УДК 004.032.26 URL: https://ptsj.bmstu.ru/catalog/icec/inf_tech/983.html

СИСТЕМА РАСПОЗНАНИЯ РАЗЛИЧНЫХ СОРТОВ ЯБЛОК НА ОСНОВЕ НЕЙРОННОЙ СЕТИ YOLOV8X

Д.А. Михеев mda18m121@student.bmstu.ru
Д.Н. Китаев kdn22mm043@student.bmstu.ru

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Проведен выбор нейросетевой модели для решения задачи детекции весового товара, рассмотрено семейство одноступенчатых моделей сверточных нейронных сетей YOLOv8, произведена первичная оценка работоспособности самой крупной модели YOLOv8x на кадрах с изображениями фруктов и овощей в продуктовом магазине, собраны и подготовлены данные для обучения оцененной сети для распознания пяти сортов яблок: Golden Delicious (Голден Делишес), Granny Smith (Гренни Смит), Gala (Гала), Honey Crisp (Медовый хруст) и Red Chief (Ред Чиф), на полученных данных обучена модель YOLOv8x с использованием трансферного обучения, проанализированы результаты работы обученной модели.

Ключевые слова: компьютерное зрение, распознание объектов, нейронные сети, сверточная нейронная сеть, одноступенчатый детектор, YOLOv8x, весовой товар, сорта яблок, розничная торговля

Введение. В погоне за оптимизацией и автоматизацией процессов люди создают новые технологии и в дальнейшем развивают их. Значительный рост вычислительной мощности компьютеров и появление новых математических моделей и алгоритмов позволило добиться значительного прогресса в области компьютерного зрения. Техническое или машинной зрение — это совокупность технологий, направленных на получение и обработку информации, путем использования камер, датчиков и вычислительного устройства.

Перспективным направлением автоматизации благодаря внедрению систем компьютерного зрения является розничная торговля. В данной работе рассмотрена нейросетевая детекция весового товара.

Существуют следующие проблемы, которые можно решить благодаря введению нейросетевых технологий в систему взвешивания товара:

1) человеческий фактор — ошибки взвешивания;

2) несколько листов с названиями и фотографиями товаров на экране весов в зале, долгий поиск товара покупателем;

3) подмена весового товара недобросовестными покупателями;

4) трата времени кассира на поиск кода весового товара в базе данных.

 ISSN 2541-8009

По статистике торговой сети «Пятерочка» [1], весовые товары входят в 30 % чеков. В течение одного часа кассир взвешивает продукты около 47 раз, каждая операция занимает в среднем 14,5 с. Предлагаемая система помогает кассирам существенно сократить это время и значительно ускорить обслуживание покупателей. В расчете на один магазин пропускная способность увеличивается на 20 000 человек в год, а годовая экономия времени на обслуживании покупателей на кассе составит не менее 1300 ч.

Таким образом, весы с компьютерным зрением позволят увеличить скорость обслуживания и уменьшить убытки, получаемые в результате невнимательности или недобросовестности покупателей.

Постановка задачи. Для интеллектуальной системы взвешивания необходимо разработать нейросетевой алгоритм распознания нескольких сортов яблок: Golden Delicious (Голден Делишес), Granny Smith (Гренни Смит), Gala (Гала), Honey Crisp (Медовый хруст), Red Chief (Ред Чиф). Все перечисленные сорта яблок представлены на рис. 1.

Рис. 1. Требуемые для распознания сорта яблок:
1 — Голден Делишес, 2 — Гренни Смит, 3 — Гала, 4 — Медовый хруст, 5 — Ред Чиф

Алгоритм должен быть достаточно быстрым, чтобы классифицировать яблоки в режиме реального времени — количество обрабатываемых изображений в секунду 30 и выше. Точность предсказания сорта яблока должна быть не ниже 0,85.

Выбор модели нейронной сети. Существующие нейросетевые модели глубокого обучения можно подразделить на одноступенчатые (one-stage detector) и двухступенчатые (two-stage detector). В одноступенчатых архитектурах входное изображение обрабатывается нейронной сетью один раз, в двухступенчатых — два. К первым относят нейронные сети YOLO, SSD, RetinaNet, ко вторым — RCNN и SPPNet, Fast RCNN и Faster RCNN, Mask R-CNN, Pyramid Networks/FPN, G-RCNN.

