

Сверточная нейронная сеть для сегментации цветков яблони на изображениях

© 2024. А. И. Кутырёв✉

ФГБНУ «Федеральный научный агроинженерный центр ВИМ», г. Москва,
Российская Федерация

В статье предложен метод оценки интенсивности цветения яблони для выполнения технологической операции прореживания, включающий подготовку набора данных (dataset) и обучение моделей сверточных нейронных сетей YOLOv8-seg (n, s, m, l, x) для сегментации изображений. В исследованиях применена техника трансферного обучения, использованы предварительно обученные модели на наборе данных датасета COCO (Common Objects in Context). Набор изображений цветков яблони собран с использованием камеры GoPro HERO 11. Аннотация (разметка) изображений была выполнена на платформе RoboFlow с применением инструментов выделения прямоугольных рамок (Bounding-box) и аннотации полигонов (Polygon Annotation and Labeling). Для расширения набора данных и повышения обобщающей способности моделей при обучении проведена аугментация исходных изображений, включающая горизонтальное отражение, горизонтальный поворот на 90°, поворот от -15° до +15°, добавление шумов до 5 % пикселей, размытие до 2,5 пикселей, горизонтальные и вертикальные сдвиги от -10° до 10°, а также изменение оттенка цветов от -15° до +15°. Метрики бинарной классификации, такие как точность Precision (точность) и Recall (полнота) использованы для оценки качества работы обученных моделей по распознаванию цветков яблони на изображениях при использовании ограничивающих рамок и масочной сегментации. Функция потерь $Loss_{Box/Mask}$ использована для оценки ошибок модели в определении ограничивающих рамок и масок сегментации объектов на изображении в процессе обучения. В результате проведенных исследований определены гиперпараметры алгоритма машинного обучения модели YOLOv8-seg для распознавания, классификации и сегментации изображений цветков яблони. Установлено, что модели YOLOv8x-seg (метрика mAP50 = 0,591) и YOLOv8l-seg (метрика mAP50 = 0,584) показывают более высокую производительность при распознавании цветков яблони. Скорость обработки кадров (Frame Rate, FR) моделями сверточных нейронных сетей составила от 10,27 (модель YOLOv8x-seg) до 57,32 кадра/с (модель YOLOv8n-seg). Средняя абсолютная ошибка моделей при распознавании цветков яблони и подсчете их количества на тестовой выборке не превышает 9 %.

Ключевые слова: трансферное обучение, распознавание цветков яблони, прореживание, машинное обучение, компьютерное зрение, цифровой мониторинг

Благодарности: работа выполнена при поддержке Минобрнауки РФ в рамках Государственного задания ФГБНУ «Федеральный научный агроинженерный центр ВИМ» (тема № FGUN-2022-0011).

Автор благодарит рецензентов за их вклад в экспертную оценку этой работы.

Конфликт интересов: автор заявил об отсутствии конфликта интересов.

Для цитирования: Кутырёв А. И. Сверточная нейронная сеть для сегментации цветков яблони на изображениях. Аграрная наука Евро-Северо-Востока. 2024;25(5):949–961. DOI: <https://doi.org/10.30766/2072-9081.2024.25.5.949-961>

Поступила: 13.05.2024

Принята к публикации: 30.09.2024

Опубликована онлайн: 30.10.2024

Convolutional neural network for segmentation of apple blossoms in images

© 2024. Alexey I. Kuttyrev✉

Federal Scientific Agroengineering Center VIM, Moscow, Russian Federation

The article provides a method for assessing the intensity of apple blossom for the thinning technological operation, including dataset preparation and training of YOLOv8-seg convolutional neural network models (n, s, m, l, x) for image segmentation. Transfer learning technique was applied in the research, utilizing pretrained models on the COCO dataset (Common Objects in Context). The apple blossom image dataset was captured using a GoPro HERO 11 camera. Image annotation was performed on the RoboFlow platform using tools for bounding box and polygon annotation and labeling. To expand the dataset and improve the models' generalization during training, augmentation of original images was conducted, including horizontal flipping, horizontal rotation by 90°, rotation from -15° to +15°, adding noise up to 5% of pixels, blurring up to 2.5 pixels, horizontal and vertical shifts from -10° to 10°, and color hue adjustment from -15° to +15°. Binary classification metrics such as Precision and Recall were used to evaluate the performance of trained models in recognizing apple blossoms in images using bounding boxes and mask segmentation. The Loss(Box/Mask) loss function was used to assess model errors in determining bounding boxes and segmentation masks of objects in images during training. The hyperparameters of the YOLOv8-seg model for image recognition, classification, and segmentation of apple blossom images were identified through the conducted research. It was determined that the YOLOv8x-seg (mAP50 metric = 0.591) and YOLOv8l-seg (mAP50 metric = 0.584) models demonstrate higher performance in apple blossom recognition. The frame processing speed (Frame Rate, FR) of convolutional neural network models ranged from 10.27 (YOLOv8x-seg model) to 57.32 (YOLOv8n-seg model). The average absolute error of the models in recognizing apple blossoms and counting their quantity in the test dataset does not exceed 9 %.

Keywords: transfer learning, apple blossom recognition, thinning, machine learning, computer vision, digital monitoring

Acknowledgements: the research was carried out under the support of the Ministry of Science and Higher Education of the Russian Federation within the state assignment of the Federal Scientific Agroengineering Center VIM (theme No. FGUN-2022-0011).

The author thanks the reviewers for their contribution to the peer review of this work.

Conflict of interest: the author stated that there was no conflict of interest.

For citations: Kuttyrev A. I. Convolutional neural network for segmentation of apple blossoms in images. *Agrarnaya nauka Evro-Severo-Vostoka* = Agricultural Science Euro-North-East. 2024;25(5):949–961. (In Russ.).

DOI: <https://doi.org/10.30766/2072-9081.2024.25.5.949-961>

Received: 13.05.2024

Accepted for publication: 30.09.2024

Published online: 30.10.2024

Интенсивность цветения яблони имеет ключевое значение для оценки последующего формирования плодов, зависит от множества факторов, включая погодные условия, состояние почвы и растения [1, 2]. Избыточное цветение и, как следствие, излишнее количество плодов у яблоневых деревьев вызвано стремлением растений максимизировать вероятность опыления и размножения в изменчивых условиях окружающей среды. Научные исследования, проведенные в данной области, подтверждают, что своевременное и систематическое проведение операции прореживания цветков яблони обеспечивает ежегодную продуктивность с прогнозируемым качеством плодов и оказывает существенное влияние на сроки начала плодоношения и устойчивость кроны [3].

