

Применение методов машинного зрения на встраиваемых системах

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

Ростовский государственный университет путей сообщения

Аннотация: В статье рассмотрено применение методов машинного зрения для встраиваемых систем с применением современных микроконтроллеров. Описаны методы машинного обучения, которые используются в встроенных системах для решения задач распознавания, а также модели нейронных сетей. Предложено использование обученных моделей для решения задач распознавания изображений в встроенных системах. Проведено сравнение архитектур нейронных сетей YOLOv3 и R-CNN. Рассмотрена аппаратная платформа Jetson TX2. Представлены результаты сравнения скорости вычисления для различных режимов устройства.

Ключевые слова: машинное зрение, нейронные сети, искусственный интеллект, встраиваемые системы, распознавание образов, YOLO, R-CNN, Jetson, Tensorflow.

Введение

Целью исследования является адаптация методов машинного зрения для распознавания образов на встраиваемых системах с применением современных микроконтроллеров.

Встраиваемые системы - это вычислительные системы специального назначения, встроенные в прикладные среды или в другие вычислительные системы и обеспечивающие специализированную поддержку. Встроенная система представляет собой комбинацию компьютерного процессора, компьютерной памяти и периферийных устройств ввода-вывода, которая имеет специальную функцию в более крупной механической или электронной системе. Она встроена как часть полного устройства, часто включающего электрическое или электронное оборудование и механические части. Примерами могут служить сетевые адаптеры для вычислительных систем и мобильных телефонов, системы управления кондиционированием воздуха, промышленные системы и встроенные системы для автомобилей, системы видеонаблюдения и другие [1].

Актуальной задачей является применение методов машинного зрения на встраиваемых системах с применением современных микроконтроллеров.

Но машинное зрение пока не входит в набор обязательных функций процессора. Машинное зрение позволяет выделять информацию из визуальных сигналов [2]. Оно может быть полезным во множестве приложений: в автомобиле - для предотвращения аварий, в системах безопасности, на сборочных конвейерах, в медицинском оборудовании, в оптических системах распознавания, и многих других, рис. 1 [3].



Рис. 1. – Примеры применения машинного зрения [3]

До сих пор эта технология была слишком дорогой для повсеместного использования. Достижения в области полупроводниковой электроники делают ее доступнее для массового потребителя. Алгоритмы машинного зрения являются очень ресурсоемкими.

Для создания программного приложения, определяющего объекты на изображениях, разработчику потребуется большое количество времени, а также вычислительных аппаратных ресурсов, поскольку следует учитывать очень большое количество всевозможных критериев отбора и классификации объектов. Но современные методы машинного обучения позволяют переложить решение данной задачи на вычислительное устройство. Тогда программисту не придется самостоятельно указывать алгоритму признаки

определения объектов. В ЭВМ необходимо только заложить в машину алгоритм принятия решения на основе некоторых входных данных. Основной проблемой такого подхода является необходимость использовать высокопроизводительные устройства и большие объёмы данных для обучения, что не позволяет использовать данные методы для решения простых задач из-за высокой стоимости реализации.

Анализ проблемы

Согласно ГОСТ Р 59276-2020 «Национальный стандарт Российской Федерации. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения» Искусственный интеллект (ИИ) – это способность технической системы имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. А также, согласно ГОСТ Р 59277-2020 «Системы искусственного интеллекта. Классификация систем искусственного интеллекта», ИИ – это комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение, поиск решений без заранее заданного алгоритма и достижение инсайта) и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека.

Таким образом, ИИ рассматривается с точки зрения поведения системы или состава системы.

Для решения задачи адаптации методов машинного зрения для распознавания образов с применением современных микроконтроллеров, требующей задействования искусственного интеллекта на базе нейронной сети, в основном используют способы [4]:

- Обучение с учителем. Для каждого события имеется пара ситуация-решение, с помощью которой корректируются значения весов;
- Обучение без учителя. Для каждого события имеются данные только о других событиях, на основе которых необходимо отнести текущее событие к определенному классу.

Основная задача машинного зрения научить устройство видеть, различать и классифицировать объекты [5, 6]. Данные возможности позволяют автоматизировать множество процессов контроля - качества или поведения, наблюдения, сообщения о нестандартных или критичных ситуациях.