В системах обнаружения объектов, работающих в режиме реального времени, применяют одноступенчатые модели, поскольку они обладают наилучшими временными показателями обработки изображения.

В настоящее время самой эффективной одноступенчатой нейронной сетью является YOLO (от англ. *You Only Look Once* — ты смотришь только раз) [2]. Последняя версия этой модели — YOLOv8, запущенная 10 января 2023 г. [3], обученная на датасете (набор размеченных данных) Microsoft COCO (от англ. *Common Objects in Context* — реальные объекты в контексте) [4].

Согласно информации, размещенной разработчиками YOLO, восьмая версия нейронной сети является самой быстрой и самой точной из всей серии, что можно увидеть на рис. 2 [2]. По оси абсцисс на графике слева указано количество параметров нейронных сетей в миллионах, на графике справа — среднее количество времени в миллисекундах, затрачиваемое нейронными сетями на обработку одного изображения на графическом процессоре NVIDIA A100 с представлением значений весов моделей в 16 бит. По оси ординат на обоих графиках отложены значения метрики mAP (от англ. *Mean Average Precision* — интерполированная средняя точность) [5] при значении IoU (от англ. *Intersection over Union* — пересечение по объединению) [4] от 0,50 до 0,95. Стрелками обозначены направления уменьшения количества параметров моделей и их времени обработки одного кадра на левом и правых графиках соответственно.

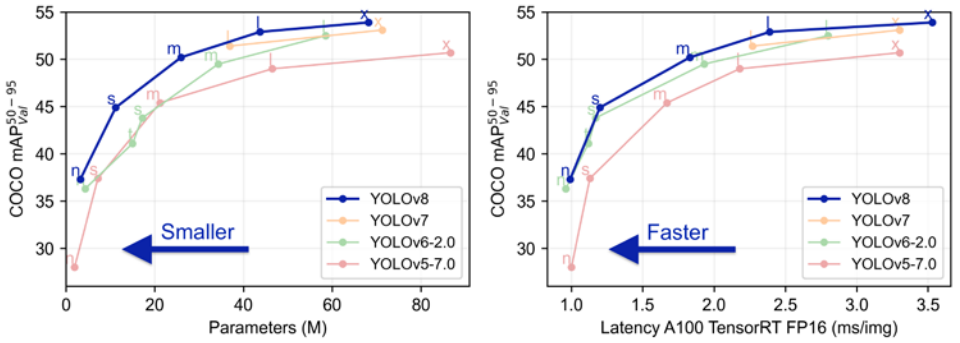

Рис. 2. Сравнительные графики последних версий YOLO [2]

Как и предыдущие версии YOLO, она включает в себя несколько вариаций, представленных в таблице.

Под размером входного изображения понимают количество пикселей изображения в ширину и высоту, информация о которых подается на вход нейронной сети.

COCO mAP^{50-95}_{Val} — метрика mAP, учитывающая, кроме точности прогноза областей объектов, еще и достоверность правильного обнаружения, что позволяет сравнивать эффективность различных алгоритмов обнаружения.

 ISSN 2541-8009

Достоверность правильного обнаружения определяется точностью (precision), характеризующей способность алгоритма обнаруживать именно нужные объекты, и полнотой (recall) обнаружения [6], характеризующей способность алгоритма находить нужный объект (аномалию) в полном объеме (в идеале все объекты заданного класса). Дополнения COCO и $^{50-95}_{\text{Val}}$ означают, что интерполированная средняя точность модели рассчитана при использовании датасета Microsoft COCO с IoU = 0,50…0,95.