В настоящее время лишние цветки удаляются механическим, химическим и ручным прореживанием [4]. Процесс прореживания характеризуется высокой степенью неточности, в значительной степени зависит от опыта агронома. Решения о прореживании принимаются на основе результатов проведения трудоемкого ручного подсчета случайно выбранных деревьев в саду и последующей экстраполяции полученных оценок на остальную площадь сада. Агрономы подсчитывают количество цветков и завязей во время цветения, приходящихся на одно дерево, и определяют размеры плодов перед сбором [5]. Проводится оценка количества цветков и завязей, которые приходится на поперечное сечение ствола дерева путем измерения окружности ствола на расстоянии 30 см над местом прививки [6]. Прореживание цветков рекомендуется проводить на начальных стадиях цветения, что позволяет распределить питательные вещества между оставшимися цветками и обеспечить более качественное плодоношение [7].

Использование RGB-камер, современных алгоритмов машинного обучения и компьютерного зрения, основанных на глубоких сверточных нейронных сетях (CNN), предоставляет эффективные решения для распознавания биологических объектов с учетом их простран-

ственного положения и формы [8]. Разработка метода, способного повысить точность оценки интенсивности цветения для управления урожайностью сада, позволит более эффективно проводить процесс прореживания.

Цель исследования – трансферное обучение (transfer learning) модели сверточной нейронной сети (Segmentation Convolutional Neural Networks, Seg-CNN) для сегментации цветков яблони на изображениях для оценки интенсивности цветения.

Научная новизна – разработан метод оценки интенсивности цветения яблони для выполнения технологической операции прореживания, включающий подготовку набора данных (dataset) с аугментацией размеченных изображений для расширения обучающего набора, анализ и выбор оптимальных гиперпараметров, обучение моделей сверточной нейронной сети YOLOv8-seg, валидацию, тестирование моделей и оценку результатов распознавания.

Материал и методы. В исследованиях ученых предложены различные способы оценки интенсивности цветения яблони с использованием систем технического зрения с различными типами сенсоров и алгоритмов обработки полученных данных. Большинство способов основано на цветовой пороговой обработке, которые усложняются переменными условиями освещения. Естественная густота цветков яблони на изображениях затрудняет процесс их распознавания по цвету из-за взаимного перекрытия или препятствий в виде листьев [9]. Предложены методы учета цветения, основанные на использовании экрана, расположенного за рядом деревьев на заднем плане [10]. На основе предположения об увеличении фотосинтетической активности плодовых деревьев в период цветения разрабатывались системы обработки спектральной информации. Значительная вариация спектральных данных от различных растений усложняет установление общих паттернов и выделения важных характеристик, требует сложных методов анализа [11, 12]. Применение метода опорных векторов

(Support vector machine, SVM), основанного на морфологических характеристиках растений, ограничено сложностью выбора признаков и интерпретацией результатов, чувствительностью к шуму и необходимостью тщательной настройки параметров. Для более точного распознавания и классификации биологических объектов в садоводстве, а также анализа сложных текстур применялись глубокие сверточные нейронные сети (Deep CNN), такие как ResNet, Inception и VGG, а также рекуррентные нейронные сети (RNN) и рекуррентные сверточные нейронные сети (RCNN) для оценки динамики развития растений [13, 14].

Существующие модели сверточных нейронных сетей (CNN), такие как YOLO (You Only Look Once), Faster R-CNN (Faster Region-based Convolutional Neural Network), SSD (Single Shot MultiBox Detector), Mask R-CNN, R-FCN (Region-based Fully Convolutional Networks), RetinaNet RFCN (Region-based Fully Convolutional Network), DetNet (Detection Network), Deformable Convolutional Networks, NAS-FPN (Neural Architecture Search for Feature Pyramid Networks) и другие позволяют распознавать биологические объекты на изображениях и классифицировать их с использованием прямоугольных ограничивающих рамок (bounding boxes). Анализ проведенных исследований показал, что ограничивающие рамки не всегда позволяют достаточно точно распознать границы биологических объектов на изображениях в кроне деревьев из-за их взаимного перекрытия или препятствий в виде листьев. Ограничивающие рамки захватывают избыточную площадь вокруг биологических объектов, что снижает точность оценки занимаемой ими площади на изображении [15, 16]. Для более эффективного подсчета количества цветков следует использовать методы масочной сегментации, которые позволяют выделить каждый цветок как отдельный объект и провести точный подсчет.

Анализ существующих моделей сверточных нейронных сетей показал, что современная модель YOLOv8-seg (You Only Look Once version 8 segmentation) обладает значительными преимуществами в производительности и в скорости обработки изображений в реальном времени. YOLOv8-seg обеспечивает детальную сегментацию объектов на изображениях различной сложности и масштабов на уровне пикселей, позволяет строить ограничивающие рамки и определять классы объектов

на изображении за один проход (One-stage), что исключает необходимость в дополнительных вычислениях и постобработке результатов [17, 18, 19].

Архитектура модели YOLOv8 состоит из трех основных частей: Backbone, Neck и Head. Часть Backbone отвечает за извлечение признаков из входного изображения, 53 сверточных слоя, использует архитектуру CSPDarknet53 с частичными межэтапными соединениями для улучшения информационного потока между различными слоями. Часть Neck объединяет карты признаков из разных этапов Backbone, чтобы захватывать информацию в разных масштабах. Использован модуль C2f, который сочетает семантические признаки высокого уровня с пространственной информацией низкого уровня, что способствует улучшению точности распознавания объектов. Для обнаружения объектов разных размеров и масштабов на изображении также использована пирамидальная сеть признаков (Feature Pyramid Network). Часть Head состоит из нескольких сверточных слоев, за которыми следует ряд полносвязных слоев. Эти слои отвечают за прогнозирование ограничивающих рамок, оценку объектности (objectness scores) и вероятности классов для объектов, распознанных на изображении (рис. 1).

