

Исследование работы методов машинного зрения в условиях изменения освещенности для встраиваемых систем

О.В. Игнатьева, Д.Н. Чупий, А.Д. Сокирка, Д.С. Журавлев

Ростовский государственный университет путей сообщения

Аннотация: В статье рассмотрено влияние освещенности и расстояния на качество распознавания для различных моделей нейронных сетей встраиваемых систем. Описаны платформы на которых проводилось тестирование, а также используемые модели. Приведены результаты исследования влияния освещенности на качество распознавания.

Ключевые слова: искусственный интеллект, машинное зрение, встраиваемые системы, распознавание образов, YOLO, Inception, Peoplenet, ESP32, Sipeed, Jetson, Nvidia. Maix.

1. Введение

Интеграция искусственного интеллекта (ИИ) во встраиваемые устройства представляет собой актуальное и стремительно развивающееся направление в современных технологиях. Поскольку встраиваемые платформы на текущий момент уже широко распространены, их модернизация становится все более актуальной. Одним из наиболее эффективных способов совершенствования таких систем является внедрение в эти системы ИИ [1].

Поскольку существуют различные алгоритмы ИИ, предназначенные для выполнения различного рода задач, то область для изучения подобного направления достаточно обширна [2]. В данной статье рассматривается часть ИИ, которая называется машинным или компьютерным зрением [3, 4].

Интеграция системы ИИ на платформу с ограниченными ресурсами представляет собой сложную задачу. Процесс распознавания, в зависимости от требуемой скорости и точности, является вычислительно и энергоемким [5].

Современные алгоритмы позволяют экономить необходимые для работы ресурсы, однако, к сожалению, в степени не достаточной для их корректной и бесперебойной работы на встраиваемых системах.

Несмотря на сложность, интеграция систем ИИ на платформы с ограниченными ресурсами остается актуальной задачей. Для ее решения были разработаны специальные алгоритмы оптимизации и компрессии существующих систем машинного зрения, включая оптимизированные модели для распознавания объектов. [6]. К сожалению, крайне тяжело добиться меньших объемов и больших скоростей при тех же параметрах точности. Поэтому качество распознавания облегченных моделей, в среднем, ниже.

Это означает, что, применив подобные алгоритмы [7], появляется возможность размещения достаточно точных и быстрых систем машинного зрения на встраиваемых устройствах [8].

Данная статья посвящена апробированию таких систем на различных встраиваемых системах в различных условиях с практической целью внедрения наиболее подходящей системы в существующий проект.

2. Описание устройств для распознавания объектов

В ранее описанной работе было проведено исследование методов машинного зрения на встраиваемых системах с применением современных микроконтроллеров [9]. В целях сравнения возможностей и производительности встраиваемых платформ использовались следующие устройства: ESP32-Wrover-Dev v1.6 (ESP32-Wrover-E), Sipeed Maix Bit, а также Nvidia Jetson TX2 DevKit.

Для каждой платформы были разработаны специализированные модели, позволяющие распознавать объекты, релевантные задачам исследования, такие как люди или лица. В этих моделях распознанные объекты соответствующего типа обозначались идентификаторами “person” и “face”.

2.1 ESP32-Wrover-DEV v1.6 (ESP32-Wrover-E)

Встраиваемая система ESP32-Wrover-DEV v1.6 (ESP32-Wrover-E), далее ESP32, создавалась как универсальный модуль, поддерживающий технологии Wi-Fi, Bluetooth и Bluetooth LE MCU.

Параметры ESP32 представлены в таблице 1.

Таблица 1

Параметры ESP32

Часть модуля	Описание
1	2
CPU	Two low-power Xtensa® 32-bit LX6 microprocessors
GPU	-
Memory	448 KB of ROM for booting and core functions. 520 KB of on-chip SRAM for data and instructions. 8 KB of SRAM in RTC, which is called RTC FAST Memory and can be used for data storage 8 KB of SRAM in RTC, which is called RTC SLOW Memory 1Kbit of eFuse 8 Мб PSRAM
Storage	4 MB SPI flash memory
Screen	-
Camera	OV2640
TF card slot	-

Для обеспечения возможности работы машинного зрения на данной встраиваемой системе используются различные инструменты, такие как: OpenCV и TensorFlow. Следует отметить, что вычислительных мощностей ESP32 недостаточно для работы полноценного фреймворка TensorFlow. Существуют специальные конвертеры, которые позволяют создать облегченный фреймворк TensorFlow Lite. В текущем проекте использовалась библиотека OpenCV с предварительно обученной моделью YOLOv3 [10, 11].

