

Использование нейронных сетей для решения задач компьютерного зрения

Б.К. Лебедев, О.Б. Лебедев, Р.И. Черкасов

МИРЭА – Российский технологический университет, Москва

Аннотация: В статье рассмотрены основные подходы к решению задач компьютерного зрения с использованием нейронных сетей. Описаны типы задач, решаемых при помощи компьютерного зрения. Представлены механизмы функционирования сверточной нейронной сети. Рассмотрена задача выбора модели обнаружения объектов, обобщающую более изученную задачу классификации объектов и оценка эффективности алгоритмов обнаружения объектов. В статье также освещены современные подходы к обучению нейронных сетей, включая использование предобученных моделей и популярных методов. Описаны преимущества и ограничения различных архитектур сверточных нейронных сетей. Рассмотрены методы работы с данными для повышения качества функционирования моделей. Приведены примеры использования алгоритмов компьютерного зрения в распознавании образов и диагностике. Оценены метрики точности. Дана оценка эффективности алгоритмов обнаружения объектов, на основе различных метрик.

Ключевые слова: компьютерное зрение, архитектура, сверточные нейронные сети, цифровое изображение, классификация объектов.

Введение

В настоящее время одним из самых сложных направлений искусственного интеллекта является решение задач, связанных с использованием компьютерного зрения [1].

Компьютерное зрение направлено на воссоздание компонентов сложной системы зрения человека, что дает возможность компьютерам распознавать и анализировать объекты на изображениях и видео также, как это делают люди [2].

Объем применения компьютерного зрения в мире постоянно увеличивается. Компьютерное зрение относится к инновационным технологиям, позволяющим автоматизировать множество процессов и изменяющим нашу повседневную жизнь. С увеличением вычислительной мощности улучшается точность распознавания, что способствует разработке более сложных систем на основе этой технологии [3,4].

1. Общие сведения о задаче классификации объектов

Задачи, которые можно решить с помощью компьютерного зрения:

- 1) классификация изображения по конкретному входному пикселю;
- 2) семантическая сегментация – выделение определенных частей изображения или определение класса изображения по отдельным пикселям;
- 3) объект распознавание – идентификация и расположение объекта на изображении;
- 4) сегментация объекта – идентификация объекта и сегментирование его пикселей по классу пикселей [1,5].

Для решения задачи модель должна уметь определять рассматриваемый объект, следовательно, подходящими моделями являются модели классификации или обнаружения объектов. Рассмотрим постановку задачи классификации. На множестве некоторых объектов X и определенном множестве меток Y , задана целевая функция $y^* : X \rightarrow Y$. Указанная функция определена только на конечном множестве объектов $X^N = \{x_1, x_2, \dots, x_N\}$. Ставится задача построения функции (алгоритма) $a : X \rightarrow Y$, прогнозирующую результаты для любых объектов из X .

Признаком объекта называется пространство $f : X \rightarrow D_f$, где D_f – множество всех меток признака. Если заданы несколько признаков f_1, f_2, \dots, f_k , то вектор $x = (f_1(x), f_2(x), \dots, f_k(x))$ называется признаковым описанием x , а множество $D = D_{f_1} \times D_{f_2} \times \dots \times D_{f_k}$ – признаковым пространством [1,2].

2. Принципы функционирования сверточной нейронной сети

Изображения воспринимаются компьютерами как последовательность пикселей, каждый из которых имеет свои собственные значения цвета и свое особое значение. Физическое представление изображения в матрице – это матрица, в которой каждый пиксель является элементом с номером (и,

следовательно, элементом изображения) [5,6]. На рисунке 1 представлена обработка рисунка, в случае распознавания введенного изображения.

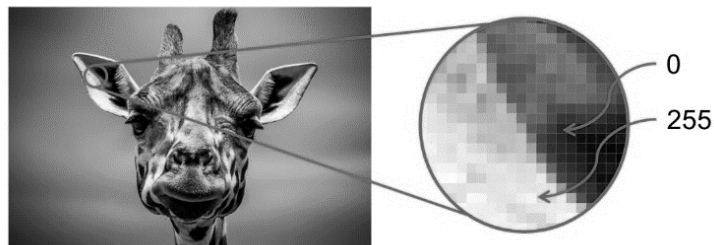


Рис. 1. – Пример интерпретации изображения в числа [6]

Алгоритм компьютерного зрения получает входные данные, а затем использует их для анализа и принятия решений при будущей обработке [6].

