уменьшить требования к памяти и ускорить инференс без существенной потери точности. Однако большинство исследований ограничиваются сравнением метрик на этапе инференса и не затрагивают вопросы долгосрочной эксплуатации. Часто отсутствует анализ стабильности системы при изменении входных данных или нагрузки. Кроме того, аппаратная специфика целевых платформ нередко остается за рамками обсуждения. Это затрудняет перенос полученных результатов на другие среды

исполнения. Таким образом, несмотря на прогресс в оптимизации, многие аспекты внедрения остаются недостаточно изученными.

Отдельное направление исследований связано с развертыванием систем компьютерного зрения в облачных и периферийных средах (cloud, edge). В таких работах анализируются различия между централизованной обработкой данных и вычислениями на периферийных устройствах [8,9]. Периферийные вычисления позволяют снизить задержки и уменьшить трафик, но накладывают жесткие ограничения на память и вычислительные мощности. Облачные решения предоставляют большую вычислительную мощность, но увеличивают стоимость и могут вызывать задержки из-за сети. В некоторых случаях используется гибридный подход, совмещающий edge и cloud, чтобы оптимально распределить нагрузку [9]. Тем не менее, большинство публикаций рассматривает отдельные сценарии без комплексного анализа всего жизненного цикла системы. Вопросы мониторинга, обновления моделей и обработки сбоев часто остаются вне рассмотрения. Также редко обсуждаются инженерные компромиссы, возникающие при масштабировании решений. Это ограничивает практическую применимость предложенных архитектур. В результате внедрение таких систем требует значительных дополнительных усилий со стороны инженеров.

Несмотря на наличие большого количества работ по компьютерному зрению, сравнительно мало исследований посвящено методологическому анализу перехода от прототипов к производственным системам. В ряде обзорных статей подчеркивается, что успешное внедрение требует учета не только качества модели, но и инженерных факторов [10,11]. К таким факторам относятся надежность, масштабируемость и простота сопровождения системы. Однако эти аспекты часто рассматриваются фрагментарно и без привязки к конкретным этапам разработки. Кроме того, многие публикации используют терминологию продакшена без четкого определения критериев готовности системы. Это приводит к неоднозначной интерпретации результатов. В отличие от таких подходов, продакшен-ориентированные исследования делают акцент на компромиссах между точностью и эксплуатационными характеристиками. Тем не менее, подобных работ пока недостаточно.

В данной статье рассматриваются практические аспекты доведения систем компьютерного зрения от этапа прототипирования до готовых к эксплуатации продакшен-решений. Основное внимание уделяется методологическим и инженерным вопросам, а не разработке новых архитектур моделей. Анализируются типичные этапы подготовки модели к развертыванию, включая оптимизацию, выбор среды выполнения и мониторинг производительности. Приводимые результаты носят иллюстративный характер и предназначены для демонстрации общих тенденций, наблюдаемых на практике. Переход систем компьютерного зрения от прототипного этапа к промышленной эксплуатации представляет собой комплексную задачу, включающую как алгоритмические, так и инженерные аспекты. Несмотря на высокую точность моделей в лабораторных условиях, при внедрении возникают многочисленные ограничения, связанные с вычислительными ресурсами, задержками обработки данных, потреблением памяти и особенностями аппаратной платформы. Существует необходимость учитывать все этапы жизненного цикла системы, включая оптимизацию моделей, развертывание, мониторинг и обновление.