В исследованиях применена техника трансферного обучения, использованы предварительно обученные модели на наборе данных датасета COCO (Common Objects in Context). Датасет COCO включает более 330 тысяч изображений с разметкой, объектную сегментацию более чем 2,5 миллиона объектов в 80 классах, а также 91 категорию фоновых элементов для сегментации, включая цветы [20]. Техника трансферного обучения позволяет ускорить обучение моделей, используя знания, полученные из предварительного обучения на других наборах данных. Для обучения использованы современные модели YOLOv8-seg (n, s, m, l, x) с различным количеством параметров: «nano» (n) имеет 3,4 млн параметров; «small» (s) – 42,6 млн параметров; «medium» (m) – 110,2 млн параметров; «large» (l) – 220,5 млн параметров; «extra-large» (x) – 344,1 млн параметров. Данные модели обладают различной средней точностью распознавания объектов на наборе данных COCO с использованием ограничивающих рамок (mAP_{box} от 36,7 до 53,4) и масок сегментации (mAP_{mask} от 30,5 до 43,4) (рис. 2).

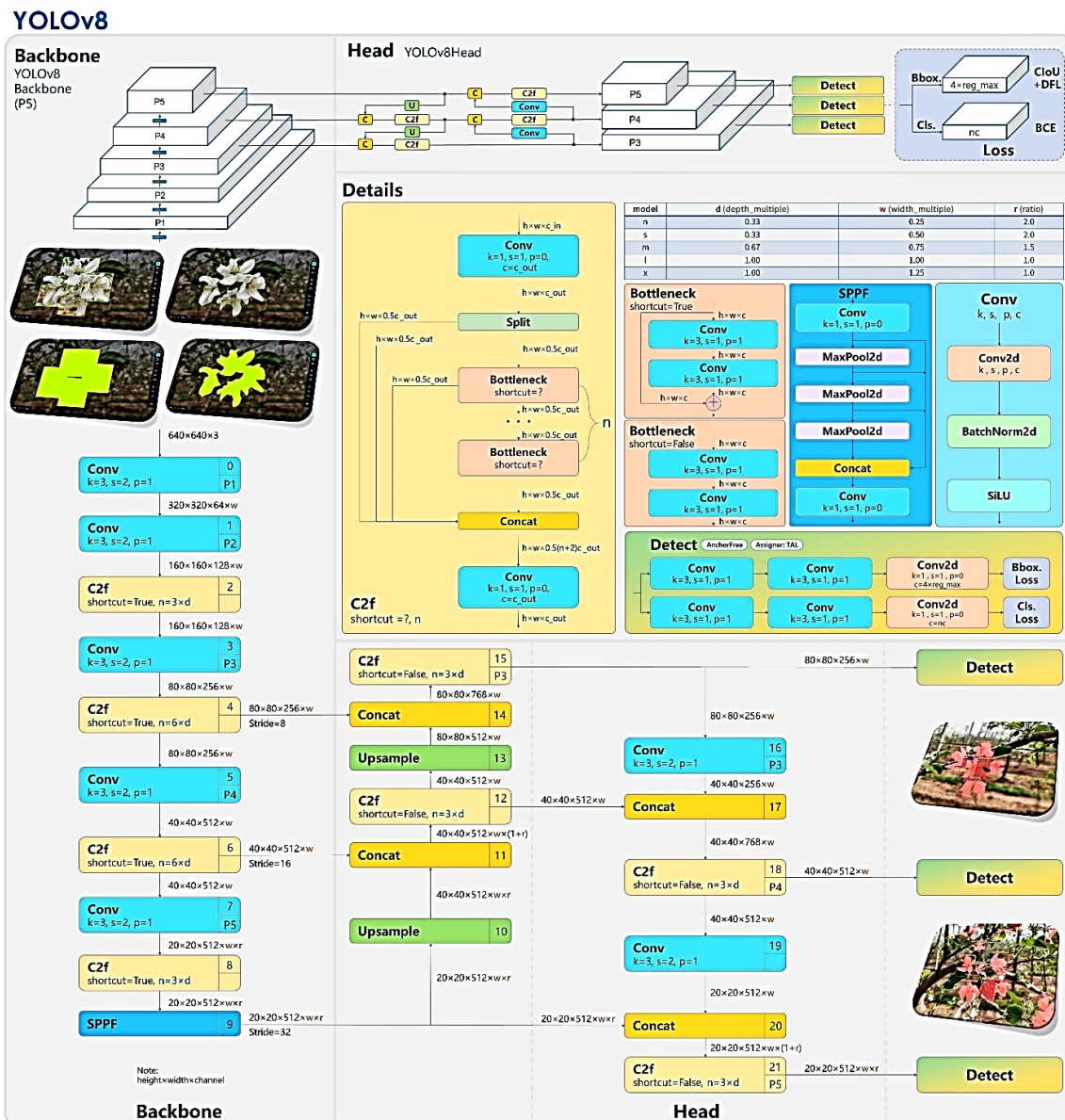


Рис. 1. Архитектура модели сверточной нейронной сети YOLOv8 /
Fig. 1. The architecture of the YOLOv8 convolutional neural network model

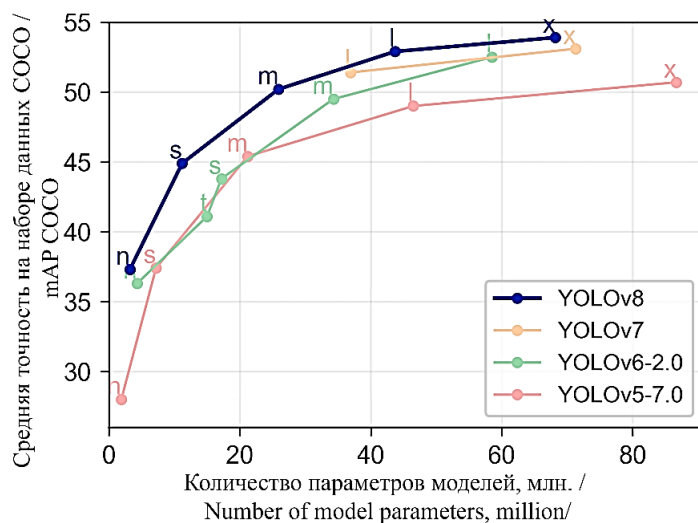


Рис. 2. Средняя точность моделей YOLOv5-YOLOv8 (n, s, m, l, x) на наборе данных COCO /

Fig. 2. The average accuracy of the YOLOv5-YOLOv8 (n, s, m, l, x) models on the COCO dataset

Набор изображений для обучения моделей нейронных сетей был собран с использованием камеры GoPro HERO 11. Матрица камеры обладает разрешением в 27 мегапикселей, разрешение фотографий 5312x2988 пикселей (широкоугольный объектив). Процесс получения изображений включал съемку с трех различных расстояний до цветущих деревьев: менее 0,2 м, от 0,2 м до 0,6 м и более 0,6 м (рис. 3).