Сверточные нейронные сети (CNN) – это архитектура глубокого обучения, революционизирующая область компьютерного зрения и выходящая за её пределы. Их эффективность в обработке изображений обусловлена уникальным механизмом работы, основанным на операциях свертки. В отличие от обычных нейронных сетей, которые обрабатывают данные как плоский вектор, CNN обрабатывают изображения в их исходной пространственной форме, что позволяет им эффективно улавливать пространственные иерархические связи между пикселями. Эта способность делает их идеальным инструментом для задач, требующих понимания контекста и пространственного расположения признаков, рисунок 2 [6,7,8].

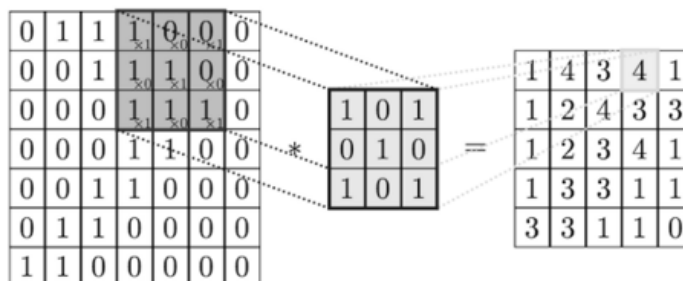


Рис. 2. – Выполнение операции свертки [6]

Основной блок CNN – это сверточный слой. Он использует фильтры (ядра), представляющие собой небольшие матрицы весов. Фильтр «сдвигается» по изображению, поэлементно умножая свои веса на соответствующие пиксели изображения в текущей области. Результаты этих умножений суммируются, давая в результате одно значение – элемент выходной карты признаков. Размерность выходной карты зависит от размера фильтра, шага сдвига (stride) и наличия/отсутствия (padding) (дополнительных пикселей по краям изображения, предотвращающих уменьшение размеров карты). Применение множества фильтров позволяет извлечь различные признаки из изображения – края, углы, текстуры и т.д., создавая многоканальную выходную карту. Этот процесс повторяется в последовательных сверточных слоях, каждый из которых обнаруживает более абстрактные и сложные признаки на основе признаков, извлеченных предыдущими слоями. Например, первый слой может обнаруживать края, второй – комбинации краев, формирующие углы и простые формы, а более глубокие слои – целые объекты. [6,7]. Схематически, полностью связанный слой представлен на рисунке 3.

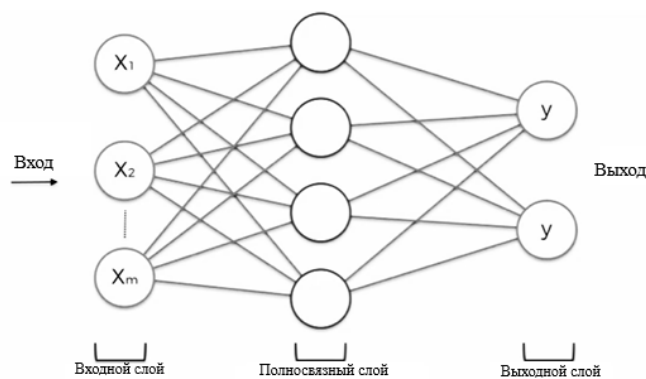


Рис. 3. – Пример полностью связного слоя [6]

После сверточных слоев часто используются пулинговые (pooling) слои, которые выполняют операцию уменьшения размерности выходной карты. Самая распространенная операция – это max-pooling, выбирающая

максимальное значение в заданной области (окне) карты признаков. Это снижает вычислительную сложность, делает модель более устойчивой к незначительным изменениям во входных данных (инвариантность к сдвигам) и помогает избежать переобучения. Кроме max-pooling существуют и другие виды пулинга, например, average-pooling, global average pooling и т.д., каждый со своими преимуществами и недостатками.

После нескольких чередующихся сверточных и пулинговых слоев следует полносвязный слой, который принимает выходной вектор из предыдущих слоев и производит окончательную классификацию или регрессию. В случае задач классификации изображений выходной слой обычно содержит столько нейронов, сколько классов нужно распознать. В случае задач обнаружения объектов архитектура может быть более сложной и включать в себя дополнительные механизмы для определения координат ограничивающих рамок (bounding boxes) объектов [8].