Данный модуль не будет участвовать в финальных тестах, поскольку результаты, хоть сколько-нибудь удовлетворяющие требованиям, не были получены, а именно – не удалось запустить систему, работающую целиком и полностью на ESP32. Часть, размещенная на модуле, при использовании библиотеки OpenCV, делает кадр и размещает его на сервере, который располагается на том же модуле. Обработку и распознавание осуществляет компьютер, подключенный к серверу модуля.

2.2 Sipeed Maix Bit

Sipeed Maix Bit – это одно из семейств устройств, ориентированных на рынок «AI at the edge». Этот модуль обеспечивает высокую производительность при небольших физических и энергетических затратах, позволяя внедрять высокоточный искусственный интеллект категории «AI at the edge».

Параметры Sipeed Maix Bit представлены в таблице 2.

Таблица 2

Параметры Sipeed Maix Bit

Часть модуля	Описание
1	2
CPU	Dual-core 64bit RISC-V/400MHz (double precision FPU integration)
GPU	-
Memory	8MiB 64bit on-chip SRAM
Storage	16MiB Flash, support micro SDXC expansion storage (max 128GB)
Screen	2.4 inch TFT, screen resolution: 320*240
Camera	200W pixels (actual use 30W), OV2640 model M12 camera
TF card slot	Multimedia resource expansion, support large-capacity storage

Рассматриваемая платформа изначально разрабатывалась для нужд развертывания искусственного интеллекта на встраиваемых системах. Она обладает большим набором подключаемой периферии, чем ESP32. Также для работы с платами Maix компанией Sipeed разработана IDE MaixPy. Она позволяет писать код на языке программирования (ЯП) microPython. Еще одной функцией является то, что MaixPy может отображать видео, получаемое с камеры Sipeed Maix, а также отображать его параметры.

Данная IDE имеет встроенные библиотеки для работы с периферией модуля, что значительно упрощает процесс разработки.

Также на официальном сайте расположена документация к работе с MaixPy, описание ее библиотек и их функций. Документация представлена на китайском и английском языках. На официальном сайте присутствует встроенный переводчик google.

Для распознавания объектов используются предварительно обученные модели, которые должны иметь формат kmodel или smodel. В данной работе использовалась предварительно обученная модель YOLOv2 [12].

2.3 Nvidia Jetson TX2 DevKit

Nvidia Jetson TX2 DevKit – это платформа для разработки, созданная компанией Nvidia, предназначенная для разработчиков, инженеров и исследователей, которые хотят создавать инновационные решения с использованием искусственного интеллекта (ИИ), глубокого обучения и робототехники.

Наиболее перспективный из представленных модулей, поскольку обладает в значительной степени, десятки раз, увеличенной производительностью, памятью и скоростью вычислений.

Таблица 3

Параметры NVIDIA Jetson TX2 DevKit

Часть модуля	Описание
1	2
CPU	ARM8 64-битный мультипроцессор с архитектурой HMP CPU: 64-битный двухъядерный процессор NVIDIA Denver 2 на частоте до 2ГГц; Четырехъядерный Arm® Cortex®-A57 MPCore на частоте до 2ГГц;
GPU	Архитектура NVIDIA Pascal™ с 256 ядрами NVIDIA CUDA® на частоте до 1.12ГГц
Memory	4ГБ LPDDR4x 128-бит 1600МГц с пропускной способностью до 51.2 Гб/с
Storage	16 Гб eMMC 5.1, 128Гб NVMe SSD
Screen	2 многорежимные DP 1.2/eDP 1.4, HDMI 2.0 2 разъема DSI (1.5 Гбит/с на канал)
Camera	До 5 камер (12 по виртуальным каналам), 12 каналов MIPI CSI-2 (3x4 или 6x2), D-PHY 1.2 (до 30 Гбит/с)
TF card slot	-

3. Результаты проведения тестов

Было проведено тестирование устройства распознавания на выбранных встроенных системах. Собранные результаты были оформлены в виде таблиц и графиков. Это обеспечит наглядную разницу между различными этапами тестирования. В случае, когда вероятность распознавания ниже 50 процентов, примем, что объект не распознан, следовательно, вероятность

равна 0. Также представленные данные актуальны в случае, когда камера расположена на уровне 90 см и направлена ровно на исследуемый объект.

3.1 Nvidia Jetson TX2 DevKit

Данные, представленные в таблицах 4 – 5, а также данные, представленные на рисунках 1 – 6, были получены при тестировании модуля NVIDIA Jetson TX2 DevKit. Результаты, полученные при использовании камеры Prestigio PWC420 HD представлены в таблице 3.

