

УДК 681.518.5

УНИВЕРСАЛЬНАЯ МОБИЛЬНАЯ ПЛАТФОРМА AUTOGENNETMOBILE ДЛЯ АВТОМАТИЗАЦИИ ПРИКЛАДНЫХ ЗАДАЧ

Ф.Д. Гапоняко, В.А. Соболевский (Санкт-Петербург)

Введение

Принятие эффективных управленческих решений в производственных условиях неразрывно связано с задачами квалиметрии [1]. Однако практическое применение методов квалиметрического анализа зачастую сдерживается разрывом между средой разработки моделей и их конечным использованием специалистами для оперативного принятия решений [2-4]. Существующие инструментальные средства автоматизации моделирования, как правило, представляют собой сложные программные комплексы, требующие стационарных рабочих станций и специальных навыков, что ограничивает их применение в реальных производственных условиях [5, 6].

Перспективным направлением для решения этой проблемы является создание системы, способной перенести эксплуатацию моделей на мобильные платформы [7, 8]. Повсеместное распространение высокопроизводительных смартфонов и планшетов открывает возможность для использования технологий на мобильных устройствах, позволяя специалистам в предметной области использовать сложные модели для поддержки принятия решений в режиме реального времени. Такой подход позволяет интегрировать результаты моделирования непосредственно в рабочие процессы, повышая их эффективность и снижая зависимость от удалённых вычислительных центров [9 -11].

Однако прямой перенос традиционных средств моделирования на мобильные устройства часто оказывается невозможным из-за их высокой вычислительной сложности и требований к ресурсам [12, 13]. Решение этой проблемы лежит в использовании нейронных сетей в качестве компактного и эффективного ядра модели [14, 15]. После этапа обучения свёрточная нейронная сеть (СНС) представляет собой высокоэффективный вычислительный инструмент анализа, что идеально подходит для развёртывания в средах с ограниченными ресурсами [16]. Оптимизированные нейросетевые модели являются легковесными и идеально подходят для развёртывания в средах с ограниченными ресурсами.

В настоящей работе представлена архитектура и реализация мобильной платформы AutoGenNetMobile, спроектированной как универсальное инструментальное средство для динамического использования различных моделей на базе СНС. Ключевое достоинство концепции заключается в том, что само приложение способно загружать и применять модели, представленные в виде СНС. Эти модели, созданные и обученные на внешнем серверном комплексе, могут решать широкий спектр задач, от контроля качества продукции до инспекции производственных площадок.

Целью данной работы является описание архитектуры такой мобильной платформы и демонстрация её практической применимости для автоматизации принятия решений по результатам моделирования. В рамках исследования будет показан полный жизненный цикл системы, от запроса списка доступных моделей и их загрузки с сервера до их исполнения на мобильном устройстве и визуализации результатов в реальном времени. В качестве экспериментального подтверждения концепции будет продемонстрировано последовательное решение двух различных задач на одном и том же приложении без внесения изменений в его программный код, что доказывает состоятельность предложенного подхода.

2. Материалы и методы решения задач, принятые допущения

2.1. Концепция и архитектура полимодельного комплекса

В основе данного исследования лежит разработка и апробация программного комплекса, реализованного в рамках клиент-серверной архитектуры. Выбор такой архитектуры продиктован различием между процессами создания сложных моделей и их последующей эксплуатации. Традиционные подходы к внедрению систем искусственного интеллекта часто представляют собой узкоспециализированные решения, что создает необходимость в создании новых систем под новые задачи. Предлагаемая концепция направлена на декомпозицию системы на два слабосвязанных, но эффективно взаимодействующих компонента, что позволяет абстрагировать всю сложность машинного обучения на стороне сервера, предоставляя специалисту в предметной области простой и гибкий инструмент для автоматизации его практической деятельности.

В рамках настоящего исследования серверным компонентом является программный комплекс AutoGenNet [17, 18], однако с точки зрения архитектуры он представляет собой абстрактный источник моделей, предоставляющий стандартизированный интерфейс для взаимодействия. Именно на серверной стороне сосредоточены все ресурсоёмкие операции по подготовке наборов данных и автоматизированному структурно-параметрическому синтезу нейросетевых архитектур, их обучению [19, 20].

2.2. Клиент AutoGenNetMobile

Клиентский компонент, реализованный в виде мобильной платформы AutoGenNetMobile, представляет собой интерактивную среду для эксплуатации прикладных моделей. Приложение спроектировано как универсальный терминал доступа к репозиторию моделей, не привязанный к какой-либо конкретной задаче или предметной области. Его архитектура состоит из трех модулей и обеспечивает динамическую загрузку, конфигурацию и использование различных, взаимозаменяемых СНС.