Обнаружение объектов – ключевая задача компьютерного зрения, которая выходит за рамки простой классификации. Цель – не только определить, какие объекты присутствуют на изображении, но и указать их местоположение. Существуют два основных подхода: одноэтапные и двухэтапные детекторы [8]. Одноэтапные детекторы, такие как YOLO (You Only Look Once) и SSD (Single Shot Detector), одновременно предсказывают классы и координаты ограничивающих рамок в одном проходе по изображению. Это обеспечивает высокую скорость работы, но часто с меньшей точностью, чем двухэтапные методы. Пример принципа обнаружения объектов представлен на рисунке 4.

Двухэтапные детекторы, такие как R-CNN (Regions with Convolutional Neural Networks) и его модификации (Fast R-CNN, Faster R-CNN), работают в два этапа [8,12]. Сначала они генерируют предложения областей (region proposals), которые потенциально содержат объекты. Затем, сверточная сеть

обрабатывает эти области, чтобы классифицировать и уточнить их границы. Двухэтапные детекторы обычно обеспечивают более высокую точность, но работают медленнее, чем одноэтапные. Выбор между одноэтапными и двухэтапными методами зависит от конкретных требований задачи – баланса между скоростью и точностью, рисунок 5.

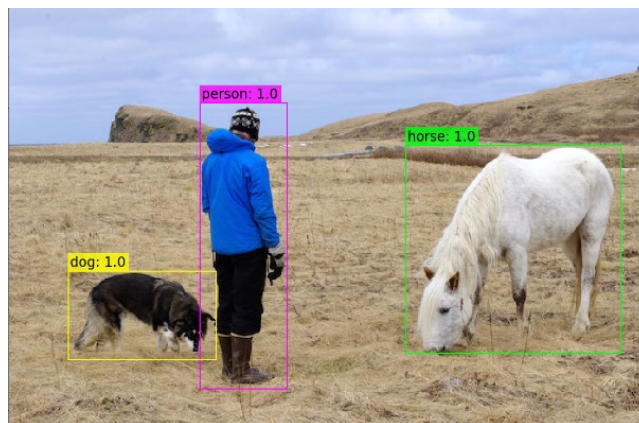


Рис. 4. – Пример работы алгоритма обнаружения объектов [8]

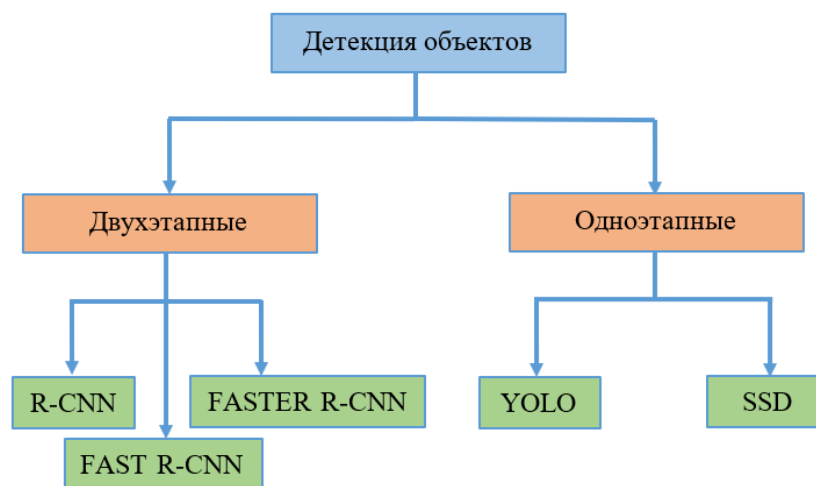


Рис. 5. – Схема детекции

Недавние достижения в области CNN привели к разработке архитектур, значительно превосходящих по производительности предыдущие модели. Например, ResNet (Residual Networks) ввели концепцию остаточных связей (skip connections), которые облегчают обучение глубоких сетей, позволяя им

обучаться эффективнее и достигать лучшей точности на сложных задачах. Другие архитектуры, такие как Inception и EfficientNet, используют различные техники для повышения эффективности и производительности.

CNN нашли применение не только в компьютерном зрении, но и в других областях, таких как обработка естественного языка (NLP) – для анализа последовательностей текста, анализ временных рядов, медицинская диагностика, автоматизированное вождение и многих других. Их адаптивность и мощь делают их незаменимым инструментом в современном мире искусственного интеллекта, и исследования в этой области продолжаются, приводя к появлению новых, еще более эффективных архитектур и алгоритмов. Дальнейшие исследования фокусируются на создании моделей, более устойчивых к шуму и неточностям данных, способных работать в режиме реального времени и требующих меньшего количества вычислительных ресурсов.