Таблица 4

Результаты проведения тестов для камеры Prestigio PWC420 HD

Модель камеры	Модель нейронной сети	Расстояние в метрах	Вероятность распознавания человека в % от освещенности в люкс			
			0	180	270	420
1	2	3	4	5	6	7
Prestigio PWC420 HD	SSD Inception V2	2.4	< 50	90	88	89
		4.8	< 50	93	83	95
		7.2	< 50	65	64	68
		9.6	< 50	85.4	85	80
		12	< 50	< 50	< 50	< 50
		14.4	< 50	< 50	< 50	< 50
	Peoplenet-Pruned	2.4	53	99.7	99.8	99.8
		4.8	60	98.9	99	99.5
		7.2	93	98.1	97.9	94.7
		9.6	95.5	98	96	96.5
		12	95	97	97.5	98.7
		14.4	90	96.9	96.4	96

Продолжение таблицы 4

Модель камеры	Модель нейронной сети	Расстояние в метрах	Вероятность распознавания человека в % от освещенности в люкс			
			0	180	270	420
1	2	3	4	5	6	7
	Peoplenet	2.4	< 50	99.9	99.8	99.9
		4.8	< 50	99.8	99.8	99.8
		7.2	< 50	99.6	99.6	99.6
		9.6	< 50	97	98.3	97
		12	< 50	99.1	99.3	99.3
		14.4	< 50	99.4	99.5	99.5

График на рисунке 1 соответствует измерениями из таблицы 3 для модели SSD Inception V2.

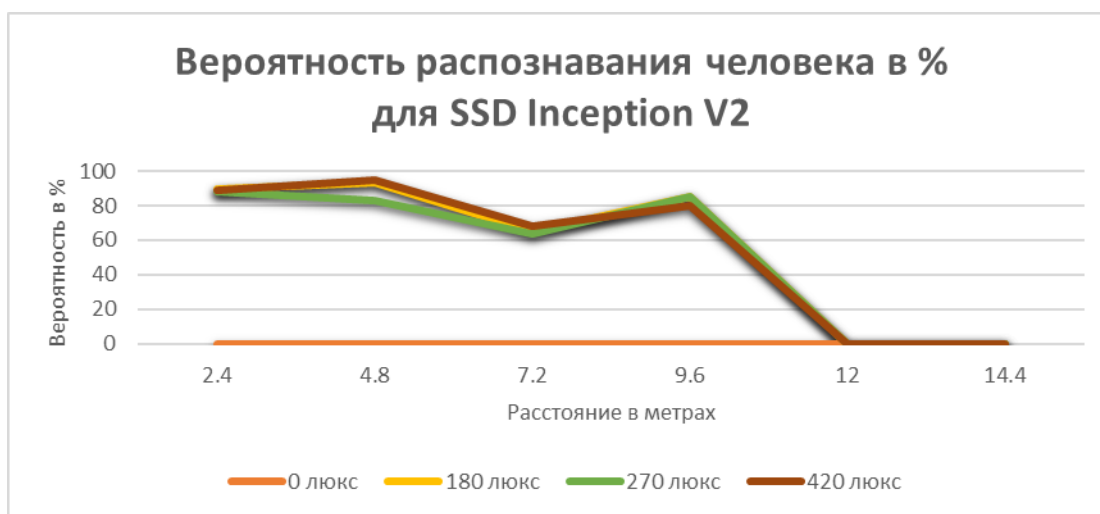


Рис. 1 - График вероятности распознавания человека при использовании модели SSD Inception V2

График на рисунке 2 соответствует измерениям из таблицы 3 для модели Peoplenet-Pruned.

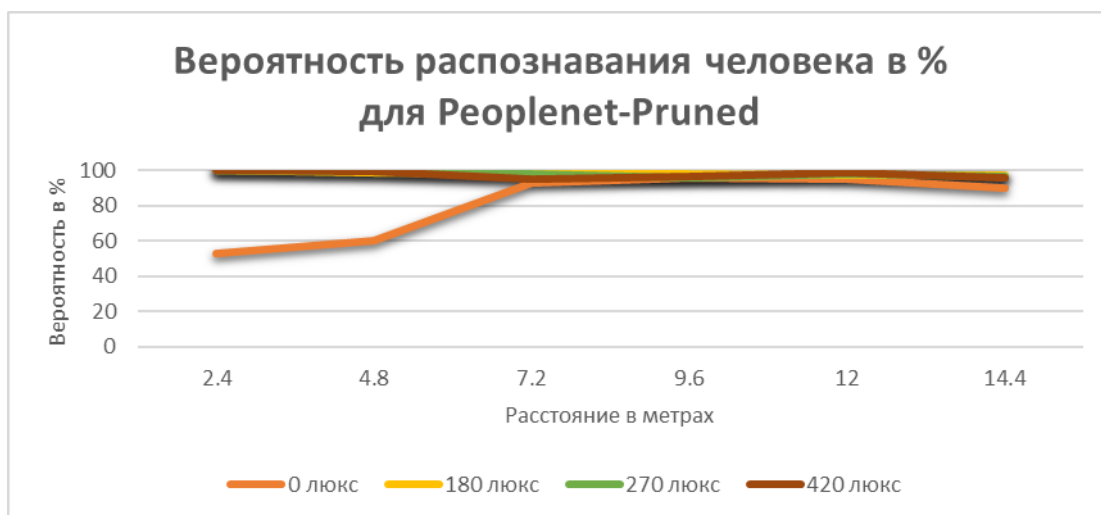


Рис. 2 - График вероятности распознавания человека при использовании модели Peoplenet-Pruned

График на рисунке 3 соответствует измерениям из таблицы 3 для модели Peoplenet.