Встраиваемые системы

Для решения поставленной задачи адаптации методов машинного зрения для распознавания образов с применением современных микроконтроллеров используется встраиваемая система – устройство на основе микроконтроллера или микропроцессора для выполнения определенной задачи или их ограниченного числа, рис. 2.



Рис. 2. – Пример встраиваемой системы

Встраиваемые системы не обладают той же гибкостью, что и обычные компьютеры, но при этом имеют куда более компактные размеры, что

позволяет их встраивать в уже существующие системы. Основным достоинством встраиваемых систем является низкое энергопотребление, стоимость и размеры, но из-за этого страдают их производительность и многофункциональность. Но, ввиду ограниченности их вычислительной мощности, реализовать машинное обучение на них в классическом смысле очень весьма затруднительно.

Использование обученных моделей

Чтобы научить устройство решать определённые задачи из области искусственного интеллекта необходимо дать ей большое количество уже обработанных данных, на основе которых она и будет учиться. Но для встраиваемых систем, которые имеют меньшую производительность относительно стационарных систем, даже сам процесс обучения может занимать большое количество времени, поэтому используется иной метод [7]. Берут уже обученную модель и внедряют её в систему, пропуская таким образом необходимость обучения, что в свою очередь позволяет сократить время реализации. Модель может быть загружена из сети, имея возможность решать ряд определённых задач или же разработчик может сам обучить свою модель. Однако неизменным остаётся тот факт, что модель обучается вне встроеной системы, а на неё передаётся только выходная модель.

Таким образом, применение методов машинного зрения на встраиваемых системах сводится к:

1. Выбору или созданию подходящей модели.
2. Переносу модели в код устройства.
3. Разработки алгоритма работы с моделью.

Для встраиваемых систем используются облегченные версии библиотек для работы с машинным обучением. Так, например, для ESP32 существует библиотека TensorFlow Lite, которая позволяет использовать модели, обученные в полноценной Tensorflow на встраиваемых системах

путём конвертации модели в сжатый плоский буфер, после чего сжатый файл можно развернуть на устройстве [8]. Также для повышения производительности можно выполнить оптимизацию путём замены 32-битных чисел с плавающей запятой, на целые 8-битные.

Часть встраиваемых систем в базовой комплектации имеет интегрированный графический процессор, который позволяет эффективнее работать с задачами машинного обучения путем задействования вычислительного блока GPU. Одними из представителей можно считать линейку NVidia Jetson, в частности Jetson Nano и Jetson TX2. Эти устройства разрабатывались специально для решения задач машинного обучения и, несмотря на то что, они также являются встраиваемыми системами, им хватает производительности для обучения модели внутри своей системы. Данные устройства поддерживают большинство полноценных библиотек машинного обучения. Jetson TX2 NX совместим с Jetson Nano, но при этом даёт в 2, 5 раз большую производительность при потреблении всего 7,5 Вт [9].

Архитектуры нейронных сетей

Для распознавания изображений существует большое количество архитектур нейронных сетей, но самыми интересными, ввиду их производительности, можно назвать сети YOLO и R-CNN. Главным их отличием является принцип нахождения объекта на изображении [10].

Так R-CNN использует регионы для обнаружения объектов на изображении. При помощи Selective Search изображение разбивается на регионы, в которых потенциально может содержаться объект [11]. Затем полученный регион преобразовывается к размеру 227 на 227 пикселей, который принимается сверточной нейронной сетью CaffeNet. Она формирует 4096-размерный вектор признаков, для которых производятся бинарные классификации с помощью метода опорных векторов. На основе полученных

параметров рамки региона с помощью линейной регрессии также подстраивается и размер охватываемого региона, рис. 3 [12].