При этом, эффективность одноэтапных методов можно объяснить скоростью обнаружения, что делает их намного быстрее. Одноэтапное обнаружение объектов в реальном времени с использованием YOLOv7 является наиболее эффективным алгоритмом перечисления [13].

YOLOv7 – это популярное семейство алгоритмов обнаружения объектов в реальном времени. Первая версия алгоритма вышла в 2015 году, автором алгоритма являлся Джозеф Редмон [14].

Принцип работы алгоритма изображен на рисунке 6 и заключается в разбиении исходного изображения на сетку из $S \times S$ ячеек. После этого относительно центра каждой ячейки рисуется серия ограничивающих рамок разного размера и различной формы. Затем определяется, какой из ограничивающих прямоугольников подходит для заданного изображения, какой объект в нем содержится и есть ли он. На рисунке 7 показана структура сети [7,8,14].

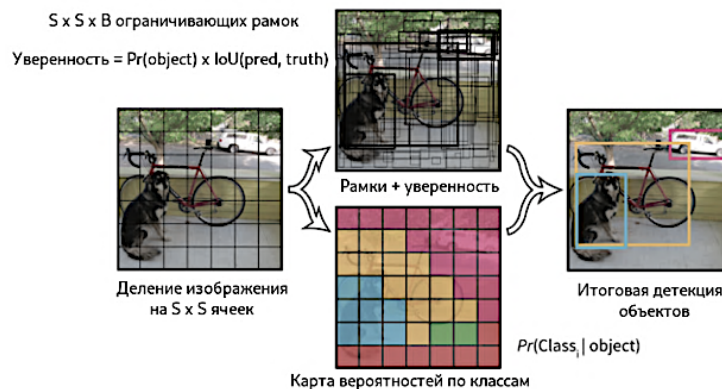


Рис. 6. – Принцип работы алгоритма YOLO [8]

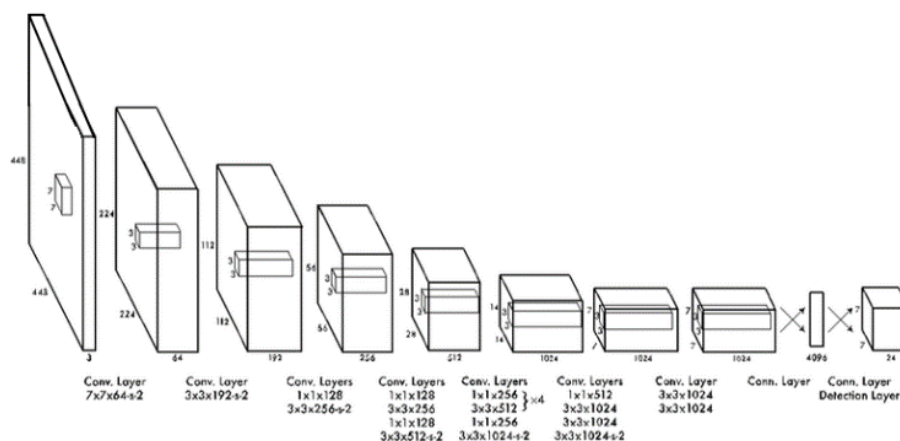


Рис. 7. – Архитектура нейронной сети для алгоритма YOLO [7]

3. Оценка эффективности алгоритмов обнаружения объектов

Правильная постановка задачи должна содержать критерий оценки работы алгоритма. При решении задачи классификации наименее трудозатратным, с точки зрения вычисления, критерием будет являться количество правильно классифицированных изображений. Пусть в представленных датасетах содержатся правильные варианты ответов и значения, полученные при помощи некоторого алгоритма. В таком случае, в качестве критерия будет использоваться процент правильно предсказанных ответов [8,10,12].

Определим ключевые показатели, которые используются для оценки алгоритмических методов обнаружения объектов, включая методики IoU, Precision score, Recall score. Обратим внимание, что представленные описания относятся исключительно к одному классу. Прямоугольная рамка, определяющая расположение одного экземпляра на изображении, называется ограничивающей рамкой. В зависимости от контекста ошибки в положении ограничительной рамки можно классифицировать как [10]:

- 1) true positives (TP) – объект присутствует на картинке и модель обнаружила его;
- 2) false negatives (FN) – объект присутствует, но модель не обнаружила его;
- 3) false positives (FP) – объект присутствует, но модель обнаружила его в неправильном месте или объект отсутствует, но модель обнаружила что-то.