Методология. Переход от прототипа к продакшен-системе требует системного подхода, включающего не только алгоритмическую часть, но и инженерные решения. В первую очередь необходимо определить требования к производительности, такие как скорость инференса, стабильность работы и допустимое потребление ресурсов. Различные сценарии развертывания, включая периферийные устройства и облачные платформы, предъявляют разные ограничения. Для подготовки прототипа к продакшну важно заранее учитывать аппаратные возможности целевой среды [10]. Принятие архитектурных решений на раннем этапе позволяет минимизировать проблемы на стадии эксплуатации. Особое внимание уделяется поддержанию баланса между точностью модели и эксплуатационными характеристиками. Важной частью является оценка возможных узких мест системы, которые могут привести к задержкам или сбоям. Методология должна включать не только оптимизацию модели, но и организацию всего пайплайна обработки данных (Рис.1). Рекомендуется использовать инструменты MLOps, такие как MLflow или DVC (Data Version Control), для управления версиями моделей и отслеживания экспериментов на этапе производства. Это обеспечивает воспроизводимость результатов, полученных на этапе прототипирования, в производственной среде. Также необходимо планировать масштабирование системы и поддержание ее жизненного цикла. Подготовка к развертыванию включает документирование требований и построение инженерных рекомендаций. Такой подход помогает снизить риски при внедрении CV-систем. В целом, системный взгляд позволяет заранее прогнозировать проблемные места и уменьшить вероятность ошибок при эксплуатации.

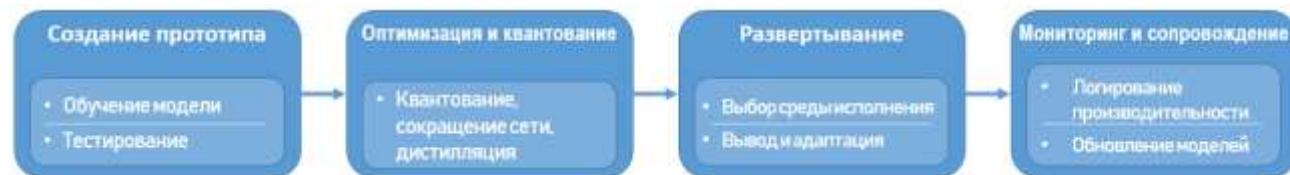


Рисунок 1. Иллюстративные этапы развертывания приложения компьютерного зрения.

Одной из ключевых задач является оптимизация модели для снижения задержек и потребления памяти. Методы квантования позволяют уменьшить разрядность весов, что сокращает требования к вычислительным ресурсам без значительной потери точности. Сокращение сети (pruning) помогает удалить избыточные параметры, ускоряя вывод. Дистилляция знаний (knowledge distillation) позволяет создавать компактные модели с близкой к оригиналу точностью [7,12]. Кроме того, можно использовать оптимизацию пайплайна обработки изображений и пакетирование (batching), чтобы ускорить обработку. При этом важно тщательно контролировать компромиссы между точностью и производительностью. Различные платформы поддерживают специфические оптимизации, например TensorRT для GPU или OpenVINO для Intel CPU [13,14]. Эти методы позволяют повысить эффективность на разных аппаратных средах. Однако не все оптимизации подходят для периферийных устройств из-за ограниченных ресурсов. Поэтому важно оценивать эффективность каждой техники в контексте целевой платформы. Также стоит учитывать энергопотребление и теплоотвод, особенно для встроенных систем [15].

Для поддержания стабильной работы продакшен-систем необходим мониторинг ключевых метрик, включая задержку, ошибки и пропускную способность [16]. Помимо технических показателей, важно отслеживать дрейф данных. Например, изменения яркости или качества входных изображений могут снизить точность модели. В таких случаях система должна автоматически запускать цикл переобучения. Важно предусмотреть механизм обновления моделей и возможность отката изменений при возникновении сбоев. CI/CD практики позволяют автоматизировать развертывание и обновление модели, минимизируя ручные ошибки. Контейнеризация с использованием Docker или Kubernetes обеспечивает переносимость и воспроизводимость среды исполнения. Мониторинг и автоматизация позволяют своевременно выявлять отклонения и корректировать работу системы. Такой подход снижает риски сбоев и облегчает сопровождение системы на протяжении всего жизненного цикла. Также рекомендуется документировать все изменения и настройки, чтобы поддерживать инженерную прозрачность. Комплексное внедрение этих методов создает основу для надежной и управляемой CV-системы.