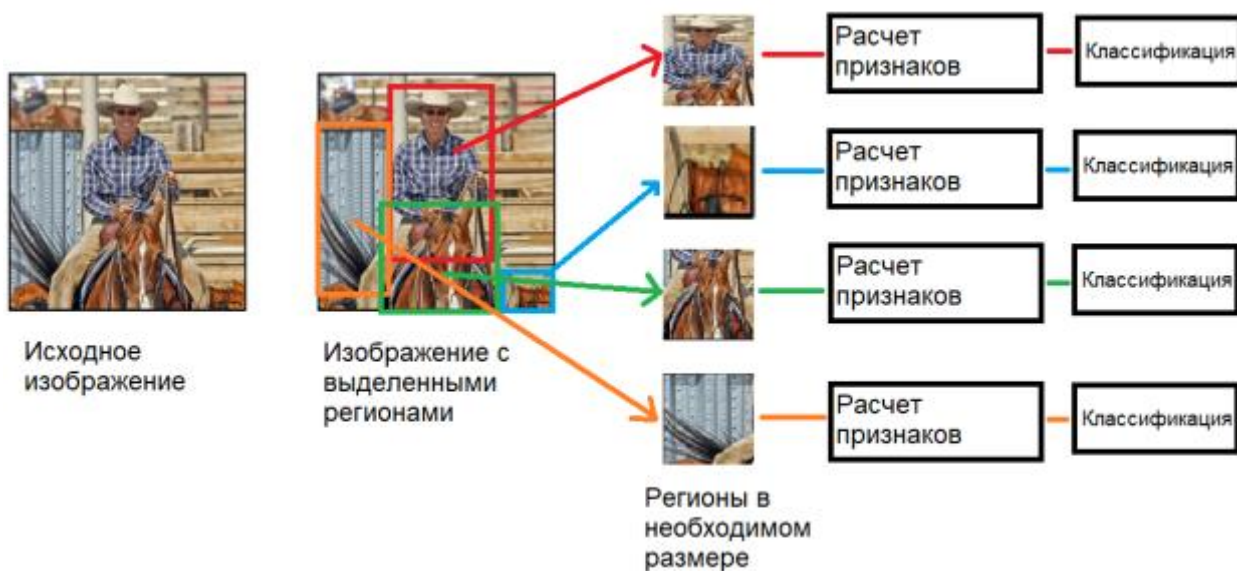
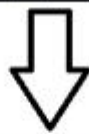


Рис. 3. – Работа R-CNN сети [12]

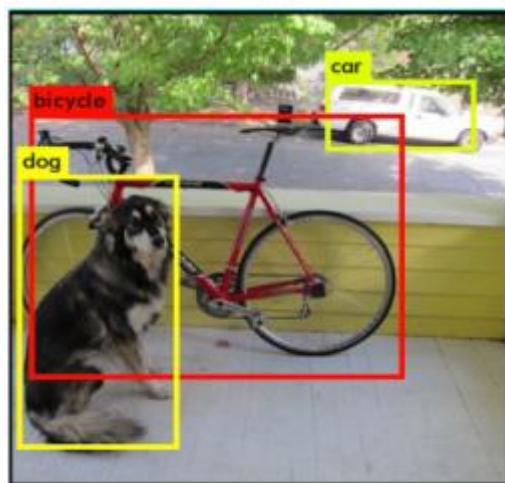
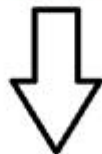
YOLO использует иной подход. Как и в R-CNN, изображение проходит через сверточную нейронную сеть, но в данном случае это сеть DarkNet-53, которая принимает на вход изображение размером 416 x 416 пикселей и на основе этого изображение выделяются признаки. Изображение делится на ячейки размером $S \times S$ [13]. Каждая ячейка отвечает за обнаружение и определение положения объекта. Также ячейки прогнозируют размеры рамки, которая ограничивает объект, относительно координат ячейки, вероятность присутствия объекта и его имя. По итогу имеется большой набор повторяющихся прогнозов, который преобразуется методом немаксимального подавления, рис. 4 [14].



Исходное изображение,
разделенное на ячейки



Все предсказанные
границы для объектов



Итоговые границы,
полученные путём
немаксимального
подавления

Рис. 4. – Получение границ объектов сетями YOLO [14]

Ввиду особенностей архитектуры, сети семейства YOLO имеют большую производительность, позволяя обработать большее количество кадров секунду. Эти особенности позволяют использовать данные сети на встраиваемых системах для распознавания объектов в реальном режиме времени.