2.2.1. Модуль управления моделями

Его функциональность включает отправку запросов на сервер для получения структурированного списка всех доступных для загрузки моделей. На основе полученной информации модуль предоставляет пользователю интерфейс для выбора необходимой модели, инициирует её загрузку и обеспечивает сохранение файла в защищенной внутренней директории приложения. Также в модуле реализован функционал для просмотра и управления уже загруженными моделями. Такой подход обеспечивает возможность работы в офлайн-режиме, используя ранее сохранённые в кэше модели.

2.2.2. Среда исполнения моделей

Вычислительным ядром приложения является среда исполнения, построенная на базе кроссплатформенного интерпретатора TensorFlow Lite [21]. Выбор данного фреймворка обусловлен его высокой степенью оптимизации для мобильных платформ, а также поддержкой аппаратного ускорения и малым размером библиотеки. Среда исполнения полностью автоматизирует предварительную обработку данных, захват кадра с камеры устройства, преобразование цветового пространства из нативного формата YUV в RGB, изменение размера изображения до разрешения, требуемого конкретной моделью, и нормализацию значений пикселей. Подготовленный тензор подаётся на вход интерпретатору, который выполняет прямой проход по графу вычислений нейронной сети и возвращает набор выходных тензоров, содержащих

результаты анализа. Важно отметить, что обращение к внутренней логике модели происходит полностью прозрачно для среды исполнения, что и обеспечивает универсальность и полимодельность всей платформы.

2.2.3. Модуль визуализации результатов

Эта подсистема выполняет задачи пост-обработки и наглядного представления результатов конечному пользователю. Модуль получает на вход необработанные тензоры от среды исполнения и преобразует их в графические примитивы, которые в реальном времени накладываются на видеопоток с камеры. Для корректной интерпретации выходных данных модели модуль использует конфигурационный файл, который загружается вместе с файлом самой модели. Этот файл сопоставляет индексы, возвращаемые нейросетью с метками, что в свою очередь обеспечивает гибкость визуализации и позволяет приложению корректно отображать результаты для любой задачи без необходимости внесения изменений в программный код.

2.3. Модели для экспериментальной апробации

Для демонстрации и верификации такого подхода, реализованного в платформе AutoGenNetMobile, были подготовлены две принципиально различные модели. Их выбор был продиктован необходимостью показать универсальность платформы на задачах из разных предметных областей промышленного контроля и агротехнологий.

Первая модель архитектуры YOLO [22], обученная на наборе данных "Construction Item Computer Vision Dataset" [23], представляет собой модель для задачи визуального аудита строительной площадки. Актуальность этой задачи обусловлена необходимостью оперативного разделения ценных вторичных ресурсов от строительных отходов, требующих утилизации, а также своевременного выявления потенциально опасных элементов. Мобильное приложение в данном контексте выступает как оперативный инструмент для инженера или прораба, позволяющий проводить экспресс-аудит непосредственно на объекте и принимать своевременные решения по управлению ресурсами и обеспечению безопасности без необходимости ручного учёта или лабораторного анализа.

Вторая модель архитектуры YOLO, обученная на наборе данных "Fruit and Vegetable Disease (Healthy vs Rotten)" [24]. Практическая значимость этой задачи заключается в предотвращении потерь на этапах хранения и транспортировки, так как своевременное выявление и удаление дефектных плодов предотвращает порчу всей партии. Мобильное приложение здесь служит инструментом для оперативного контроля качества на этапах оценки состояния урожая в поле, инспекции партий на складе или при приёмке товара от поставщика. Это позволяет принимать быстрые и объективные решения.

Эти две модели будут поочередно использоваться в рамках одной и той же инсталляции приложения AutoGenNetMobile для экспериментального подтверждения его гибкости, универсальности и практической применимости в качестве среды исполнения. Для оценки и верификации предложенной мобильной платформы, а также для подтверждения состоятельности предложенного подхода была разработана и проведена серия экспериментальных исследований.

2.4. Цель и задачи эксперимента

Основной целью экспериментальной апробации является не столько оценка качества конкретных нейросетевых моделей, сколько валидация самой мобильной платформы AutoGenNetMobile как универсальной и эффективной инструментальной средства для использования разнородных прикладных моделей. Первостепенно было необходимо продемонстрировать функциональную состоятельность подхода, то есть экспериментально доказать, что одно и то же приложение способно без модификации

программного кода динамически загружать и корректно выполнять две принципиально разные модели, предназначенные для решения задач из различных предметных областей. А также оценить качество работы каждой из загруженных моделей с использованием стандартных метрик для задач детекции объектов.