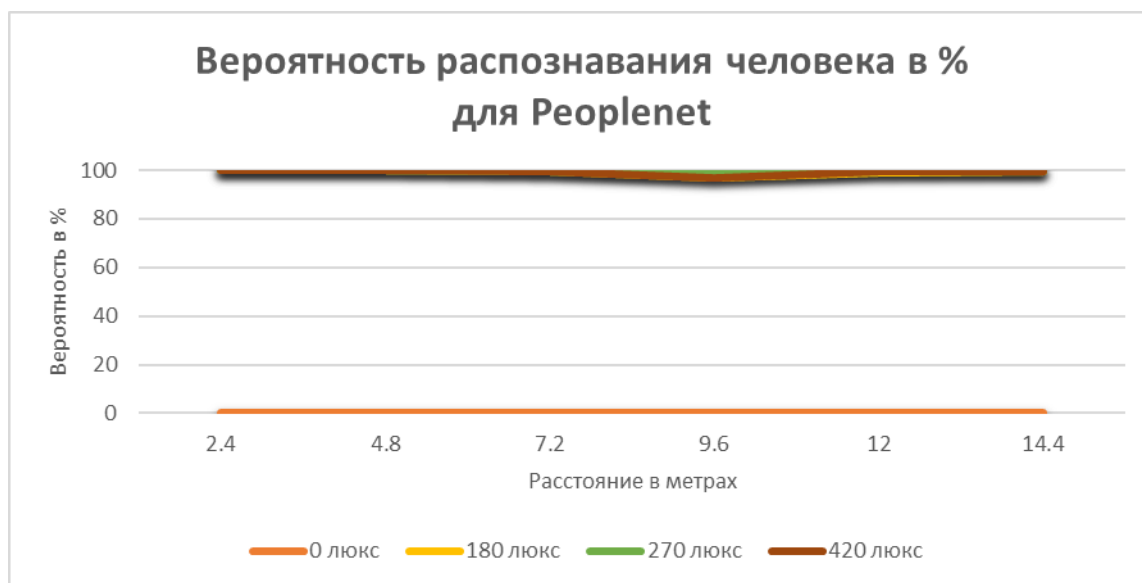


Рис. 3 - График вероятности распознавания человека при использовании модели Peoplenet

С камерой Prestigio PWC420 HD наилучший результат в среднем показала модель Peoplenet за исключением случая с освещенностью 0 люкс,

для которого лучший результат показала модель Peoplenet-Pruned. Результаты обеих моделей нейросетей показывают уровень эффективности, достаточный для достижения цели исследования – подсчет количества людей в кадре в реальном времени. При этом для распознавания моделью Peoplepruned требуется в среднем в 2 раза меньше вычислительных ресурсов. Однако в условиях недостаточной освещенности – 0 люкс, полноценная, неурезанная модель Peoplenet не смогла распознать силуэт человека, что смогла сделать ее облегченная версия Peoplenet-Pruned.

Результаты, полученные при использовании основной камеры Samsung Galaxy A54 представлены в таблице 4. Камера смартфона использует не только аппаратные возможности, но и программные, что позволяет улучшить качество изображения. Однако это означает не только более качественную картинку, но и увеличенные ресурсозатраты на ее обработку. Поэтому приложение, передающее видеопоток со смартфона на плату NVIDIA Jetson TX2 DevKit, снижает качество изображения до 1280x720.

Таблица 5

Результаты проведения тестов для основной камеры Samsung Galaxy A54

Камера	Модель нейронной сети	Расстояние в метрах	Вероятность распознавания человека в % от освещенности в люкс			
			0	180	270	420
1	2	3	4	5	6	7
Samsung Galaxy A54 Основная камера	SSD Inception V2	2.4	99.2	< 50	99	99.3
		4.8	63	76	88	72
		7.2	< 50	< 50	< 50	< 50
		9.6	70	< 50	< 50	75
		12	< 50	< 50	< 50	< 50
		14.4	< 50	< 50	< 50	< 50

Продолжение таблицы 5.