Аннотация (разметка) изображений была выполнена на платформе Roboflow с применением инструментов выделения прямоугольных

рамочек (Bounding-box) и аннотации полигонов (Polygon Annotation and Labeling). Выделены контуры цветков яблони, указаны сегменты цветков на изображениях (рис. 4).

Аннотации представлены в форме бинарных масок, где каждый пиксель отмечен значением 1, если он соответствует объекту, и значением 0, если фоновый пиксель. Всем размеченным цветкам яблони был присвоен класс «apple-flowers». Данные, включая координаты и класс, сохранены в формате JSON для последующего использования их при обучении моделей.



Рис. 3. Примеры изображений, полученных с различных расстояний до цветущих деревьев яблони /
Fig. 3. Examples of images captured from various distances to flowering apple trees

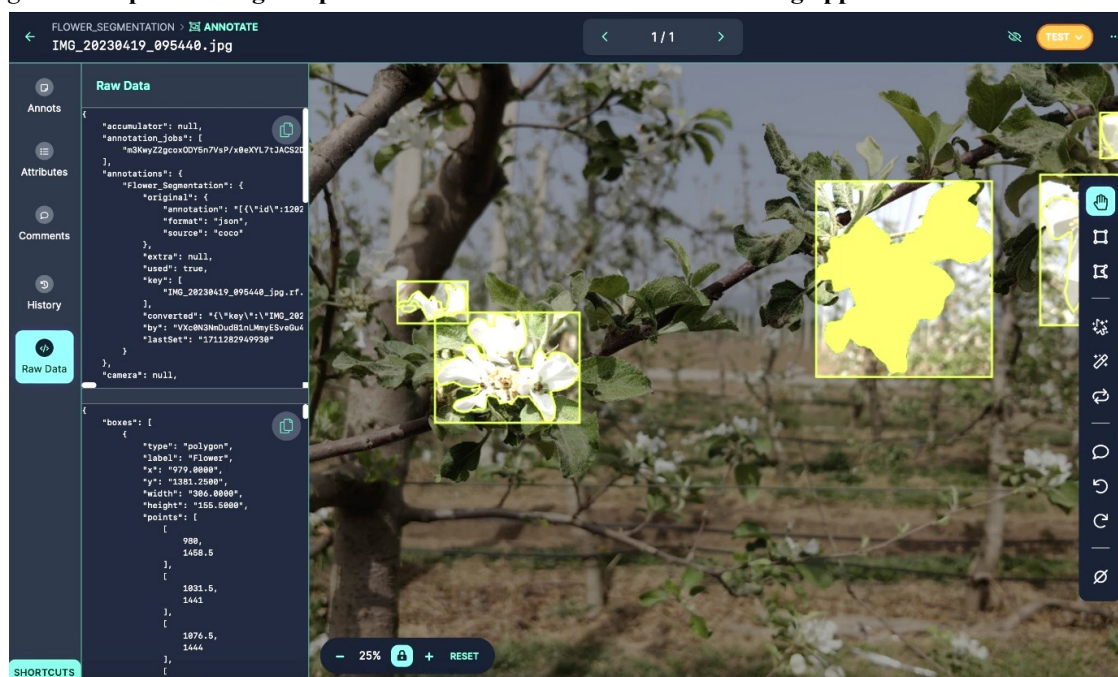


Рис. 4. Разметка изображений с применением инструмента выделения прямоугольных рамок (Bounding-box) и инструмента аннотации полигонов (Polygon Annotation and Labeling) /
Fig. 4. Image annotation using the bounding-box tool and the polygon annotation and labeling tool

Для расширения набора данных и повышения обобщающей способности моделей при обучении была проведена аугментация размеченных изображений, размер которых был уменьшен до 640x640 пикселей в соответствии с требованиями архитектуры моделей YOLO (Resize). Этот процесс включал горизонтальное

отражение (Flip), горизонтальный поворот на 90° (Rotate), поворот от -15° до +15° (Rotation), добавление шумов до 5 % пикселей (Noise), размытие до 2,5 пикселей (Blur), горизонтальные и вертикальные сдвиги от -10° до 10° (Shear), а также изменение оттенка цветов от -15° до +15° (Hue) (рис. 5).