Варианты моделей-детекторов YOLOv8 и их характеристики [2]

Вариант модели	Размер входного изображения, пикселей	COCO $\text{mAP}^{50-95}_{\text{Val}}$	Speed CPU ONNX, мс	Speed Latency A100 TensorRT FP16, мс/изображение	Количество параметров, млн	FLOPs, млрд
YOLOv8n (nano)	640	37,3	80,4	0,99	3,2	8,7
YOLOv8s (small)	640	44,9	128,4	1,20	11,2	28,6
YOLOv8m (medium)	640	50,2	234,7	1,83	25,9	78,9
YOLOv8l (large)	640	52,9	375,2	2,39	43,7	165,2
YOLOv8x (extra large)	640	53,9	479,1	3,53	68,2	257,8

Пересечение по объединению между размеченным и предсказанным объектом IoU определяют как отношение пересечения размеченного и предсказанного объектов к их объединенной площади. Точность и полноту рассчитывают по следующим формулам:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}};$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

где TP — true positive, количество верных предсказаний; FP — false positive, количество ложных предсказаний; FN — false negative, количество не предсказанных, но присутствующих на изображении объектов.

Значение AP получают вычислением площади под графиком зависимости точности от полноты. Значение mAP получают вычислением среднего значения из полученных AP для каждого класса.

Speed CPU ONNX — метрика производительности, которая измеряет скорость компьютерной системы при выполнении модели ONNX на цен-

тральном процессоре. ONNX (от англ. *Open Neural Network Exchange* — открытый обмен нейронными сетями) [7] — открытая библиотека программного обеспечения для построения нейронных сетей глубокого обучения.

Speed Latency A100 TensorRT FP16 — количество времени, которое требуется компьютерной системе для обработки заданной задачи с использованием графического процессора NVIDIA A100 Tensor Core в сочетании с программным обеспечением TensorRT с использованием 16-битного представления данных с плавающей точкой FP16.

Под параметрами нейронной сети обычно понимают веса и смещения, которые определяют силу связей между нейронами и порог активации каждого нейрона соответственно.

FLOP (от англ. *FLoating-point OPerations per Second* — число операций с плавающей точкой в секунду) [8] — единица измерения производительности вычислительных систем, характеризующая максимальную вычислительную мощность системы для операций с плавающей точкой.

Поскольку разрабатываемая система не обладает высокой динамикой, а разница между некоторыми сортами яблок выражена не ярко, решено использовать самую большую и точную модель YOLOv8x.

Описание модели YOLO. Нейронная сеть YOLOv8, как и все модели семейства YOLO, относится к одноступенчатым детекторам, архитектура которых состоит из нескольких частей (рис. 3): input (входной слой), backbone (экстрактор признаков), neck (комбинатор признаков), dense prediction (выходной предсказывающий слой).

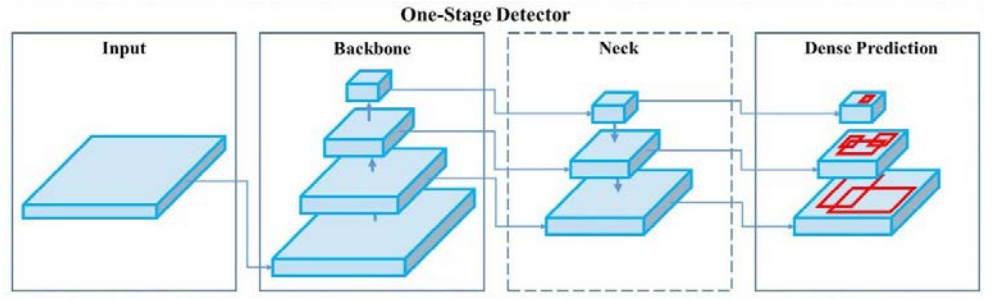

Рис. 3. Архитектура одноступенчатой нейронной сети

Входной слой принимает изображение размером 640×640 пикселей и передает его backbone. Экстрактор признаков извлекает feature maps (набор признаков) из поступившего изображения и передает его комбинатору. В YOLO изображение дублируется несколькими с различными масштабами

 ISSN 2541-8009

для отдельного извлечения крупных и мелких деталей. Комбинатор признаков получает от экстрактора набор элементарных признаков, комбинирует их в более сложные и передает их выходному предсказывающему слою. Выходной слой предсказывает объект на изображении по полученному набору признаков. На выходе нейронной сети получается тензор, содержащий в себе координаты центра и размер ограничивающей объект рамки, вероятность нахождения в этой рамке какого-либо объекта и вероятность для каждого предсказываемого класса.