Камера	Модель нейронной сети	Расстояние в метрах	Вероятность распознавания человека в % от освещенности в люкс			
			0	180	270	420
1	2	3	4	5	6	7
Samsung Galaxy A54 Основная камера	Peoplenet-Pruned	2.4	99	99.4	99.9	99.9
		4.8	97.9	98.3	98	98.4
		7.2	98	98.7	95.4	96
		9.6	98.7	97.6	98.6	98.6
		12	97	97.5	98.8	97.3
		14.4	95.6	96	95.7	96.5
	Peoplenet	2.4	99.9	99.9	99.9	99.9
		4.8	98	98.9	98.5	98.9
		7.2	99.4	99	99.2	99.6
		9.6	97	99.5	98.6	98.6
		12	98	98	99.2	99.5
		14.4	98	98	96	99.3

График на рисунке 4 соответствует измерениям из таблицы 4 для модели SSD Inception V2.

График на рисунке 5 соответствует измерениям из таблицы 4 для модели Peoplenet-Pruned.

График на рисунке 6 соответствует измерениям из таблицы 4 для модели Peoplenet.

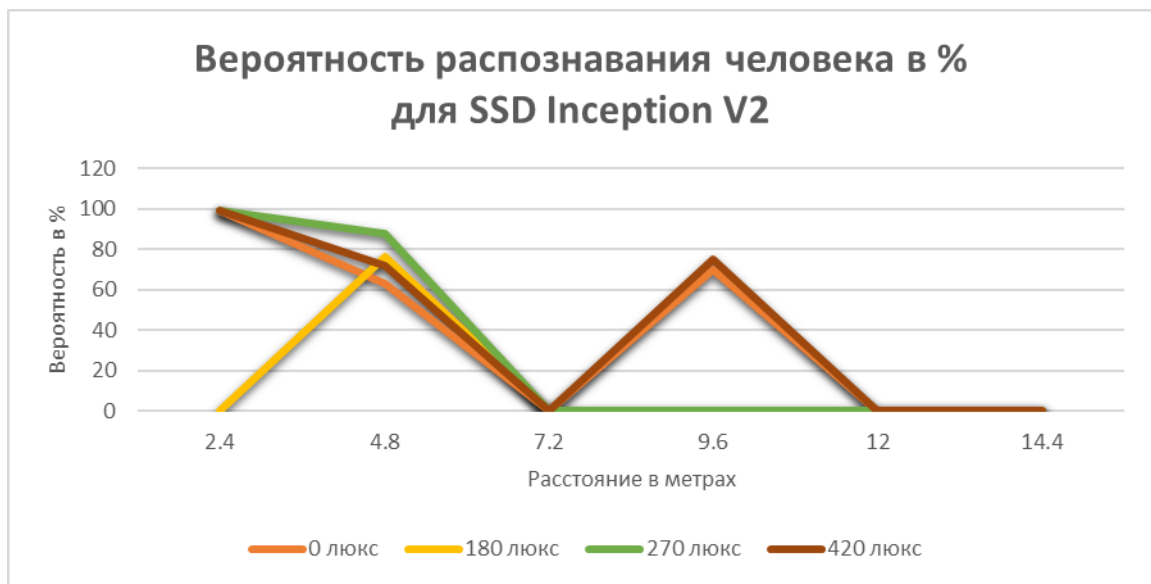


Рис. 4. - График вероятности распознавания человека при использовании модели SSD Inception V2