2.5. Протокол проведения эксперимента

Все экспериментальные исследования проводились на мобильном устройстве Samsung Galaxy A15 с процессором MediaTek Helio G99, 8 ГБ оперативной памяти, под управлением операционной системы Android 14. Для проведения экспериментов первоначально использовался доступ к серверной части, которая предоставляла интерфейс для загрузки двух предварительно обученных моделей компьютерного зрения. После того как модели были загружены и сохранены в локальном хранилище устройства, все последующие операции по их оценке и исполнению проводились в полностью автономном режиме, без необходимости дальнейшего сетевого взаимодействия с сервером. Такой протокол эксперимента позволяет оценить как функциональность клиент-серверного взаимодействия на этапе получения моделей, так и производительность приложения в его основном, офлайн-режиме работы.

Для количественной оценки качества работы моделей использовался комплексный подход, включающий как метрики классификации, так и метрики детекции объектов.

Метриками для оценки качества классификации послужили:

- Precision – характеризует долю верных положительных предсказаний среди всех объектов, которые модель определила как положительные;
- Recall – определяет долю верных положительных предсказаний среди всех реально существующих положительных объектов;
- F1-score – гармоническое среднее между precision и recall, служит сбалансированной метрикой для оценки классификационной способности модели;
- матрица ошибок – инструмент для детального анализа ошибок, который показывает, между какими именно классами модель испытывает трудности.

Метрикой оценки качества являлась mAP (mean Average Precision) – основная интегральная метрика для оценки качества моделей детекции. Она рассчитывается как среднее значение метрики Average Precision (AP) по всем классам и комплексно отражает способность модели как правильно классифицировать и точно локализовывать объекты. В работе приводятся значения для mAP50 и mAP50-95, которые оценивают производительность модели при разных требованиях к точности локализации.

3. Результаты

В результате была проведена последовательная загрузка и тестирование двух прикладных моделей на одной и той же инсталляции мобильной платформы AutoGenNetMobile. Всесторонняя оценка качества и производительности для каждой модели была выполнена на независимых тестовых выборках. Полученные количественные результаты сведены в табл. 1.

Таблица 1. Сводные результаты оценки результатов моделей

Метрика	YOLOv11 Construction Item	YOLOv11 vegetables
Precision	0,850	0,949
Recall	0,644	0,902
F1-score	0,720 (при 0.552)	0,920 (при 0.739)
mAP50	0,716	0,971
mAP50-95	0,584	0,968

Как видно из табл. 1 модель YOLOv11_Construction_Item продемонстрировала умеренное качество. Общий F1-score (рис. 1) составил 0,72, что отражает значительный дисбаланс между показателями точности и полноты. Precision достигла высокого значения 0,85, в то время как Recall оказалась значительно ниже – 0,644. Детальный анализ матрицы ошибок (рис. 2) показывает, что причиной этого является неверная классификация большого числа реальных объектов как фона, то есть их пропуск. Это свидетельствует о том, что модель консервативна и делает предсказания только при высокой степени уверенности.