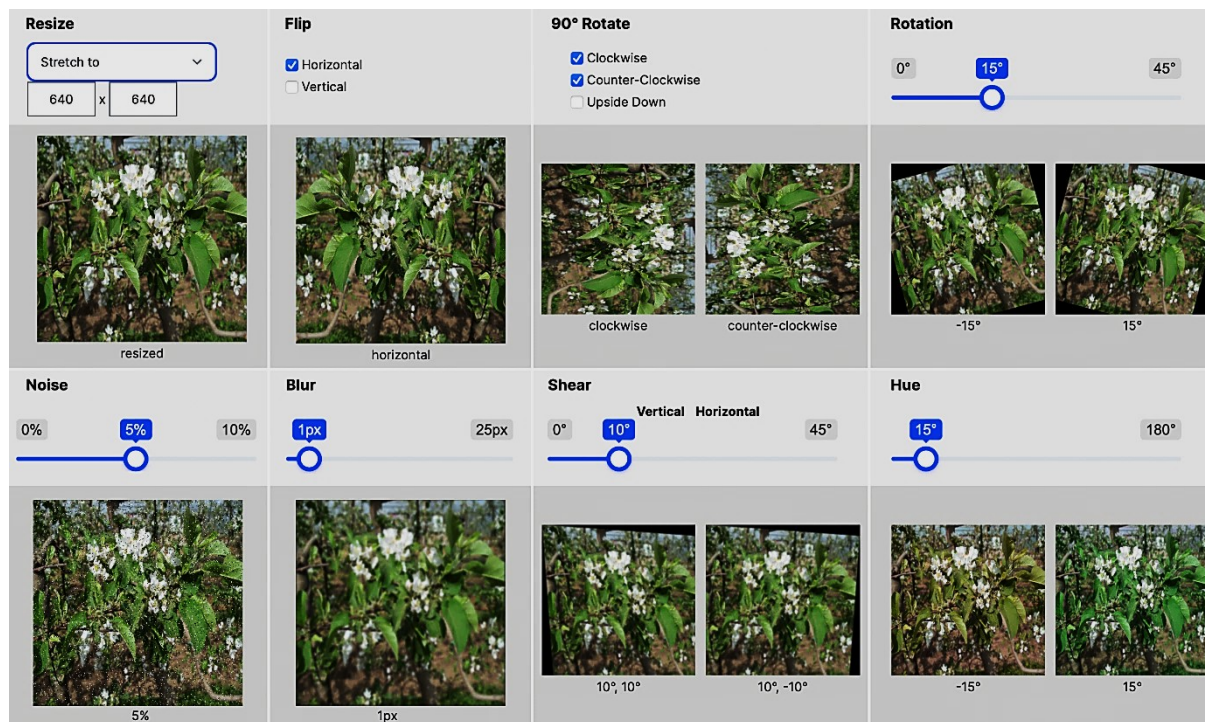


Рис. 5. Аугментация изображений цветков яблони /
Fig. 5. Augmentation of apple blossom images

После получения размеченного набора данных (2500 шт. изображений) проведено их разделение на три независимые подвыборки. Обучающая выборка (70 %, 1750 шт. изображений) использована для обучения моделей. Валидационная выборка (20 %, 500 шт. изображений) использована для настройки гиперпараметров моделей и оценки производительности во время обучения. Тестовая выборка (10 %, 250 шт. изображений) использована для окончательной оценки производительности модели после её обучения (рис. 6).

Такое разделение данных помогает оценить производительность моделей на разных этапах их разработки и настроить гиперпараметры для достижения оптимальной производительности. Построена карта разметки данных (Annotation hitmap), которая показывает плотность разметки объектов на изображении, что помогает анализировать распределение объектов в наборе данных и оценивать качество разметки.

Метрики бинарной классификации, такие как Precision (точность) и Recall (полнота) [21] использованы для оценки качества работы обученных моделей по распознаванию цветков яблони на изображениях при применении ограничивающих рамок и масочной сегментации, рассчитаны по формулам (1) и (2) (расчеты проведены в соответствии с ГОСТ Р 70462.1-2022¹):

$$\text{Precision (точность)}_{\text{Box/Mask}} = \frac{\text{TP}_{\text{Box/Mask}}}{\text{TP}_{\text{Box/Mask}} + \text{FP}_{\text{Box/Mask}}}, \quad (1)$$

$$\text{Recall (полнота)}_{\text{Box/Mask}} = \frac{\text{TP}_{\text{Box/Mask}}}{\text{TP}_{\text{Box/Mask}} + \text{FN}_{\text{Box/Mask}}}, \quad (2)$$

где TP (true positive)_{Box/Mask} – количество верно обнаруженных ограничивающих рамок или пикселей масок объектов;

FP (false positive)_{Box/Mask} – количество ложноположительных ограничивающих рамок или пикселей масок объектов;

FN (false negative)_{Box/Mask} – количество ложноотрицательных ограничивающих рамок или пикселей масок объектов.

¹ГОСТ Р 70462.1-2022. Информационные технологии. Интеллект искусственный. Оценка робастности нейронных сетей. Ч. 1. Обзор. М.: Российский институт стандартизации, 2022. 32 с. URL: <https://files.stroyinf.ru/Data/790/79058.pdf>

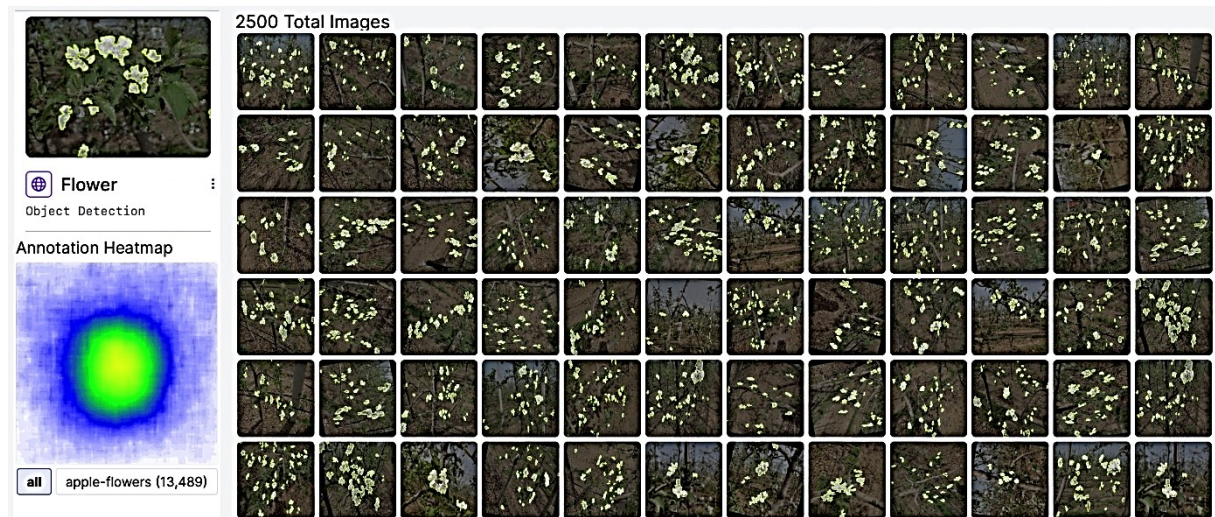


Рис. 6. Датасет размеченных изображений цветков яблони /

Fig. 6. Labeled dataset of apple blossom images

Метрика Mean Average Precision (mAP) использована для расчета средней точности, установлена путем определения площади под кривой Precision-Recall (точность-полнота, формула 3).

$$mAP_{Box/Mask} = \frac{1}{n} \sum_{i=1}^N AP_{Box,i/Mask,i}, \quad (3)$$

где N – количество классов объектов; $AP_{Box,i/Mask,i}$ – площадь под кривой Precision-Recall (точности-полноты) для каждого класса i при использовании ограничивающих рамок ($AP_{Box,i}$) или масок сегментации ($AP_{Mask,i}$).