Использование модели YOLOv3 на встраиваемой системе Jetson TX2

Jetson TX2 – система разработанная компанией NVIDIA специально для работы с искусственным интеллектом и глубоким обучением. Для решения этих задач он оснащён аппаратным ускорителем NVIDIA Pascal с 256 ядрами CUDA. Также устройство оснащено такими интерфейсами, как: Gigabit Ethernet, HDMI, USB 3.0, USB 2.0, CSI, DSI, UART, I2C и поддерживает беспроводные технологии WiFi и Bluetooth, рис. 5.

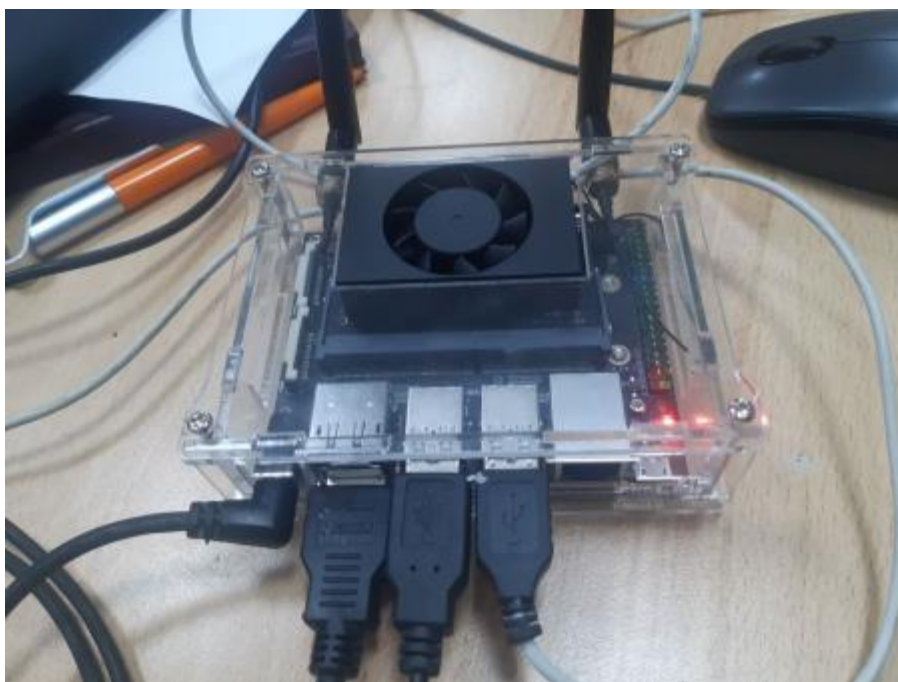


Рис. 5. – Внешний вид JetsonTX2

Несмотря на то, что Jetson TX2 поддерживает работу с библиотекой Tensorflow, которая создана для настольного ПК, его производительность

имеет очень низкий уровень, потому что библиотека не задействует CUDA ядра. Для решения этой проблемы компания создала свою библиотеку TensorflowRT. Но даже эта библиотека имеет ряд ограничений, одним из которых являются используемые модели. TensorflowRT не позволяет напрямую использовать модель YOLOv3, поэтому её придётся конвертировать в формат ONNX. Этот формат является открытым представлением нейронной сети, который позволяет связать модели из разных фреймворков. Полученный ONNX-файл в дальнейшем также проходит через ещё одну конвертацию уже средствами TensorflowRT и в результате получаем модель, которая способна использовать все вычислительные ресурсы встраиваемой системы. Помимо смены библиотеки Jetson TX2, позволяет улучшить результаты с помощью изменения режима питания (таблица 1), от которого зависит количество используемых ядер, а также их частота.

Таблица №1

Режимы питания Jetson TX2

Режим работы	Ядра Denver	Ядра ARM	Частота
Max-N	2	4	2.0 GHz
Max-Q	0	4	1.2 GHz
Max-P-Core-All	2	4	1.4 GHz
Max-P-ARM	0	4	2.0 GHz

Тестирование скорости работы YOLOv3 для Jetson TX2.