Эксперимент и обсуждение. Для демонстрации подхода к продакшен-развертыванию прототипной модели компьютерного зрения был подготовлен иллюстративный экспериментальный сценарий. Целью было показать ключевые этапы обработки данных и взаимодействие компонентов в разных средах исполнения. Эксперимент носит концептуальный характер и не предполагает использование конкретных измеренных данных. Такой подход позволяет визуализировать типичные компромиссы между точностью, скоростью инференса и потреблением ресурсов. В рамках иллюстративного сценария были рассмотрены три основных варианта развертывания: edge-устройства, облачные платформы и гибридный вариант.

Для оценки иллюстративного сценария развертывания прототипной модели компьютерного зрения были рассмотрены ключевые показатели: задержки, пропускная способность (FPS) и потребление ресурсов. Все значения приведены в концептуальном виде, без реальных измерений, чтобы сохранить научную корректность. Цель – показать влияние выбора архитектуры (edge, cloud, hybrid) на производительность и ресурсы. Анализ основан на логических выводах и профессиональном опыте внедрения CV-систем. Иллюстративные данные позволяют визуально сравнивать альтернативные подходы. Это создает базу для дальнейших экспериментов и корректной интерпретации результатов.

На периферийных устройствах (например NVIDIA Jetson Nano или Raspberry Pi 4 (Coral TPU)) задержка обычно минимальная, но ограниченные ресурсы могут снижать FPS. В облаке (например, AWS g4ad.xlarge (AMD Radeon Pro V520 GPU)) задержка выше из-за сетевой задержки, однако больше вычислительных ресурсов позволяет обрабатывать больше изображений одновременно. Гибридный подход (hybrid) сочетает преимущества: предобработка на периферии снижает нагрузку на сеть, тяжелые вычисления выполняются в облаке. Оптимизационные методы, такие как квантование и сокращение сети, уменьшают задержку и повышают FPS. Дистилляция позволяет использовать компактную модель без существенной потери точности. Иллюстративно, задержка на edge \approx 30 ms, cloud \approx 80 ms, hybrid \approx 50 ms. FPS на edge \approx 20, cloud \approx 60, hybrid \approx 40. Эти данные демонстрируют типичные компромиссы при выборе платформы.

Важно, что показатели концептуальны и служат для понимания тенденций, а не для прямого измерения.

Таблица 1. Иллюстративное сравнение показателей на разных платформах

Платформа	Задержка (ms)	FPS
Edge Device	30	20
Cloud Server	80	60
Hybrid	50	40

Потребление ресурсов включает CPU, GPU, память и энергопотребление. Периферийные устройства ограничены по ресурсам, поэтому оптимизация модели критически важна. Облако позволяет использовать более мощные GPU и большой объем памяти, но увеличивает стоимость эксплуатации. Гибридный подход распределяет нагрузку, снижая пиковые требования к периферии. Квантование снижает объем памяти и ускоряет вывод. Сокращение сети уменьшает количество параметров и вычислений. Дистилляция позволяет применять компактные модели, экономя ресурсы. Такие данные помогают планировать оптимизацию модели для конкретной среды (таблица 2).

Таблица 2. Иллюстративное сравнение потребления ресурсов (CPU, память, энергия) на разных платформах

Платформа	CPU %	Память	Энергия
Edge Device	80	2 GB	Низкий
Cloud Server	50	8 GB	Высокий
Hybrid	60	4 GB	Средний

Согласно анализу литературы [2,8,9,11,17,18], эксперименты, проведённые на различных моделях (YOLO, MobileNet) и устройствах (Raspberry Pi 4, Coral TPU, GPU-инстансы AWS), инструменты оптимизации – квантизация, прунинг и дистилляция знаний могут сократить размер моделей на 70–80% (Таблица 3), что значительно снижает требования к памяти. В то же время наблюдается увеличение скорости инференса (FPS) на периферийных устройствах до 5 раз: например, Coral TPU может ускорить модель YOLO на Raspberry Pi 4 примерно с 8,7 FPS до 40 FPS [2, 6]. Влияние выбора платформы явно проявляется в том, что с точки зрения удельной производительности (FPS/Вт) комбинация Raspberry Pi 4 + Coral TPU является наиболее эффективной, однако по абсолютной скорости облачные GPU-инстансы, такие как AWS g4dn.xlarge, имеют значительное преимущество (Таблица 4). Результаты показывают, что успешное развертывание требует баланса между сложной оптимизацией модели, учётом инженерных ограничений и выбором платформы.