Метрика mAP учитывает точность и полноту при обнаружении объектов, оценивая обобщенную производительность модели. Более высокое значение mAP свидетельствует о более точном распознавании объектов заданных классов на изображениях.

Метрика IoU (Intersection over Union) вычислена путем деления площади пересечения ограничивающих рамок или масок на площадь их объединения, по формуле:

$$IoU_{Box/Mask} = \frac{A_{\text{Intersection Box/Mask}}}{A_{\text{Union Box/Mask}}}, \quad (4)$$

где $A_{\text{Intersection Box/Mask}}$ – площадь области; образованной пересекающимися сторонами или пикселями; $A_{\text{Union Box/Mask}}$ – сумма площадей предсказанных и истинных областей, найденная по формуле:

$$A_{\text{Union Box/Mask}} = A_{\text{predicted Box/Mask}} + A_{\text{ground truth Box/Mask}} - A_{\text{Intersection Box/Mask}}. \quad (5)$$

Если коэффициент пересечения между предсказанной и истинной областями превышает заранее заданный порог, например 50 %, то распознавание считается успешным.

Функция потерь $Loss_{Box/Mask}$ использована для оценки ошибок модели в определении ограничивающих рамок или масок объектов на изображении в процессе обучения с целью минимизации расхождений между предсказанными и фактическими (размеченными) значениями:

$$Loss_{Box/Mask} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} \cdot \text{MSE}(x_i, \hat{x}_i, y_i, \hat{y}_i, w_i, \hat{w}_i, h_i, \hat{h}_i / M_i, \hat{M}_i), \quad (6)$$

где λ_{coord} – коэффициент для балансировки компонентов функции потерь; S – размер сетки (grid size); B – количество предсказанных ограничивающих рамок на каждой ячейке сетки; i – индекс ячейки сетки; j – индекс предсказанной ограничивающей рамки в ячейке сетки; 1_{ij}^{obj} – индикаторная функция принимает значение, равное 1, если ограничивающая рамка j в ячейке сетки i отвечает за объект, MSE – среднеквадратичная ошибка (Mean Squared Error) между предсказанными и истинными значениями координат ($x_i, \hat{x}_i, y_i, \hat{y}_i, w_i, \hat{w}_i, h_i, \hat{h}_i$) или сегментационных масок ($M_i, \hat{M}_i$).

Скорость обработки кадров (Frame Rate, FR) моделями сверточных нейронных сетей определяли по формуле:

$$FR = \frac{1}{T_{\text{cp}}}, \quad \text{где} \quad (7)$$

T_{cp} – среднее время обработки одного кадра, мс.

Для обучения сверточных нейронных сетей и анализа данных использовали различные программные инструменты и библиотеки. Фреймворки CUDA 12.2, TensorFlow 2.15.0, TensorBoard 2.15.0 и Keras 2.15.0 были использованы для реализации и обучения моделей нейронных сетей. PyArrow 10.0.1 применяли для обработки и анализа больших объемов

данных. Библиотеки JAX 0.4.23 и JAXLib 0.4.23 обеспечивали возможность численных вычислений с автоматическим дифференцированием и ускорением на GPU. Библиотека NumPy использовалась для работы с многомерными массивами и математическими функциями, такими как вычисление площади, библиотека OpenCV – для обработки изображений, преобразования цветовых пространств, поиска контуров, библиотека matplotlib – для визуализации данных в виде графиков и диаграмм, библиотека tqdm – для отображения прогресса выполнения итераций обучения.

Обучение моделей нейронных сетей проводили на компьютере с процессором Intel Core i9-10900X с тактовой частотой 3.70 ГГц, видеокартой GeForce RTX 2080TI с 11 ГБ памяти (2 шт.), материнской платой GIGABYTE X299 UD4 Pro, SSD-накопителем Intel PCI-E на 1 ТБ и модулем памяти Kingston DDR4 DIMM емкостью 32 ГБ (4 шт.).



Каждый распознанный цветок яблони на изображении был выделен путем создания ограничивающей рамки (box), после чего проведена классификация и сегментация. Каждый пиксель цветка яблони сегментирован, отмечен определенным цветом, что позволяет визуализировать

Результаты и их обсуждение. В результате экспериментальных исследований были определены оптимальные параметры для обучения моделей нейронных сетей YOLOv8n-seg – YOLOv8x-seg. Размер изображений, подаваемых на вход, составил 640x640 пикселей, количество эпох обучения установлено равным 100. Параметр терпимости (Patience), который определяет количество эпох, в течение которых обучение модели продолжается без изменений в производительности на валидационном наборе данных, составил 10 эпох. Размер пакета данных (Batch Size), определяющий количество изображений, обрабатываемых параллельно на каждой итерации обновления весов модели, был также установлен на уровне 10. Примеры распознавания, классификации и сегментации цветков яблони на изображениях тестового набора данных модели YOLOv8x-seg представлены на рисунке 7.

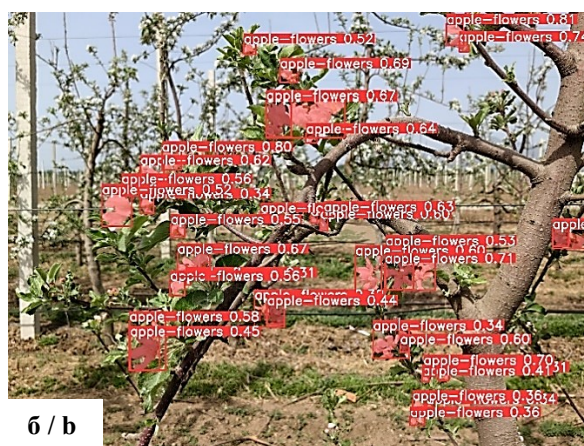


Рис. 7. Результаты распознавания, классификации и сегментации цветков яблони на изображениях при различных расстояниях до кроны: а – менее 0,2 м, б – от 0,2 м до 0,6 м, в – более 0,6 м /

Fig. 7. The results of apple blossom recognition, classification and segmentation on images at various distances from the canopy: a – less than 0.2 m, b – from 0.2 m to 0.6 m, c – more than 0.6 m

результаты работы модели. Полученные графики mAP50(Box)-Epoch и mAP50(Mask)-Epoch показывают, как изменяется средняя точность (mAP) при 50 %-ном пороговом значении IoU (Intersection over Union) в зависимости от эпохи обучения (рис. 8).