На устройстве Jetson TX2 было протестировано распознавание объектов на изображении для различных режимов питания, а также с использованием библиотек Tensorflow RT и Tensorflow. Результаты тестирования приведены в таблице 2.

Таблица №2

Сравнение производительности Jetson TX2

Режим работы	Tensorflow, сек	TensorflowRT, сек
Max-N	3,56	0,042
Max-Q	6,5	0,083
Max-P-Core-All	4,86	0,063
Max-P-ARM	3,68	0,05

Тестирование проводилось на фотографии размером 608 на 608 в формате jpeg, рис. 6.



Рис. 6 - Фотография для тестирования производительности

На выходе получаем изображение, на котором нейронная сеть предсказала границы объекта и его класс, рис. 7.

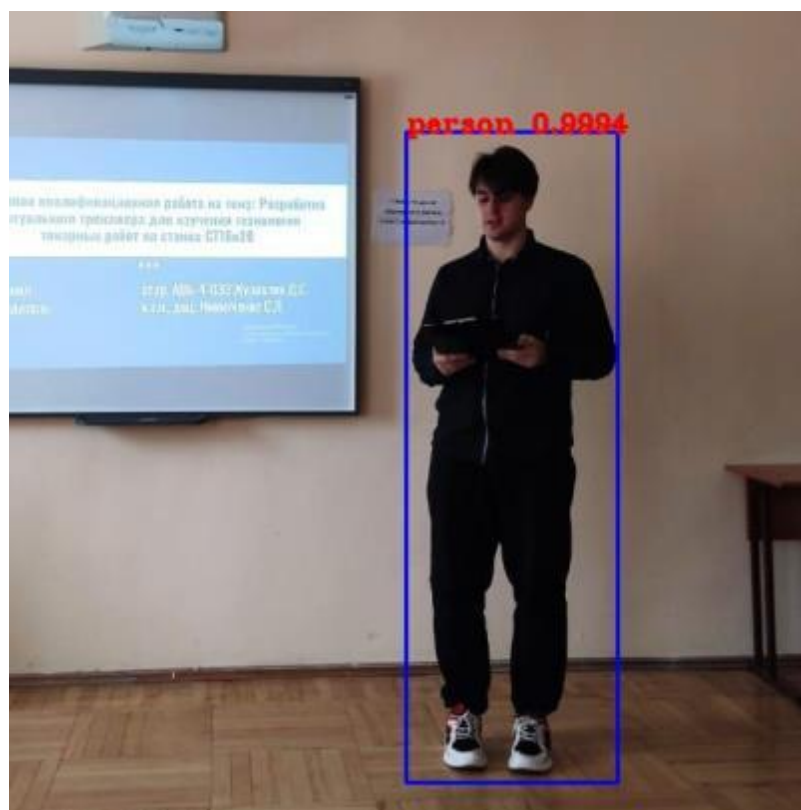


Рис. 7. – Фотография после прохождения через нейронную сеть

По результатам тестирования можно заключить, что Jetson TX2 способен распознавать объекты на изображении вне зависимости от его режима питания и даже от используемой библиотеки, но если выбирать оптимальные параметры, а именно - режим питания MAX-N и библиотеку TensorflowRT, то можно значительно увеличить производительность работы устройства и получить скорость, с которой можно будет производить обнаружение не на фотографии, а на потоковом видео.

Выводы

Использование машинного зрения требует больших вычислительных ресурсов от машины, но если подобрать правильную архитектуру сети, которая будет заниматься обработкой видеопотока, то можно добиться приемлемого уровня работы даже на встраиваемых системах, которые, в

свою очередь, открывают возможность к созданию небольших энергоэффективных и недорогих устройств.

Литература

1. Malamas E.N., Petrakis E.G.M., Zervakis M., Petit L., Legat J.D. A survey on industrial vision systems, applications and tools. *Image and Vision Computing*. 2003. V. 21. № 2. Pp. 171-188.
2. Kokaia G. About computer vision with machine learning. *Computer Sciences and Telecommunications*. 2022. № 2 (62). Pp. 10-13.
3. Applications of Embedded System Based Real-Time Projects. <https://www.elprocus.com/real-time-applications-of-embedded-systems/>
4. 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.
5. Качков М.С. Создание нейронной сети для решения различных прикладных задач // *Известия Тульского государственного университета. Технические науки*. 2023. № 2. С. 339-343.
6. Татарникова Т.М., Пойманова Е.Д., Богданов П.Ю., Краева Е.В., Вережкин С.А. Программные среды для изучения основ нейронных сетей // *Программные продукты и системы*. 2021. № 1. С. 145-150.
7. Лебедев И.С. Адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации // *Информационно-управляющие системы*. 2022. № 3 (118). С. 20-30.
8. 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.