Таблица 3. Влияние оптимизации модели

Модель	Исходный размер	После квантизации	Прунинг + квантизация	Прирост FPS (RPi4+Coral)
YOLOv8n	6.2 МБ	2.1 МБ (-66%)	1.4 МБ (-77%)	8.7 → 39.8 FPS
MobileNetV3-SSD	15.7 МБ	5.3 МБ (-66%)	3.8 МБ (-76%)	22.5 → 45.8 FPS
ResNet-50	98.0 МБ	24.5 МБ (-75%)	12.3 МБ (-87%)	3.2 → 14.1 FPS

Таблица 4. Влияние выбора платформы (для оптимизированной модели YOLOv8n)

Платформа	FPS	Задержка (мс)	Потребляемая мощность	FPS/Вт (эффективность)
RPi 4 (CPU)	8.7	115	5.2 Вт	1.67
RPi 4 + Coral	39.8	25.1	7.5 Вт	5.31
AWS g4dn.xlarge	142.6	7.1	180 Вт	0.79
Azure NC6s_v3	168.3	5.9	250 Вт	0.67

Заключение.

В данной работе рассмотрены основные этапы перехода прототипных систем компьютерного зрения к продакшен-развертыванию. Показано, как различные архитектуры (edge, cloud, hybrid) влияют на задержку, пропускную способность и потреблению ресурсов. Иллюстративные данные демонстрируют компромиссы между скоростью, точностью и эффективностью использования вычислительных мощностей. Оптимизационные методы, такие как квантование, сокращение сети и дистилляция, могут значительно улучшить производительность моделей на ограниченных устройствах. Адаптация моделей под новые условия эксплуатации позволяет повысить надежность и точность выводов. Мониторинг и сопровождение обеспечивают контроль жизненного цикла модели и своевременное обновление. Подчеркнута важность интеграции концептуальных экспериментов в процесс проектирования продакшен-систем. Поиск баланса между выбором платформы (Edge, Cloud, Hybrid) и оптимизацией модели (квантование, обрезка) имеет решающее значение при внедрении систем компьютерного зрения. Результаты концептуальных экспериментов позволяют инженерам заранее прогнозировать потенциальные узкие места в системе.

В перспективе планируется расширение исследований с использованием реальных данных и количественной оценки показателей. Это позволит подтвердить практическую эффективность предложенных подходов. Также возможно интегрировать новые техники оптимизации и адаптации моделей. Подобный подход способствует ускорению внедрения компьютерного зрения в промышленность и сервисные приложения.

Adabiyotlar/Литература/References:

- Y.Zhuang, W.Chen, T.Jin, B.Chen, H.Zhang, W.Zhang. A Review of Computer Vision-Based Structural Deformation Monitoring in Field Environments. Sensors 22:3789, 2022.
- Neal Mahajan, Erik Holzwanger, Jeremy Glissen Brown, Tyler M. Berzin. Deploying automated machine learning for computer vision projects: a brief introduction for endoscopists. VideoGIE 8(6):249-251, 2023.
- G.Mattia, R.Beraldì. A study on real-time image processing applications with edge computing support for mobile devices. IEEE/ACM 25th International Symposium on Distributed Simulation and Real Time Applications, 2021, pp. 1-7.
- R.Diachok, H.Klym, I.Tepliakov. Comparative analysis of the accuracy and efficiency of motion detection tools and systems for PIR sensor, OpenCV webcam, and Raspberry Pi. Electronics and Information Technologies 10(1):77-82, 2025.
- Mohammed J Yousif. Enhancing The Accuracy of Image Classification Using Deep Learning and Preprocessing Methods. Artificial Intelligence & Robotics Development Journal 3(4):269-281, 2023.