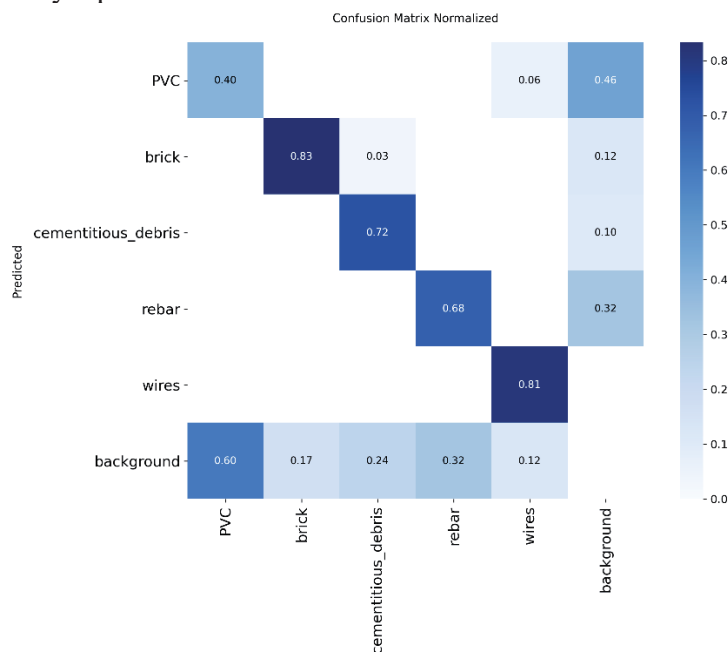


Рис. 1. Матрица ошибок для модели YOLOv11_Construction_Item

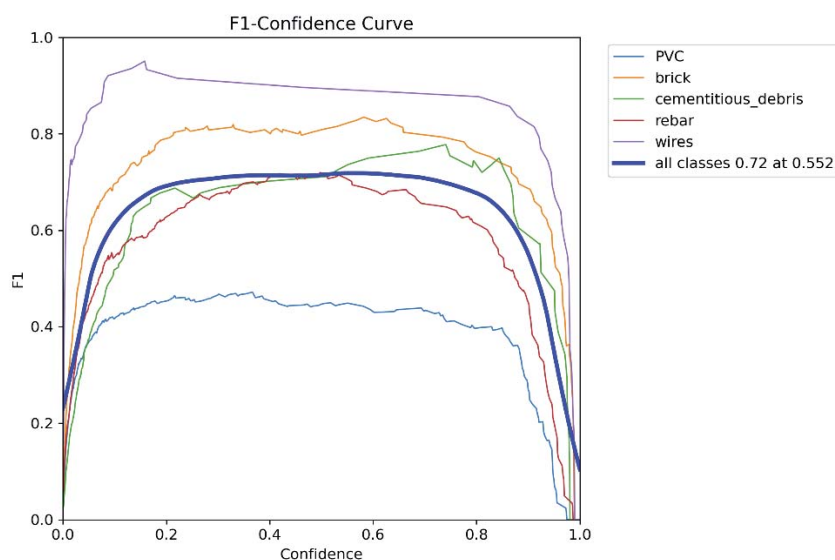


Рис. 2. График F1-score для модели YOLOv11_Construction_Item

В отличие от предыдущей, модель YOLOv11_vegetables продемонстрировала значительно более высокие и сбалансированные результаты. Общее значение F1-score (рис. 4) составило 0,92, что является высоким показателем качества. Комплексная метрика детекции mAP50 достигла 0,971, подтверждая способность модели не только правильно классифицировать, но и точно локализовывать объекты. Матрица ошибок (рис.3) показывает стабильно высокие результаты для большинства классов за

исключением некоторого снижения полноты для `healthy_potato`, что связано с его визуальным сходством с дефектными экземплярами.