В архитектуре YOLOv8 neck и dense prediction объединены в одну часть — head.

Обучение. При обучении глубоких нейронных сетей с ненастроенными весами возникает две проблемы: необходимость создания большого обучающего набора размеченных изображений и подключение больших вычислительных ресурсов для завершения работы в срок.

Решить эти проблемы помогает методика transfer learning (трансферное обучение), представляющая собой дообучение pre-training model (предварительно обученная модель). Такой подход позволяет повторно натренировать последний слой CNN-сети (от англ. *convolutional neural network* — сверточная нейронная сеть) с помощью собственного набора изображений в приемлемое время, не изменяя веса других слоев и достигая необходимой точности.

Был проведен анализ исходной модели. Он показал, что нейронная сеть в том виде, в котором она сейчас существует, не способна различить яблоки разных сортов, поскольку в ней отсутствуют соответствующие классы (рис. 4). Помимо этого модель допускает ошибки распознавания (на одном из изображений груши обозначены как яблоки), в некоторых случаях дублирует ограничивающие объект рамки.

Необходимо подготовить и разметить изображения яблок заданных сортов, внести новые классы в архитектуру сети и удалить неиспользуемые.

Было получено 837 изображений пяти сортов яблок, в среднем 167 изображений на один класс. Для разметки объектов использовался онлайн сервис roboflow [9]. Результатом разметки в данном случае был набор одноименных изображений и текстовых файлов, а также файл data.yaml. В текстовом файле построчно для каждого объекта представлена информация, разделенная пробелом: номер класса (0 — Gala, 1 — Golden, 2 — Granny Smith, 3 — Honey Crisp, 4 — Red Chief), относительная координата центра объекта по оси X, относительная координата центра объекта по оси Y, относительная ширина объекта и его относительная ширина. Относительные величины имеют значения от 0 до 1 и рассчитываются как отношение их модулей к соответствующим размерам изображения. Файл data.yaml содержит информацию о рас-

положении изображений и текстовых файлов, количество классов и расшифровку номера каждого из них. Визуальный результат разметки и пример одного размеченного изображения представлены на рис. 5.