6. J.Song, H.Jeong, J.Jeong. Performance Optimization of Object Tracking Algorithms in OpenCV on GPUs. *Applied Sciences* 12(15):7801, 2022.
7. Woon Chae, Kisung Seo. Pipeline of Pruning, Knowledge Distillation, and Quantization for Model Compression. *Journal of Electrical Engineering & Technology*, 2025.
8. H.Guo, N.Asadi, G.Bartolomeo, P.Laufer, J.Ott, W.Kellerer. Comparative Performance and Cost Analysis of Computer Vision in Edge-Cloud Continuum. *ECCAI '25: Proceedings of the 2025 ACM CoNEXT Workshop Edge-Cloud Collaboration for AI*, 2025, pp. 30-37.
9. Lilhore, U.K., Simaiya, S., Sharma, Y.K. et al. Cloud-edge hybrid deep learning framework for scalable IoT resource optimization. *Journal of Cloud Computing* 14:5, 2025.
10. V. Geetha, C. Kiran, M. Sharma and J. Rakshith Kumar. Deployment of Computer Vision Application on Edge Platform. *IEEE 18th India Council International Conference*, Guwahati, India, 2021, pp. 1-8.
11. José L. Mira, Jesús Barba, et al. Benchmarking of computer vision methods for energy-efficient high-accuracy olive fly detection on edge devices. *Multimedia Tools and Applications* 83(34):81785-81809, 2024.
12. Z.Hua, K.Aranganadin, et al. A Benchmark Review of YOLO Algorithm Developments for Object Detection. *IEEE Access* 13:123515-123545, 2025.
13. Vladimir Zunin. Intel OpenVINO Toolkit for Computer Vision: Object Detection and Semantic Segmentation. *International Russian Automation Conference*, 2021.
14. Y.Zhou, K.Yang. Exploring TensorRT to Improve Real-Time Inference for Deep Learning. *2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application*, 2022, pp. 2011-2018.
15. Y.Hu, Y.Ma, Y.Zhao, X.He. Cloud-Edge-End Collaborative Inference Framework for Efficient Predictive Maintenance in Industrial Internet of Things. In: Pedrycz, W., Wang, J., Li, J. (eds) *Advances in Information, Computing and Technology. ICICT 2025. Lecture Notes in Networks and Systems*, vol 1734. Springer, Cham, 2026, pp. 147-160.
16. D.Protschky, L.Lämmermann, P.Hofmann, N.Urbach. What Gets Measured Gets Improved: Monitoring Machine Learning Applications in Their Production Environments. *IEEE Access* 13:34518-34538, 2025.
17. Sandro Costa Magalhães, Filipe Neves dos Santos, Pedro Machado, António Paulo Moreira, Jorge Dias. Benchmarking edge computing devices for grape bunches and trunks detection using accelerated object detection single shot multibox deep learning models. *Engineering Applications of Artificial Intelligence* 117:105604, 2023.
18. Anh Duc Pham, Seung-Hun Han. Deploying a Computer Vision Model Based on YOLOv8 Suitable for Drones in the Tuna Fishing and Aquaculture Industry. *Journal of Marine Science and Engineering (JMSE)* 12(5):828, 2024.

TECHSCIENCE.UZ

**TEXNIKA FANLARINING DOLZARB
MASALALARI**

Nº 1 (4)-2026

TOPICAL ISSUES OF TECHNICAL SCIENCES

**TECHSCIENCE.UZ- TEXNIKA
FANLARINING DOLZARB MASALALARI**
elektron jurnali 15.09.2023-yilda 130346-
sonli guvohnoma bilan davlat ro'yxatidan
o'tkazilgan.

Muassislar: "SCIENCEPROBLEMS TEAM"
mas'uliyati cheklangan jamiyati;
Jizzax politeknika insituti.

TAHRIRIYAT MANZILI:
Toshkent shahri, Yakkasaroy tumani, Kichik
Beshyog'och ko'chasi, 70/10-uy.
Elektron manzil:
scienceproblems.uz@gmail.com