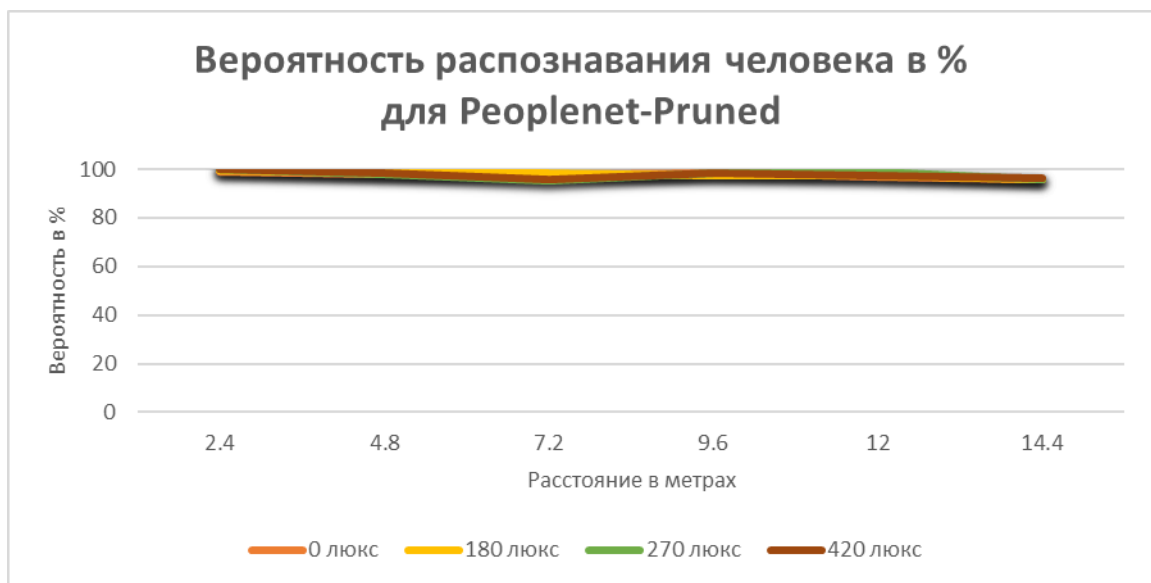


Рис. 5. - График вероятности распознавания человека при использовании модели Peoplenet-Pruned

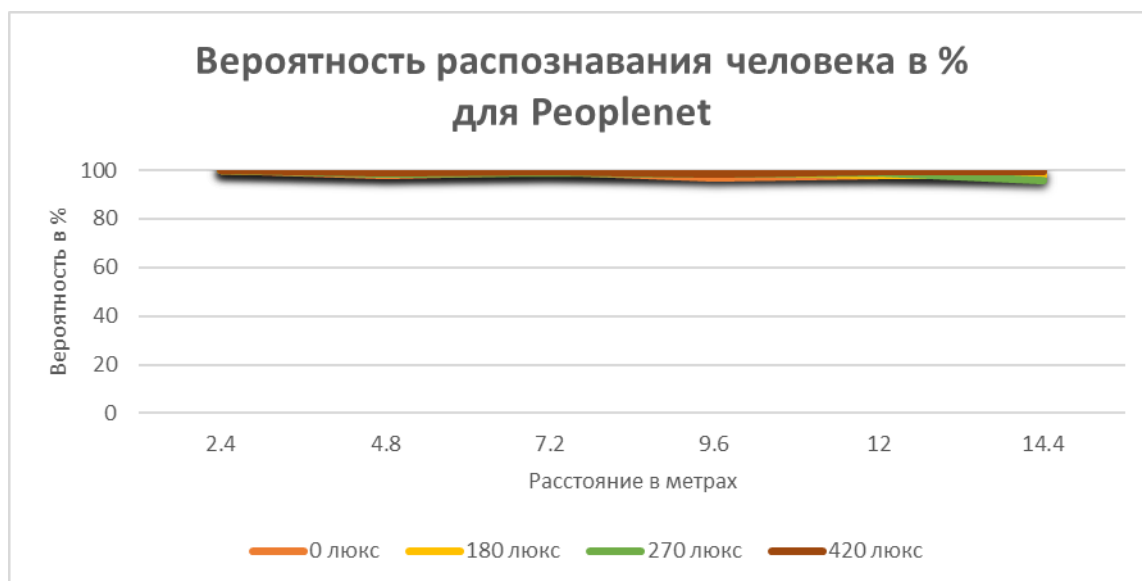


Рис. 6. - График вероятности распознавания человека при использовании модели Peoplenet

Для основной камеры смартфона Samsung A54 наилучшие результат показала модель Peoplenet. При этом модель Peoplenet-pruned в среднем показывает эффективность на 1% меньше, но которая остаётся на уровне, достаточном для выполнения поставленной задачи.

3.2 Sipeed Maix Bit

Данные, представленные в таблице 6, а также данные, представленные на рисунке 7, были получены при тестировании модуля Sipeed Maix Bit. Данный модуль основан на чипе K210, для которого существует не так много моделей, чтобы их использовать. Тесты проведены с использованием модели нейросети YOLO V2[13]. Также на официальном сайте Maix существует возможность обучить собственную модель под определенные нужды [14]. Однако модели, обученные при помощи предлагаемого инструмента, не обладают достаточной точностью, чтобы использовать их в общих тестах, поскольку точность составила менее 60% в лучшем случае [15].

Таблица 6

Результаты проведения тестов для Sipeed Maix Bit

Камера	Модель нейронной сети	Расстояние в метрах	Вероятность распознавания человека в % от освещенности в люкс			
			0	180	270	420
1	2	3	4	5	6	7
OV2640	YOLO V2	2.4	< 50	98	99	99
		4.8	< 50	96	96	97
		7.2	< 50	91	91	96
		9.6	< 50	73	62	64
		12	< 50	< 50	< 50	< 50
		14.4	< 50	< 50	< 50	< 50

График на рисунке 7 соответствует измерениям из таблицы 6 для модели YOLO V2.