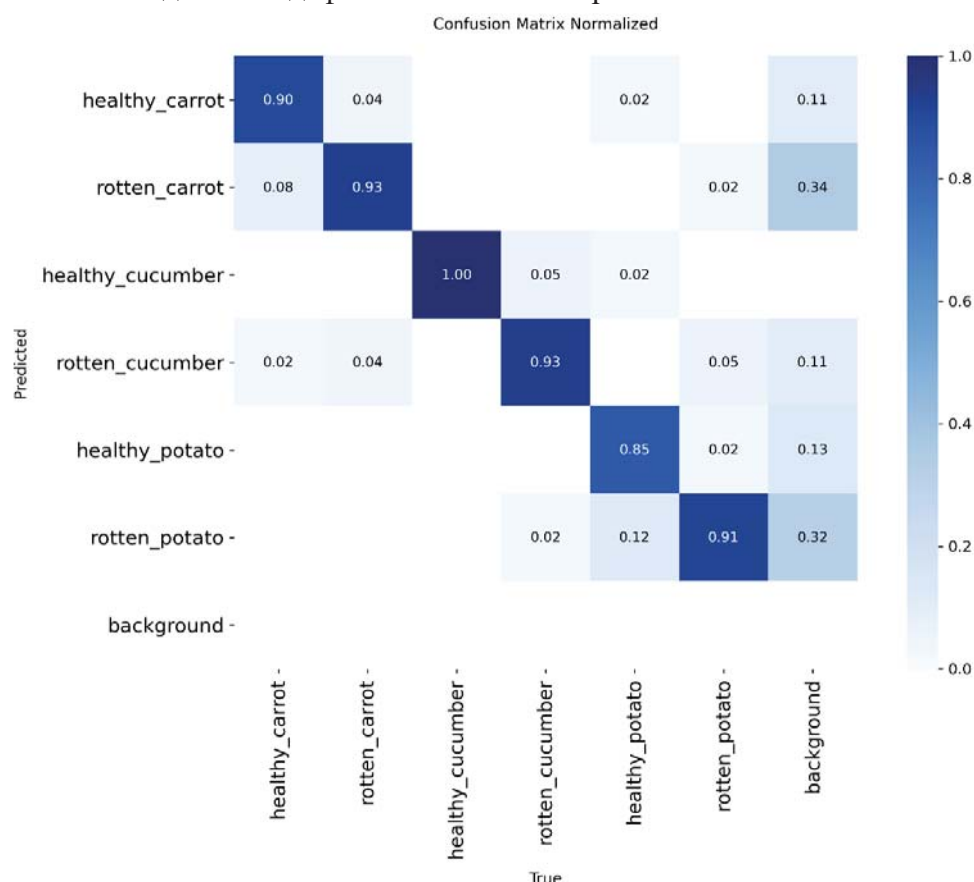


Рис. 3. Матрица ошибок для модели YOLOv11_vegetables

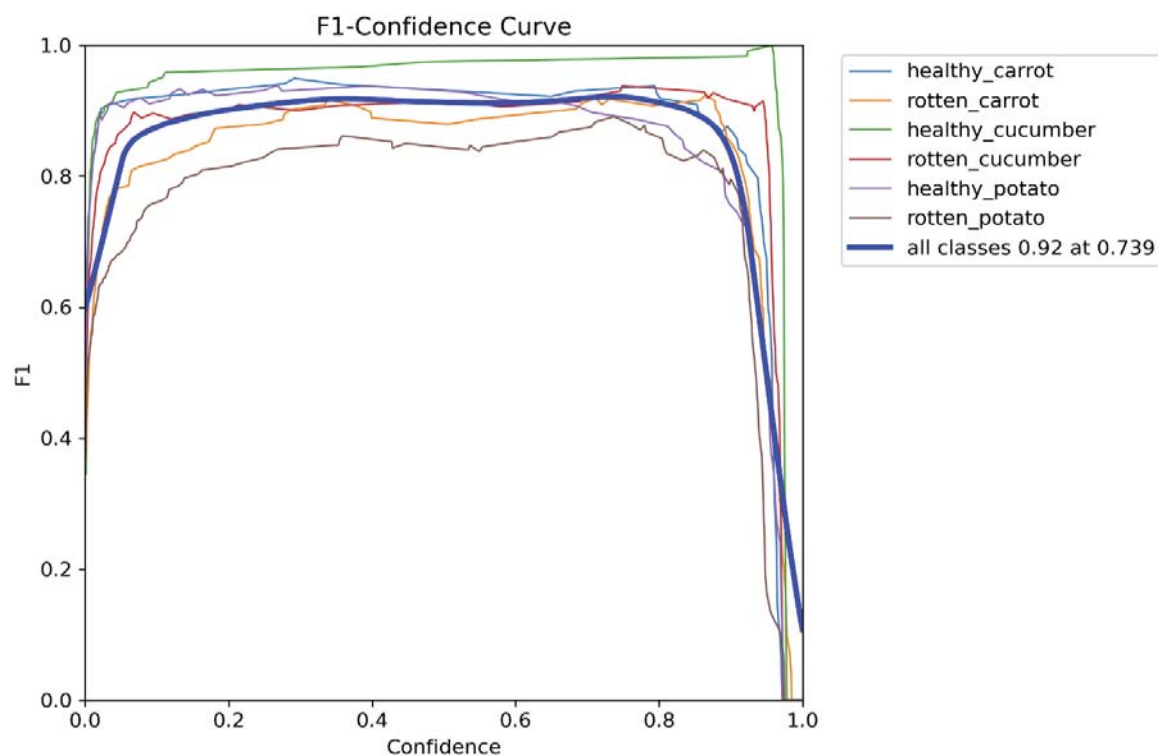


Рис. 4. График F1-score для модели YOLOv11_vegetables

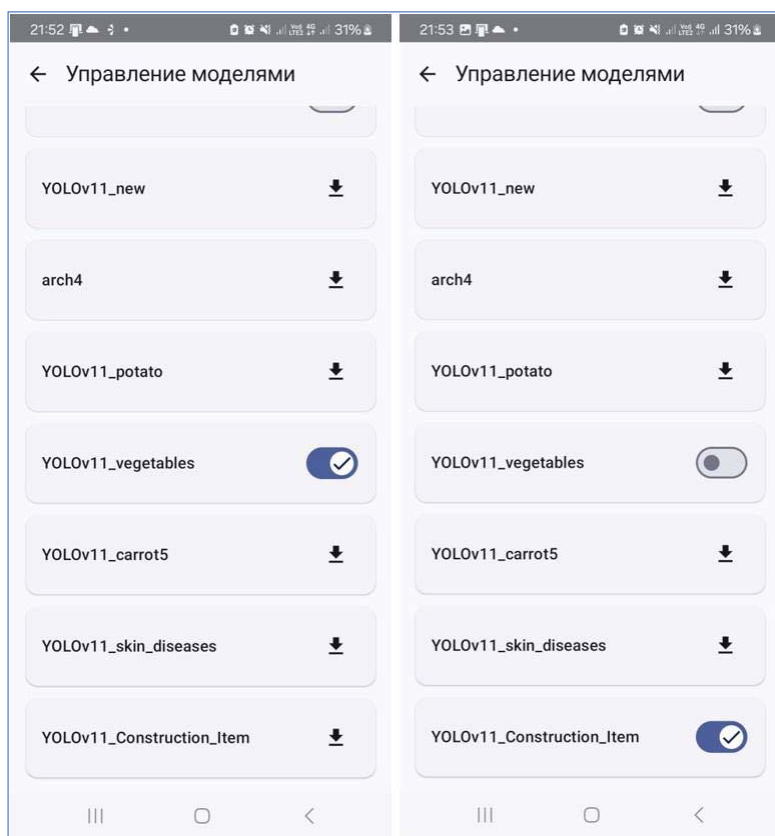


Рис. 5. Интерфейс выбора моделей AutoGenNetMobile

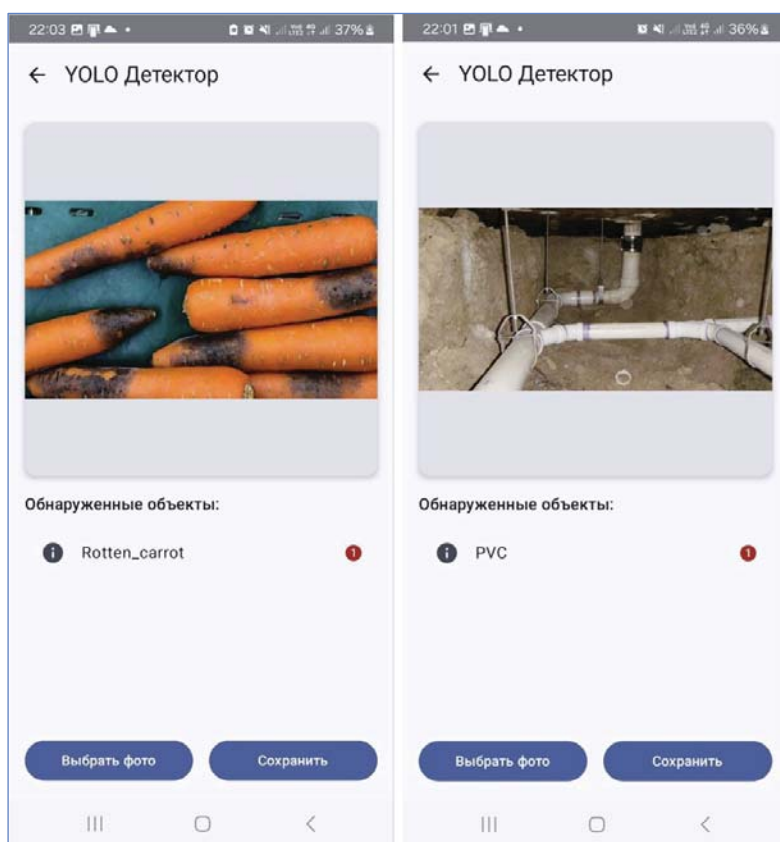


Рис. 6. Пример работы программы AutoGenNetMobile

Сравнение результатов двух экспериментов позволяет сделать ключевой вывод. Разница в итоговых метриках обусловлена не недостатками самой мобильной платформы, а сложностью и качеством исходных данных. Набор данных со строительными материалами является значительно более сложным для анализа, нежели чем с овощами.

Тот факт, что одна и та же мобильная платформа смогла без каких-либо модификаций успешно загрузить и исполнить обе модели, корректно отразив их результаты, доказывает её гибкость и универсальность. Система продемонстрировала, что способна служить средой для исполнения как высокоточных, так и менее совершенных аналитических моделей, адекватно отражая их внутреннее качество. Это подтверждает, что предложенная архитектура является эффективным и универсальным инструментом для практического применения широкого спектра прикладных моделей.