IoU – отношение площадей предсказанных и истинных ограничивающих рамок.

IoU принимает диапазон значений от 0 до 1.

- 0, если рамки не пересекаются;
- 1, если происходит полное наложение, рисунок 8.



Рис. 8. – Примеры влияния положения рамок

Precision score – это отношение правильно предсказанных примеров ко всем предсказанным, в интервале от 0 до 1. Чем выше значение, тем больше вероятность того, что расположение ограничивающих рамок правильно.

Recall score – отношение правильно предсказанных примеров ко всем истинным, так же находится в интервале от 0 до 1. Чем выше значение, тем больше вероятность того, что расположение ограничивающих рамок покрывают все истинные рамки [8,11]. Соответственно метрики Precision и Recall дополняют друг друга, интерпретировать их показания можно так:

1) высокий Recall, но низкий Precision – означает что предсказания корректны, но часть истинных рамок была пропущена;

2) низкий Recall, но высокий Precision – означает что большая часть истинных рамок была обнаружена, но при этом было много ложных срабатываний [11].

Большая часть работы алгоритмов машинного обучения строится на нахождении ответов наибольшей точности при обращении к обучающей выборке. В связи с этим эффективность работы таких алгоритмов следует проверять на отложенной выборке, которая не подвергалась обучению.

Заключение

В статье рассмотрены основные подходы к решению задач компьютерного зрения с использованием нейронных сетей. Описаны типы задач, решаемых при помощи компьютерного зрения. Представлены механизмы функционирования сверточной нейронной сети. Рассмотрена задача выбора модели обнаружения объектов, обобщающая более изученную задачу классификации объектов. Также приведена оценка эффективности алгоритмов обнаружения объектов, на основе представленных в работе метрик.

Описан алгоритм обнаружения объектов в реальном времени YOLOv7. Принцип работы алгоритма заключается в разбиении исходного изображения на сетку из $S \times S$ ячеек. Представлена схема работы данного алгоритма и архитектура его нейронной сети.

Литература

1. Николенко С., Кадурич А., Архангельская Е. Глубокое обучение. – СПб.: Питер, 2021. – 480 с.
2. Davies E.R., Turk M.A. Advanced Methods and Deep Learning in Computer Vision. – Academic Press, 2022. – 690 p. – ISBN 978-0-12-822109-9.
3. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. – 2nd ed. – Sebastopol: O'Reilly Media, 2019. – 851 p. – ISBN 978-1-492-03264-9.
4. Горячкин Б.С., Китов М.А. Компьютерное зрение // E-Scio. – EDN: EBYPIO, № 9(48), 2020. – С. 317–345.
5. Gonzalez R.C., Woods R.E. Digital Image Processing. – 3rd ed. – Upper Saddle River: Prentice Hall, 2008. – 976 p. – ISBN 978-0-13-168728-8.
6. Кочанов Д.Н. Актуальные достижения в области компьютерного зрения // Сборник научных тезисов и статей «Хроники цифровых трансформаций». – Волгоград: Издательский дом «Сириус». 2022. – С. 29-36.
7. Султанова А.И., Павлов А.Д., Фаттахова К.А., Мокшин В.В. Определение нарушений техники безопасности на основе методов распознавания образов // Сборник материалов XI Всероссийской научно-практической конференции «Профессиональные коммуникации в научной среде – фактор обеспечения качества исследований». – СПб: Издательство ООО «Скифия-принт», 2022. – С. 190-194.
8. Зотов С.С., Яковлев А.А., Колчинцев Д.А. Обнаружение объектов в реальном времени с помощью алгоритмов распознавания YOLO // Синергия наук. № 26, 2018. – С. 388-404.
9. Лебедев В.Б., Лебедев О.Б. Синтез разделяющих функций при распознавании образов с использованием композитной архитектуры многоагентной системы // Материалы X Международной научно-технической

конференции «Технологии разработки информационных систем». – Таганрог: Южный федеральный университет, 2020. – С. 194-200.

10. Дударев Д.С., Дударев К.С., Мотайленко Л.В. Ретроспектива компьютерного зрения // Столыпинский вестник. – 2023. – Т. 5, № 2. – DOI 10.55186/27131424_2023_5_2_12. – EDN SAAUOE.