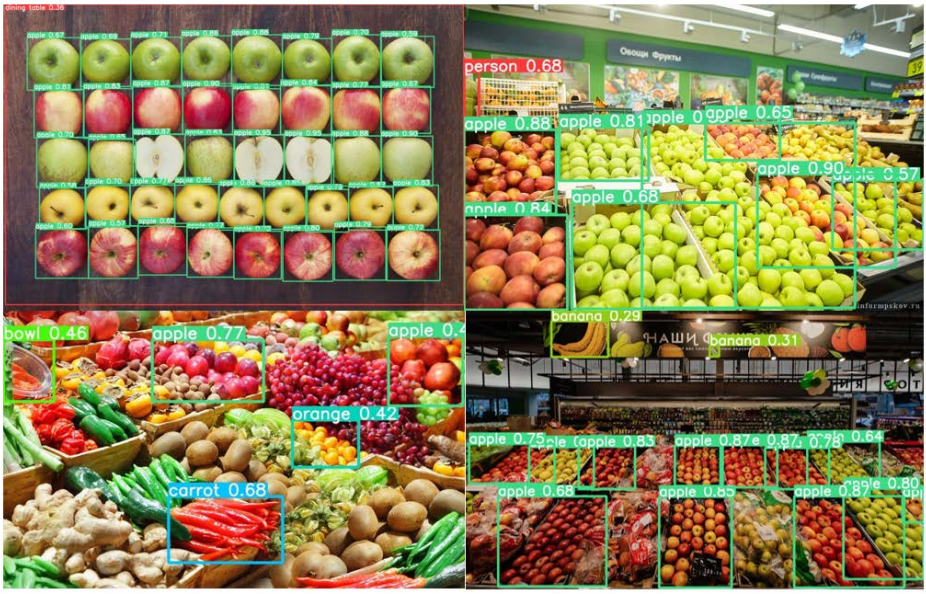

Рис. 4. Обнаружение яблок с помощью YOLOv8x

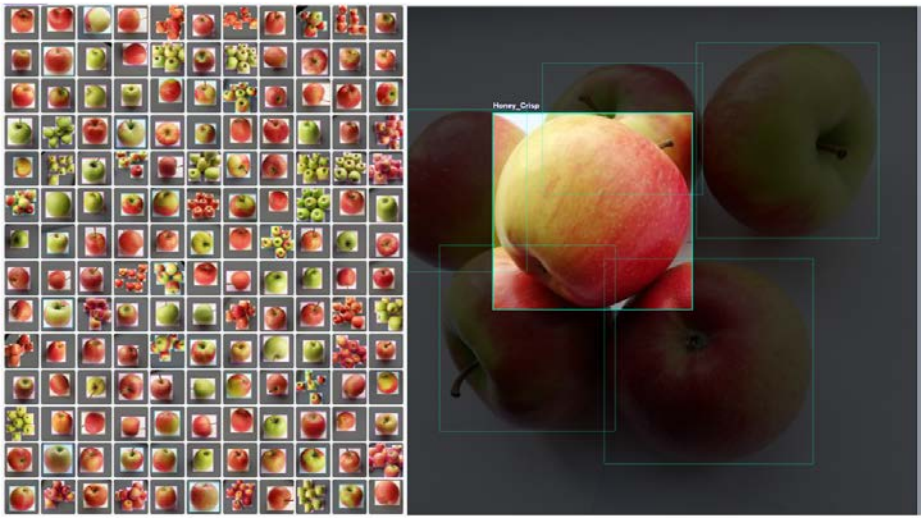

Рис. 5. Размеченные изображения

 ISSN 2541-8009

Для достижения наибольшей точности нейронные сети обучают на большом количестве данных. Самый эффективный метод, используемый для искусственного расширения датасета, — data augmentation («аугментация», расширение данных).

Существует несколько вариантов аугментации данных [10]:

– отражение изображения относительно заданных осей;

– изменение цветовых характеристик изображения (изменение контрастности, яркости, выделение одного цвета и др.);

– обрезка изображения;

– поворот изображения на заданную величину;

– размытие, смещение изображения;

– введение в изображение шумов.

В данной работе для аугментации данных использовались инструменты из библиотеки Albumentations [11]. Аугментация проводилась над тренировочным набором данных, его количество увеличилось в 10 раз и стало равно 5850. Примеры аугментированных изображений приведены на рис. 6.

Рис. 6. Изображения после процедуры агументации

Обучение проводилось с использованием программно-аппаратной архитектуры параллельных вычислений CUDA (от англ. *Compute Unified Device Architecture* — архитектура унифицированных вычислительных устройств) [12] для переноса вычислений на графический процессор Nvidia GA104. Обучение сети продлилось 140 эпох (однократное прохождение всего обучающего набора данных через нейронную сеть). По результатам обучения были построены графики ошибки (рис. 7), метрик mAP при IoU = 0,50 и 0,50…0,95, точности и полноты (рис. 8), а также матрица ошибок (рис. 9).

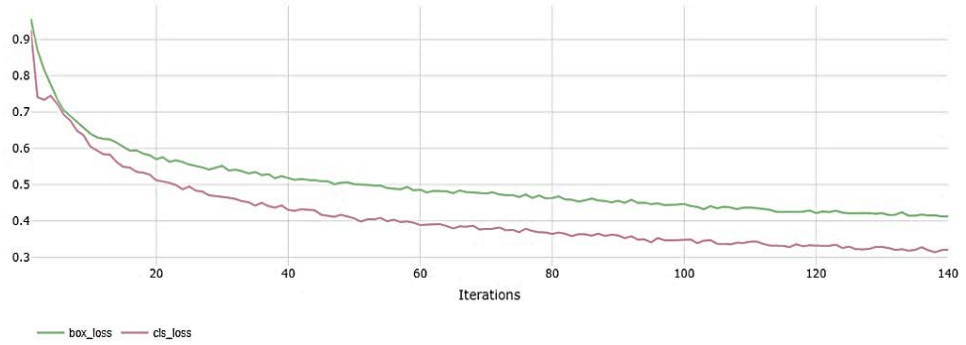

Рис. 7. График ошибки позиционирования объекта (*зеленый* — box_loss) и график ошибки предсказания класса объекта (*розовый* — cls_loss). Ось *X* — количество прошедших эпох обучения (Interations), ось *Y* — значение ошибки