9. Ковалев В.В., Сергеев Н.Е. Реализация сверточных нейронных сетей на встраиваемых устройствах с ограниченным вычислительным ресурсом // Известия ЮФУ. Технические науки. 2021. № 6 (223). С. 64-72.

10. Тимошкин М.С., Миронов А.Н., Леонтьев А.С. Сравнение YOLO V5 и Faster R-CNN для обнаружения людей на изображении в потоковом режиме // Международный научно-исследовательский журнал. 2022. № 6-1 (120). С. 137-146.

11. Линдигрин А.Н. Искусственные нейронные сети как основа глубокого обучения // Известия Тульского государственного университета. Технические науки. 2019. № 12. С. 468-472.

12. Hyungtae Lee, Sungmin Eum, Heesung Kwon. ME R-CNN: Multi-Expert R-CNN for Object Detection. Computer Science «Computer Vision and Pattern Recognition». IEEE Transactions on Image Processing. 2022. arXiv:1704.01069v3.

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

14. Arka Prava Jana, Abhiraj Biswas, Mohana Mohana. YOLO based Detection and Classification of Objects in video records. 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). 2018. Pp. 2448-2452.

References

1. Malamas E.N., Petrakis E.G.M., Zervakis M., Petit L., Legat J.D. Image and Vision Computing. 2003. V. 21. № 2. Pp. 171-188.

2. Kokaia G. Computer Sciences and Telecommunications. 2022. № 2 (62). Pp. 10-13.

3. Applications of Embedded System Based Real-Time Projects.
<https://www.elprocus.com/real-time-applications-of-embedded-systems/>
 4. ZHilenkov A.A., Silkin A.A., Serebryakov M.YU., Kolesova S.V. Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki. 2022. № 10. Pp. 109-112.
 5. Kachkov M.S. Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki. 2023. № 2. Pp. 339-343.
 6. Tatarnikova T.M., Pojmanova E.D., Bogdanov P.YU., Kraeva E.V., Programmnye produkty i sistemy. 2021. № 1. Pp. 145-150.
 7. Lebedev I.S. Informacionno-upravlyayushchie sistemy. 2022. № 3 (118). Pp. 20-30.
 8. Pchelincev S.YU., Kovaleva O.A., Suslin A.A. Vestnik kibernetiki. 2022. № 2 (46). Pp. 6-13.
 9. Kovalev V.V., Sergeev N.E. Izvestiya YUFU. Tekhnicheskie nauki. 2021. № 6 (223). Pp. 64-72.
 10. Timoshkin M.S., Mironov A.N., Leont'ev A.S. Mezhdunarodnyj nauchno-issledovatel'skij zhurnal. 2022. № 6-1 (120). Pp. 137-146.
 11. Lindigrin A.N. Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki. 2019. № 12. S. 468-472.
 12. Hyungtae Lee, Sungmin Eum, Heesung Kwon. ME R-CNN: Multi-Expert R-CNN for Object Detection. Computer Science «Computer Vision and Pattern Recognition». IEEE Transactions on Image Processing. 2022. ArXiv: 1704.01069v3.
 13. Hongxu Q., Bohush R., Guangdi Ma., Yang W., Ablameyko S. People detecting and tracking in video by CNN YOLO and Strongsort combined algorithm. Nonlinear Phenomena in Complex Systems. 2023. V. 26. № 1. Pp. 83-97.
-



14. Arka Prava Jana, Abhiraj Biswas, Mohana Mohana. YOLO based Detection and Classification of Objects in video records. 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). 2018. Pp. 2448-2452.

Дата поступления: 24.11.2023

Дата публикации: 3.01.2024