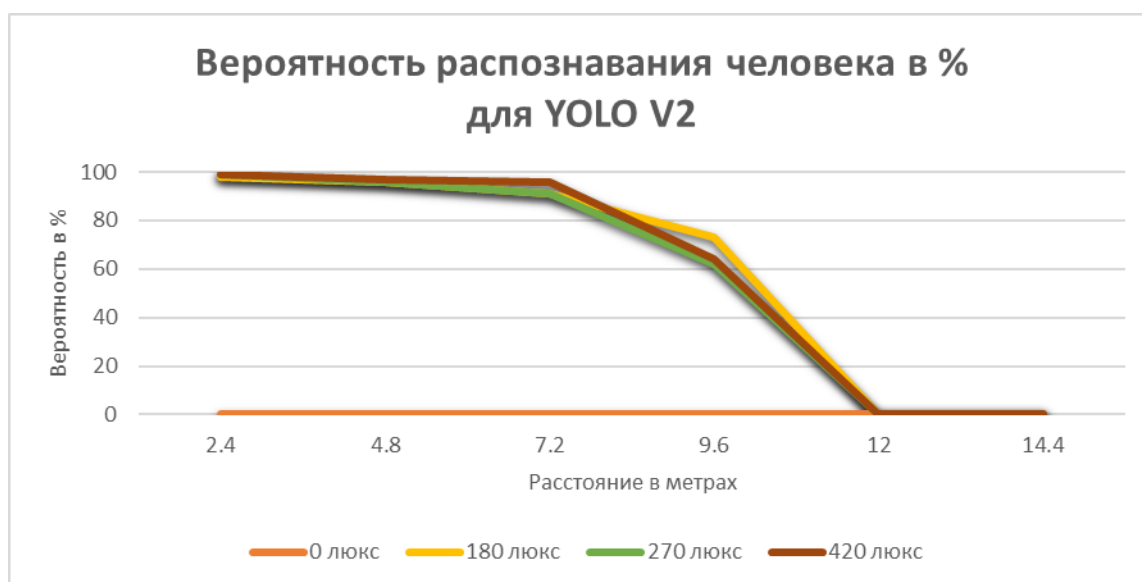


Рис. 6 - График вероятности распознавания человека при использовании модели YOLO V2

Для данной платформы не представляется возможность провести сравнительный анализ работы различных моделей, поскольку не удалось подобрать большее количество моделей для тестирования.

4. Заключение

При использовании NVIDIA Jetson TX2 DevKit, для обеих камер лучшие результаты показала модель PeopleNet. Основное различие заключается в том, что камера смартфона автоматически настраивает светочувствительность камеры. Это позволяет распознавать объекты в слабоосвещённых местах. Модель PeopleNet-pruned показала результаты близкие к неурезанной модели PeopleNet, но при этом, в среднем, потребляла в два раза меньше ресурсов. SSD-Inception-v3 показала худший результат. Независимо от удаленности, наблюдалась неустойчивость в выдаче результатов моделью, причем при достижении объектом расстояния в 12 метров от источника изображения происходила полная потеря способности к распознаванию.

Как и при использовании моделей SSD Inception V2 и PeopleNet с камерой Prestigio PWC420 HD, при недостаточном уровне освещенности модуль не способен распознать объект на всей выборке расстояний. Однако, в отличие от модели NVIDIA Jetson TX2 DevKit с моделью SSD Inception V2 и камерой Prestigio PWC420 HD, Sipeed Maix Bit с Yolo V2 показал большую линейность при изменении степени освещенности и расстояния.

Литература

1. Ветохин В.В., Ракитин Я.Е., Нестеренко И.А. Выбор средств реализации при разработке автоматизированной информационной системы с применением технологий машинного зрения // Столыпинский вестник. 2022. № 4. С. 1958 – 1966.

2. Рапаков Г.Г., Малыгин Л.Л., Пчелкина О.С. Технологии компьютера зрения и методы машинного обучения в задаче разработки системы промышленной видеоаналитики // Вестник Череповецкого государственного университета. 2022. № 6 (111). С. 79 – 88.

3. Горячкин Б.С., Китов М.А. Компьютерное зрение. E-Scio. 2020. №9 (48). С. 317 – 345.

4. Kokaia G. About computer vision with machine learning. Computer Sciences and Telecommunications. 2022. №2 (62). pp. 10 – 13.

5. Scherer R. Computer Vision Methods for Fast Image Classification and Retrieval. Czestochowa: Springer, 2019. – 137 p.

6. Rousseau J.-J. Kapralos B. Pattern Recognition. Computer Vision, and Image Processing. Montreal: Springer Nature, 2023. – 732 p.