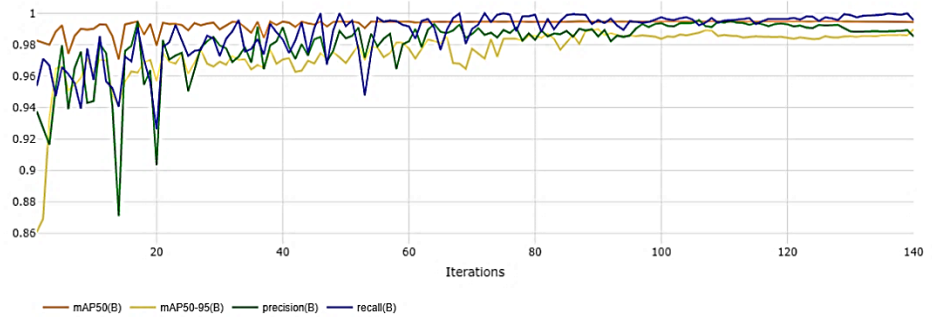

Рис. 8. График метрики mAP при IoU = 0,50 (*оранжевый* — mAP50), график метрики mAP при IoU = 0,50…0,95 (*желтый* — mAP50-95), график точности обнаружения объекта (*зеленый* — precision), график полноты обнаружения объектов (*синий* — recall). Ось *X* — количество прошедших эпох обучения (Interations), ось *Y* — значение от 0 до 1

На рис. 8 видно, что ошибки с каждой эпохой уменьшалась, и чем ближе к концу обучения, тем меньше изменялось значение ошибки, графики устремлены к константным значениям. По рис. 9 можно заключить, что метрики mAP, точность и полнота обнаружения для полученной модели составляет приблизительно 0,98. Согласно матрице ошибок (см. рис. 9), в целом все классы предсказываются верно, лишь в редких случаях с вероятностью от 0,01 до 0,33 некоторые объекты предсказываются ошибочно.

На рис. 10 приведены некоторые результаты работы обученной модели на валидационном наборе изображений, служащим для оценки предсказания полученной нейронной сети.