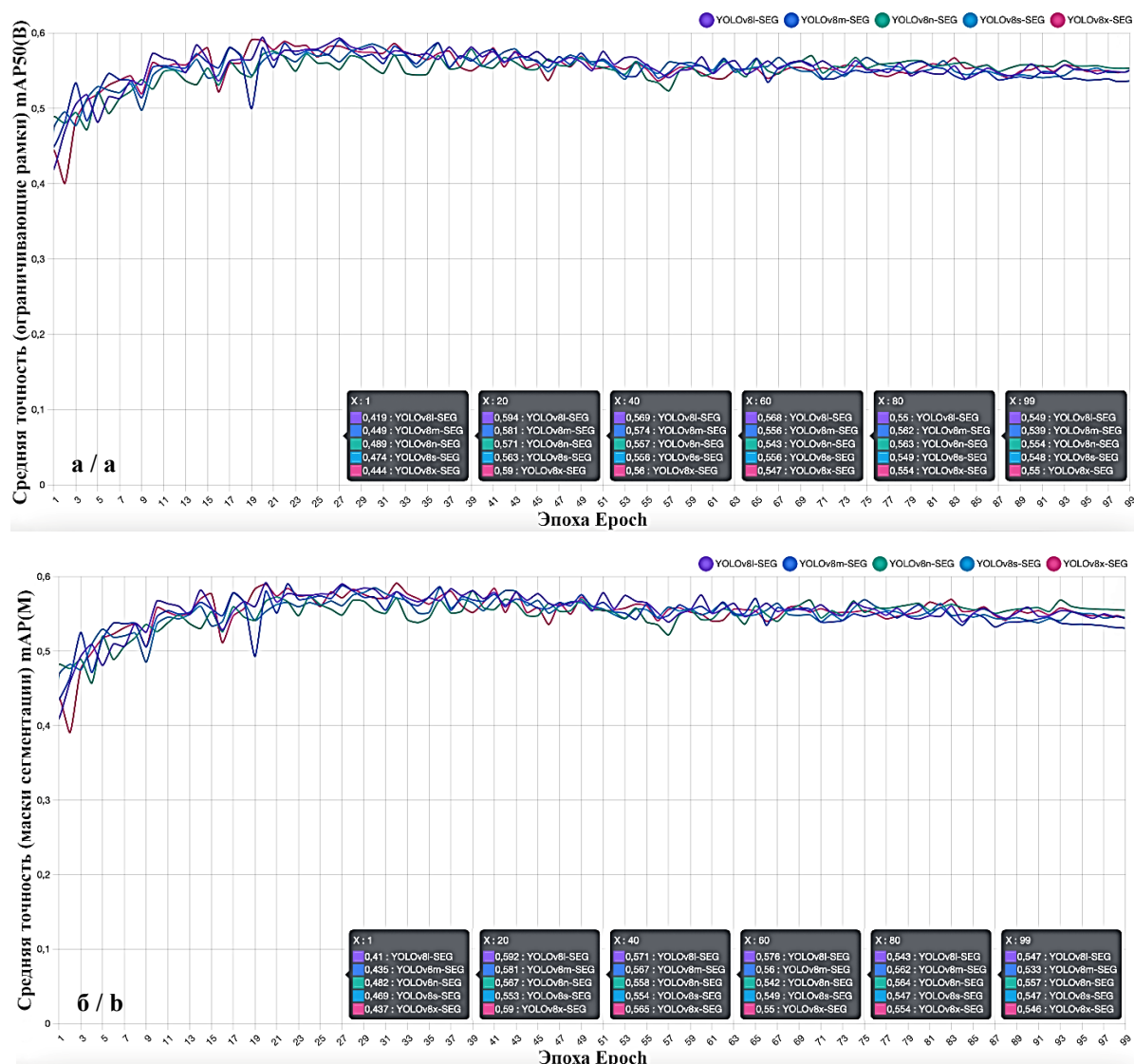


Рис. 8. Кривые оценки метрики средней точности в зависимости от эпохи обучения: а – средняя точность при распознавании ограничивающих рамок (mAP50 bounding boxes) цветков яблони; б – средняя точность при распознавании сегментирующих масок (mAP50 segmentation masks) цветков яблони /

Fig. 8. Curves for estimating the average accuracy metric depending on the learning epoch: a – average accuracy in recognizing mAP50 bounding boxes apple tree blossoms; b – average accuracy in recognizing mAP50 segmentation masks apple tree blossoms

Для каждой итерации обучения провели оценку точности распознавания с применением ограничивающих рамок и сегментационных масок. После анализа полученных результатов определили оптимальную эпоху обучения (36), при которой модели достигали наилучшей производительности.

В процессе обучения моделей были построены кривые потерь для ограничивающих рамок и сегментационных масок, отражающие динамику изменений функции потерь в течение каждой итерации обучения (рис. 9).

Полученные кривые Loss(Box/Mask) – Epoch позволяют определить целесообразность дальнейшего обучения моделей, отражают способность моделей снижать ошибки

в распознавании объектов и помогают предотвратить переобучение (overfitting), при котором модель подстраивается под обучающие данные, теряя способность обобщать знания на новые данные.

Установлено, что функция потерь достигает минимума на 34-й эпохе обучения и остается практически неизменной с 34 по 47 эпохи. Начиная с 47-й эпохи, наблюдается увеличение потерь. Оптимальное количество эпох обучения было определено с использованием стратегии ранней остановки обучения (early stopping), которая прекращает процесс обучения, когда производительность модели на валидационной выборке перестает улучшаться в течение заданного числа итераций.