11. Коломеец М.В., Чечулин А.А., Дойникова Е.В. и др. Методика визуализации метрик кибербезопасности // Известия вузов. Приборостроение, Т. 61. № 10, 2018. – С. 873-880.

12. Дьяченко Р.А., Довгаль В.В., Гура Д.А. К вопросу сравнения эффективности нейронных сетей YOLOv7 и U-Net в задачах сегментации территориальных объектов // Научно-технический и научно-производственный журнал «Информационные технологии». – М.: Издательство «Новые технологии», Том 30, № 9, 2024. – С. 480-485.

13. Матвеев Ю., Кухарев Г., Каменская Е. и др. Методы обработки и распознавания изображений лиц в задачах биометрии. – СПб.: Издательство Политехника, 2013. – 198 с.

Wang C.-Y., Vochkovskiy A., Liao H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors // arXiv preprint arXiv:2207.02696. – 2022. URL: researchgate.net/publication/361807900_YOLOv7_Trainable_bag-of-freebies_sets_new_state-of-the-art_for_real-time_object_detectors.

References

1. Nikolenko S., Kadurin A., Arkhangle'skaya YE. Glubokoye obucheniye [Deep learning]. SPB. Piter, 2021. 480 p.

2. Davies E.R., Turk M.A. Advanced Methods and Deep Learning in Computer Vision. Academic Press, 2022. 690 p. ISBN 978-0-12-822109-9.

3. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems. 2nd ed. Sebastopol O'Reilly Media, 2019. 851 p. ISBN 978-1-492-03264-9.
4. Goryachkin B.S., Kitov M.A. E-Scio. EDN EBYPIO, № 9(48), 2020. pp. 317–345.
5. Gonzalez R.C., Woods R.E. Digital Image Processing. 3rd ed. Upper Saddle River: Prentice Hall, 2008. 976 p. ISBN 978-0-13-168728-8.
6. Kochanov D.N. Aktual'nyye dostizheniya v oblasti komp'yuternogo zreniya. Sbornik nauchnykh tezisov i statey «Khroniki tsifrovyykh transformatsiy». Volgograd: Izdatel'skiy dom «Sirius». 2022. pp. 29-36.
7. Sultanova A.I., Pavlov A.D., Fattakhova K.A., Mokshin V.V. Opredeleniye narusheniy tekhniki bezopasnosti na osnove metodov raspoznavaniya obrazov. Sbornik materialov XI Vserossiyskoy nauchno-prakticheskoy konferentsii «Professional'nyye kommunikatsii v nauchnoy srede – faktor obespecheniya kachestva issledovaniy». SPb: Izdatel'stvo «Skifiya-print», 2022. pp. 190-194.
8. Zotov S.S., Yakovlev A.A., Kolchintsev D.A. Sinergiya nauk № 26, 2018. pp. 388-404.
9. Lebedev V.B., Lebedev O.B. Materialy X Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii «Tekhnologii razrabotki informatsionnykh sistem». Taganrog: Yuzhnyy federal'nyy universitet, 2020. pp. 194-200.
10. Dudarev D.S., Dudarev K.S., Motaylenko L.V. Stolypinskiy vestnik. 2023. V. 5, № 2. DOI 10.55186/27131424_2023_5_2_12. EDN SAAUOE.
11. Kolomeyets M.V., Chechulin A.A., Doynikova YE.V. Izvestiya vuzov. Priborostroyeniye, V. 61. № 10, 2018. pp. 873-880.
12. D'yachenko R.A., Dovgal' V.V., Gura D.A. Nauchno-tekhnicheskii i nauchno-proizvodstvennyy zhurnal «Informatsionnyye tekhnologii». M. Izdatel'stvo «Novyye tekhnologii», Tom 30, № 9, 2024. pp. 480-485.



13. Matveyev YU., Kukharev G., Kamenskaya YE. and other. Metody obrabotki i raspoznavaniya izobrazheniy lits v zadachakh biometrii [Methods of processing and recognizing facial images in biometrics problems]. SPb. Izdatel'stvo Politekhnik, 2013. 198 pp.

14. Wang C.-Y., Bochkovskiy A., Liao H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696. 2022. URL: researchgate.net/publication/361807900_YOLOv7_Trainable_bag-of-freebies_sets_new_state-of-the-art_for_real-time_object_detectors.

Дата поступления: 27.12.2024

Дата публикации: 5.02.2025