 ISSN 2541-8009

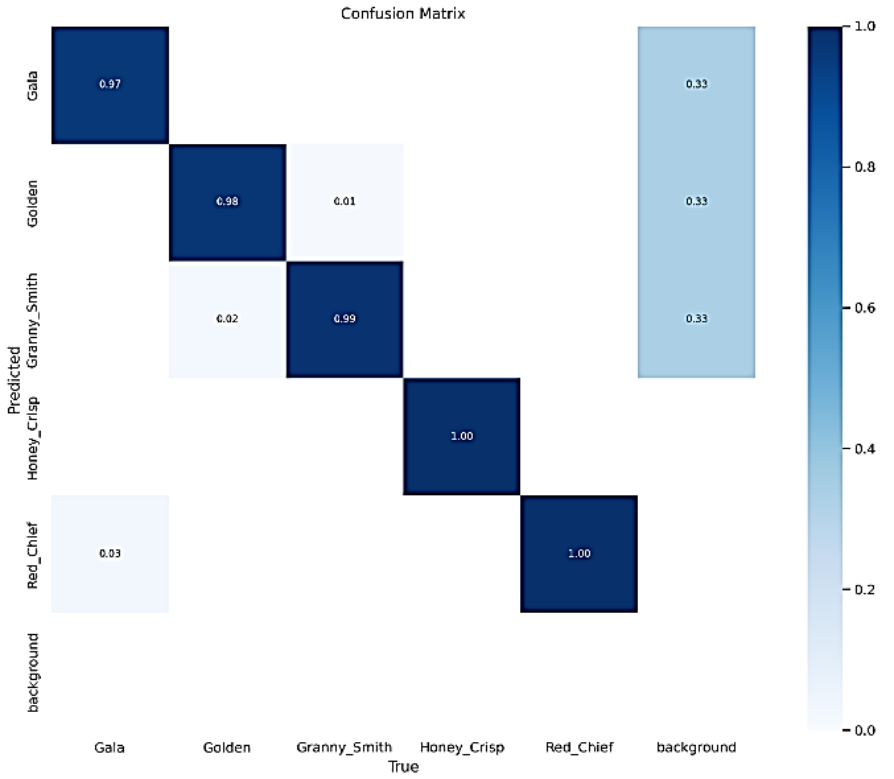

Рис. 9. Матрица ошибок. Слева по вертикали — предсказываемый объект, снизу по горизонтали — предсказанный объект, справа по вертикали — цветовая шакала вероятности предсказания

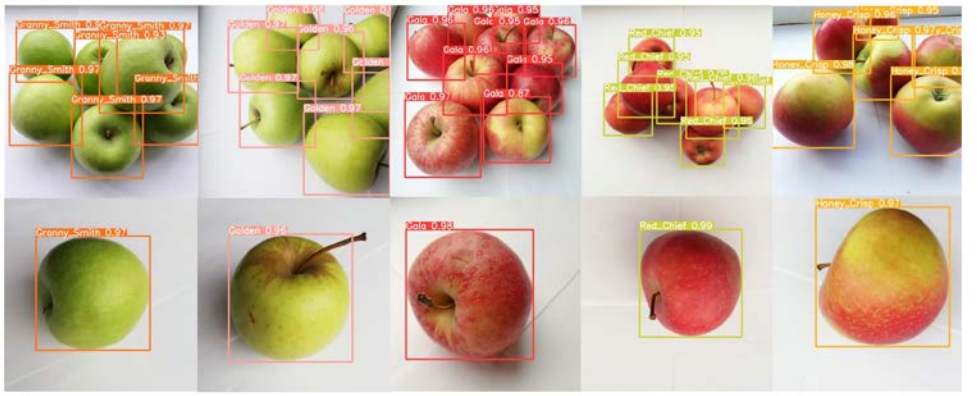

Рис. 10. Результат работы дообученной модели

На рис. 10 видно, что модель отлично распознает заданные сорта яблок как в единичном экземпляре, так и в группе с перекрытием. Также обученная модель с высокой точностью определяет границы объекта. Обработка одного изображения занимает в среднем 33 мс при запуске на графическом процессоре NVIDIA GeForce RTX 3060 Ti или приблизительно 30 кадров в секунду.

Заключение. Проведен выбор нейросетевой модели для решения задачи детекции весового товара. Рассмотрено семейство одноступенчатых моделей сверточных нейронных сетей YOLOv8. Подготовлен набор целевых кадров для обучения нейронной сети, пять сортов яблок: Golden Delicious (Голден Делишес), Granny Smith (Гренни Смит), Gala (Гала), Honey Crisp (Медовый хруст) и Red Chief (Ред Чиф). На полученных данных обучена модель YOLOv8x с использованием трансферного обучения и аугментации. Значение метрики mAP при IoU = 0,50…0,95, точность и полнота обнаружения обученной модели составили приблизительно 0,98. Скорость работы сети составила 33 мс на один кадр или 30 кадров в секунду на графическом процессоре NVIDIA GeForce RTX 3060 Ti.

Литература

[1] *X5 Retail Group N.V. Умные весы X5 определяют товар без участия человека.* URL: https://www.x5.ru/ru/PublishingImages/Pages/Media/News/x5_smart_weight.pdf (дата обращения 15.04.2024).

[2] *GitHub, ultralytics repository, YOLOv8*. URL: https://github.com/ultralytics/ultralytics (accessed April 15, 2024).

[3] *Roboflow, Jacob Solawetz, Francesco “What is YOLOv8? The Ultimate Guide”*. Available at: https://blog.roboflow.com/whats-new-in-yolov8/ (accessed April 15, 2024).