7. Панарин Р.Н., Соловьев А.А., Хворова Л.А. Применение технологий искусственного интеллекта и компьютерного зрения при решении задач автоматизации обработки и распознавания биологических объектов. Известия Алтайского государственного университета. 2022. №1 (123). С. 101 – 107.

8. Ковалев В.В., Сергеев Н.Е. Реализация сверточных нейронных сетей на встраиваемых устройствах с ограниченным вычислительным ресурсом. Известия ЮФУ. Технические науки, 2021. №6. С. 64 – 72.

9. Игнатьева О.В., Сокирка А.Д., Журавлев Д.С. Применение методов машинного зрения на встраиваемых системах. Инженерный вестник Дона, 2024, №1. URL: ivdon.ru/ru/magazine/archive/n1y2024/8918.

10. Bohush R., Guangdi Ma., Yang W., Ablameyko S., Hongxu Q. People detecting and tracking in video by CNN YOLO and Strongsort combined algorithm. Nonlinear Phenomena in Complex Systems. 2023. Vol.26. No.1. Pp. 83 – 97.

11. Сичкар В.Н., Колюбин С.А. Детектирование и классификация дорожных знаков в реальном времени на основе алгоритма YOLO версии 3. Научно-технический вестник информационных технологий, механики и оптики. 2020. № 3. С. 418 – 424.

12. Aria K., Kapoor S. *Advances in Computer Vision Volume 2*. Saga: Springer Nature, 2020. – 767 p.

13. Пчелинцев С.Ю., Ковалева О.А., Суслин А.А. Тестирование и анализ фреймворков, предназначенных для глубокого обучения нейросетевых моделей. Вестник кибернетики. 2022. №2(46). С. 6 – 13.

14. Жиленков А.А., Силкин А.А., Серебряков М.Ю., Колесова С.В. Сравнительный анализ систем глубокого обучения с подкреплением и систем обучения с учителем. Известия Тульского государственного университета. Технические науки. 2022. № 10. С. 109 – 112.

15. Качков М.С. Создание нейронной сети для решения различных прикладных задач. Известия Тульского государственного университета. Технические науки. 2023. № 2. С. 339 – 343.

References

1. Vetohin V.V., Rakitin YA.E., Nesterenko I.A. *Stolypinskij vestnik*. 2022. V 4. Pp. 1958 – 1966.

2. Rapakov G.G., Malygin L.L., Pchelkina O.S. *Vestnik Cherepoveckogo gosudarstvennogo universiteta*. 2022. V.6 (111). Pp. 79 – 88.

3. Goryachkin B.S., Kitov M.A. *E-Scio*. 2020. V.9 (48). pp. 317 – 345.

4. Kokaia G. About computer vision with machine learning. *Computer Sciences and Telecommunications*. 2022. №2 (62). Pp. 10 – 13.

5. Scherer R. *Computer Vision Methods for Fast Image Classification and Retrieval*. Czestochowa: Springer, 2019. 137 p.

6. Rousseau J.-J., Kapralos B. *Pattern Recognition, Computer Vision, and Image Processing*. Montreal: Springer Nature, 2023. 732 p.



7. Panarin R.N., Solov'ev A.A., Hovorova L.A. Izvestiya Altajskogo gosudarstvennogo universiteta. 2022. V.1 (123). Pp. 101 – 107.
8. Kovalev V.V., Sergeev N.E. Izvestiya YUFU. Tekhnicheskie nauki, 2021. V.6. Pp. 64 – 72.
9. Ignat'eva O.V., Sokirka A.D., ZHuravlev D.S. Inzhenernyj vestnik Dona, 2024. V.1. URL: ivdon.ru/ru/magazine/archive/n1y2024/8918.
10. Bohush R., Guangdi Ma., Yang W., Ablameyko S., Hongxu Q. People detecting and tracking in video by CNN YOLO and Strongsort combined algorithm. Nonlinear Phenomena in Complex Systems. 2023. Vol.26. No.1. Pp. 83 – 97.
11. Sichkar V.N., Kolyubin S.A. Nauchno-tehnicheskij vestnik informacionnyh tekhnologij, mekhaniki i optiki. 2020. V.3. Pp. 418 – 424.
12. Aria K., Kapoor S. Advances in Computer Vision Volume 2. Saga: Springer Nature, 2020. 767 p.
13. Pchelincev S.YU, Kovaleva O.A., Suslin A.A. Vestnik kibernetiki. 2022. V.2 (46). Pp. 6 – 13.
14. ZHilenkov A.A., Silkin A.A., Serebryakov M.YU., Kolesova S.V. Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki. 2022. V.10. Pp. 109 – 112.
15. Kachkov M.S. Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki. 2023. V.2. Pp. 339 – 343.

Дата поступления: 20.03.2024

Дата публикации: 26.04.2024