Разработанная мобильная платформа выступает в роли инструмента для автоматизации задач квалитметрии. Она решает проблему трудоёмкого ручного контроля и субъективной оценки, заменяя их быстрым и объективным анализом. Автоматизируя этот процесс, приложение позволяет за секунды получить количественные и качественные характеристики объекта, которые служат основанием для оперативного принятия решений: отсортировать партию товара, задокументировать состояние объекта или оценить объём материалов. Таким образом, платформа обеспечивает переход от трудоёмких ручных процедур к оперативной поддержке принятия решений, основанной на актуальной информации, получаемой непосредственно с физического объекта в реальном времени.

4. Заключение

В настоящей работе была представлена и экспериментально апробирована мобильная платформа AutoGenNetMobile, основной вклад которой заключается в снижении порога входа для практического применения технологий компьютерного зрения специалистами в предметной области.

Ключевое достоинство разработанной архитектуры заключается в том, что она полностью автоматизирует и абстрагирует от пользователя всю техническую сложность. Функциональность динамической загрузки и автоматической конфигурации превращает узкоспециализированные нейросетевые модели в простые, взаимозаменяемые инструменты. Специалисту не требуется знание форматов моделей или процедур их настройки, он просто выбирает из списка задачу, которую необходимо решить.

Экспериментальная проверка подтвердила, что платформа является универсальной и способна без модификации программного кода выполнять разнородные задачи. Таким образом, исследование доказывает, что предложенная архитектура является эффективным решением для оперативного развертывания средств автоматизированного анализа в реальных производственных условиях, передавая сложный аналитический инструмент непосредственно в руки специалистов. Дальнейшие направления развития включают расширение библиотеки доступных моделей для решения новых прикладных задач

Благодарности. Исследование, проведенное в данной работе, частично финансировалось бюджетной темой FFZF-2025-0020.

Литература

1. **Микони С.В., Соколов Б.В., Юсупов Р.М.** (2018). Квалитметрия моделей и полимодельных комплексов. Москва: РАН. 314 с. ISBN 978-5-907036-32-1. <https://doi.org/10.31857/S9785907036321000001>.

2. **Angira Sharma, Edward Kosasih, Jie Zhang, Alexandra Brintrup, Anisoara Calinescu.** Digital Twins: State of the art theory and practice, challenges, and open research questions, *Journal of Industrial Information Integration*, Volume 30, 2022, 100383, ISSN 2452-414X, <https://doi.org/10.1016/j.jii.2022.100383>.
3. **Fiona Kattenstroth, Jan-Philipp Disselkamp, Jonas Lick, Roman Dumitrescu.** Challenges in the implementation of simulation models for the digital factory twin - a systematic literature review, *Procedia CIRP*, Vol. 128, 2024, P.442-447, ISSN 2212-8271, <https://doi.org/10.1016/j.procir.2024.07.052>.
4. **Fei Yu, Chen Zheng.** Tools, application areas and challenges of factory simulation in Small and Medium-Sized Enterprises – A Review, *Procedia CIRP*, Volume 104, 2021. P.399-404, ISSN 2212-8271, <https://doi.org/10.1016/j.procir.2021.11.067>.
5. **Tom Lloyd Garwood, Ben Richard Hughes, Michael R. Oates, Dominic O'Connor, Ruby Hughes.** A review of energy simulation tools for the manufacturing sector, *Renewable and Sustainable Energy Reviews*, Volume 81, Part 1, 2018. P. 895-911, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2017.08.063>.
6. **Chris J Turner, Wolfgang Garn.** Next generation DES simulation: A research agenda for human centric manufacturing systems, *Journal of Industrial Information Integration*, Volume 28, 2022, 100354, ISSN 2452-414X, <https://doi.org/10.1016/j.jii.2022.100354>.
7. **Aadharsh Roshan Nandhakumar, Ayush Baranwal, Priyanshukumar Choudhary, Muhammed Golec, Sukhpal Singh Gill.** EdgeAISim: A toolkit for simulation and modelling of AI models in edge computing environments, *Measurement: Sensors*, Volume 31, 2024, 100939, ISSN 2665-9174, <https://doi.org/10.1016/j.measen.2023.100939>.
8. **R. S. Lais, J. Fitzner, Y.-K. Lee, V. Struckmann.** Open-sourced modeling and simulating tools for decision-makers during an emerging pandemic or epidemic – Systematic evaluation of utility and usability: A scoping review update, *Dialogues in Health*, Vol. 5, 2024, 100189, ISSN 2772-6533, <https://doi.org/10.1016/j.dialog.2024.100189>.
9. **Leonardo Maretto, Maurizio Faccio, Daria Battini.** The adoption of digital technologies in the manufacturing world and their evaluation: A systematic review of real-life case studies and future research agenda, *Journal of Manufacturing Systems*, Volume 68, 2023, P. 576-600, ISSN 0278-6125, <https://doi.org/10.1016/j.jmsy.2023.05.009>.
10. **E. Hoos, C. Gröger, B. Mitschang.** Mobile Apps in engineering: a process-driven analysis of business potentials and technical challenges, *Procedia CIRP*, Vol. 33, 2015, Pages 17-22, ISSN 2212-8271, <https://doi.org/10.1016/j.procir.2015.06.005>.
11. **Mohd Javaid, Abid Haleem, Ravi Pratap Singh, Shanay Rab, Rajiv Suman.** Upgrading the manufacturing sector via applications of Industrial Internet of Things (IIoT), *Sensors International*, Vol. 2, 2021, 100129, ISSN 2666-3511, <https://doi.org/10.1016/j.sintl.2021.100129>.
12. **Yingchun Wang, Jingyi Wang, Weizhan Zhang, Yufeng Zhan, Song Guo, Qinghua Zheng, Xuanyu Wang.** A survey on deploying mobile deep learning applications: A systemic and technical perspective, *Digital Communications and Networks*, Vol. 8, Issue 1, 2022, P. 1-17, ISSN 2352-8648, <https://doi.org/10.1016/j.dcan.2021.06.001>.
13. **Charafeddine Mechlikh, Abd El Heq Silem, Zahra Safavifar, Fatemeh Golpayegani.** Quality matters: A comprehensive comparative study of edge computing simulators, *Simulation Modelling Practice and Theory*, Vol. 138, 2025, 103042, ISSN 1569-190X, <https://doi.org/10.1016/j.simpat.2024.103042>.
14. **D. Hou, R. Evins.** A protocol for developing and evaluating neural network-based surrogate models and its application to building energy prediction, *Renewable and Sustainable Energy Reviews*, Vol. 193, 2024, 114283, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2024.114283>.