[4] Андриянов Н.А., Дементьев В.Е., Ташлинский А.Г. Обнаружение объектов на изображении: от критериев Байеса и Неймана — Пирсона к детекторам на базе нейронных сетей EfficientDet. *Компьютерная оптика*, 2022, т. 46, № 1, с. 139–159. https://doi.org/10.18287/2412-6179-CO-922

[5] Hui J. *mAP (mean Average Precision) for object detection*. Available at: https://jonathan-hui.medium.com/map-meanaverage-precision-for-object-detection-45c121a31173 (accessed April 15, 2024).

[6] Powers D. Evaluation: From precision, recall and f-measure to ROC, Informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2011, vol. 2 (1), pp. 37–63.

 ISSN 2541-8009

[7] *Open Neural Network Exchange*. Available at: https://onnx.ai/ (accessed April 15, 2024).

[8] *Computer Hope, Dictionary, Definitions, "FLOPS"*. Available at: https://www.computerhope.com/jargon/f/flops.htm (accessed April 15, 2024).

[9] *Roboflow, "Quickly Label Training Data and Export to Any Format"*. Available at: https://roboflow.com/annotate (accessed April 15, 2024).

[10] Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection. *Proc IEEE Conf on Computer Vision and Pattern Recognition (CVPR)*, 2016, vol. 1, pp. 779–788. https://doi.org/10.48550/arXiv.1506.02640

[11] *Albumentations, "Computer vision tool that boosts the performance of deep convolutional neural networks"*. Available at: https://albumentations.ai/ (accessed April 15, 2024).

[12] Očkay M., Harakal M., Líška M. Compute Unified Device Architecture (CUDA) GPU programming model and possible integration to the parallel environment. *Science & Military*, no. 2, vol. 3, 2008, pp. 64–68.

Поступила в редакцию 24.04.2024

Михеев Дмитрий Александрович — студент магистратуры кафедры «Робототехнические системы и мехатроника», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Китаев Дмитрий Николаевич — студент магистратуры кафедры «Робототехнические системы и мехатроника», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Рубцов Василий Иванович, доцент кафедры «Робототехнические системы и мехатроника», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:

Михеев Д.А., Китаев Д.Н. Система распознания различных сортов яблок на основе нейронной сети YOLOv8x. *Политехнический молодежный журнал*, 2024, № 03 (92). URL: https://ptsj.bmstu.ru/catalog/icec/inf_tech/983.html

A SYSTEM FOR IDENTIFYING DIFFERENT APPLE VARIETIES BASED ON THE YOLOV8X NEURAL NETWORK

D.A. Mikheev mda18m121@student.bmstu.ru
D.A. Kitaev kdn22mm043@student.bmstu.ru

Bauman Moscow State Technical University, Moscow, Russian Federation

The paper presents results of selecting a neural network model to solve the problem of identifying the weighed products. It considers a family of the single-stage models of the YOLOv8 convolutional neural networks and assesses at the initial stage performance of the largest YOLOv8x model on the frames with images of fruits and vegetables in a grocery store. Data were collected and prepared for the assessed network learning to recognize five apple varieties: Golden Delicious, Granny Smith, Gala, Honey Crisp and Red Chief. The obtained data was introduced to learn the YOLOv8x model using the transfer learning; results of the learned model operation were analyzed.

Keywords: computer vision, object identification, neural networks, convolutional neural network, single-stage detector, YOLOv8x, weighed products, apple varieties, retail

Received 24.04.2024

Mikheev D.A. — Master's Program Student, Department of Robotic Systems and Mechatronics, Bauman Moscow State Technical University, Moscow, Russian Federation.

Kitaev D.N. — Master's Program Student, Department of Robotic Systems and Mechatronics, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — Rubtsov V.I., Associate Professor, Department of Robotic Systems and Mechatronics, Bauman Moscow State Technical University, Moscow, Russian Federation.

Please cite this article in English as:
Mikheev D.A., Kitaev D.N. A system for identifying different apple varieties based on the YOLOV8X neural network. *Politekhnicheskiy molodezhnyy zhurnal*, 2024, no. 03 (92). (In Russ.). URL: https://ptsj.bmstu.ru/catalog/icec/inf_tech/983.html

 ISSN 2541-8009