15. **Parviz Ghafariasl, Alireza Mahmoudan, Mahmoud Mohammadi, Aria Nazarpour, Siamak Hoseinzadeh, Mani Fathali, Shing Chang, Masoomeh Zeinalnezhad, Davide Astiaso Garcia.** Neural network-based surrogate modeling and optimization of a multigeneration system, *Applied Energy*, Vol. 364, 2024, 123130, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2024.123130>.
16. **N. Kanwisher, M. Khosla, K. Dobs.** Using artificial neural networks to ask ‘why’ questions of minds and brains, *Trends in Neurosciences*, Volume 46, Issue 3, 2023, P. 240-254, ISSN 0166-2236, <https://doi.org/10.1016/j.tins.2022.12.008>.
17. **Sobolevskii V.A.** (2024) Using AutoML technologies to solve monitoring problems. *Informatization and Communication*, 15(1), pp. 90-97. doi:10.34219/2078-8320-2024-15-1-90-97.
18. **Sobolevskii V.** (2025) Creation of automatic monitoring systems using AutoML technologies. In: *Digital Systems and Information Technologies in the Energy Sector. Lecture Notes in Networks and Systems*. Springer, pp. 401-410.
19. **A. Raj Munappy, J. Bosch, H. Holmström Olsson, A. Arpteg, B. Brinne,** Data management for production quality deep learning models: Challenges and solutions, *Journal of Systems and Software*, Vol. 191, 2022, 111359, ISSN 0164-1212, <https://doi.org/10.1016/j.jss.2022.111359>.
20. **Xin Liu, Jie Li, Jianwei Zhao, Bin Cao, Rongge Yan, Zhihan Lyu.** Evolutionary neural architecture search and its applications in healthcare, *CMES - Computer Modeling in Engineering and Sciences*, Volume 139, Issue 1, 2023, P. 143-185, ISSN 1526-1492, <https://doi.org/10.32604/cmes.2023.030391>.
21. Google Developers. Announcing TensorFlow Lite. Available at: <https://developers.googleblog.com/en/announcing-tensorflow-lite/> (23 September 2025).
22. **Redmon J., Divvala S., Girshick R. and Farhadi A.** (2016) You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 779-788.
23. Roboflow. Construction Item Computer Vision Dataset. Available at: <https://universe.roboflow.com/convr2022/construction-item> (23 September 2025).
24. Kaggle. Fruit and Vegetable Disease (Healthy vs Rotten). Available at: <https://www.kaggle.com/datasets/muhammad0subhan/fruit-and-vegetable-disease-healthy-vs-rotten> (accessed 23 September 2025).