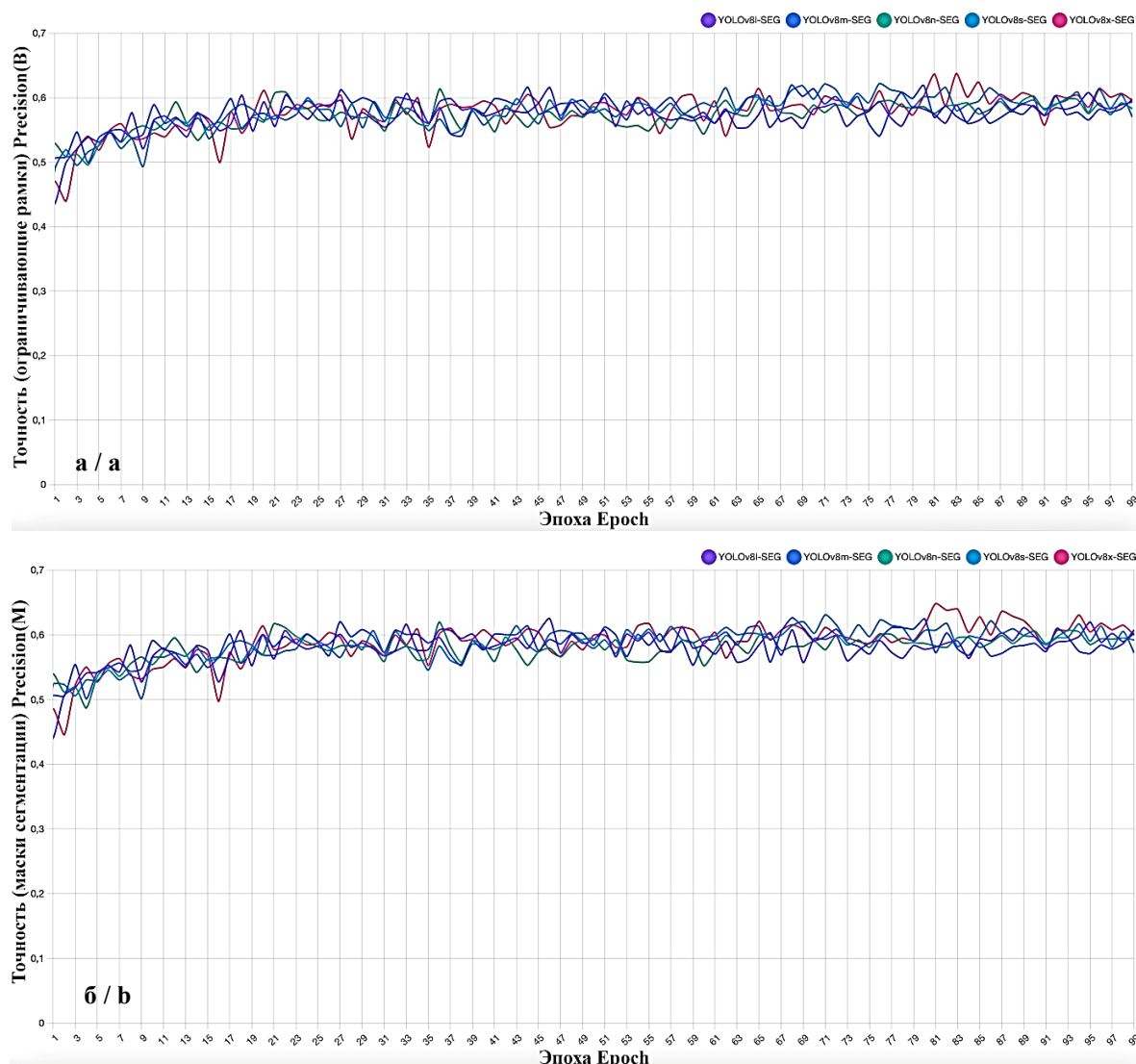


Рис. 9. Кривые изменения функции потерь в зависимости от эпохи обучения: а – потери при распознавании ограничивающих рамок (Box Loss) цветков яблони; б – потери при распознавании сегментирующих масок (Seg Loss) цветков яблони /

Fig. 9. Curves of loss function changes depending on the training epochs: а – losses in bounding box recognition (Box Loss) apple tree blossoms; б – losses in segmentation mask recognition (Seg Loss) apple tree blossoms

Полученные в процессе обучения данные о точности (Precision) и полноте (Recall) моделей использовали для вычисления средней точности (mAP50) как при использовании ограничивающих рамок (Box), так и сегментирующих масок (Mask) (табл.).

Установлено, что модели YOLOv8x-seg (mAP50 0,591) и YOLOv8l-seg (mAP50 0,584) показывают более высокую производительность по всем метрикам, чем модель YOLOv8n-seg (mAP50 0,577), что связано с количеством используемых параметров данных моделей. Большинство моделей показывают незначительное снижение точности распознавания масок сегментации (Mask) по сравнению с точностью распознавания ограничивающих рамок (Box), что объясняется более сложным

процессом получения маски для каждого пикселя изображения. Полученные результаты подтверждают эффективность предложенного метода и его способность к точному распознаванию цветков яблони. Сравнение результатов эффективности работы моделей YOLOv8-seg (n, s, m, l, x) на наборе данных COCO (рис. 1) с результатами, полученными в данном исследовании, показывает превосходство моделей, обученных с использованием метода трансферного обучения в средних показателях точности при распознавании цветков яблони.

Модель YOLOv8n-seg показала наиболее высокую среднюю скорость обработки кадров (57,32 кадра/с), что является основным показателем при работе на менее мощных вычислительных системах. Средняя скорость обра-

ботки кадров модели YOLOv8s-seg составила 32,56 кадра/с, YOLOv8m-seg 18,92 кадра/с, YOLOv8l-seg 15,6 кадра/с и YOLOv8x-seg 10,27 кадра/с.

Таблица – Результаты вычисления метрик для бинарной классификации при использовании ограничивающих рамок и масок сегментации /

Table – The results of calculating metrics for binary classification using bounding boxes and segmentation masks

Метод распознавания / Recognition method	Метрика точности / Precision Metric	Метрика полноты / Recall metric	Метрика средней точности / mAP50 metric
Модель YOLOv8n-seg / Model YOLOv8n-seg			
Ограничивающие рамки (Box) / Bounding Boxes (Box)	0,583	0,576	0,579
Маски сегментации (Mask) / Segmentation Masks (Mask)	0,572	0,578	0,577
Модель YOLOv8s-seg / Model YOLOv8s-seg			
Ограничивающие рамки (Precision Box) / Bounding Boxes (Precision Box)	0,57	0,579	0,580
Маски сегментации (Mask) / Segmentation Masks (Mask)	0,607	0,566	0,579
Модель YOLOv8m-seg / Model YOLOv8m-seg			
Ограничивающие рамки (Box) / Bounding Boxes (Box)	0,597	0,564	0,582
Маски сегментации (Mask) / Segmentation Masks (Mask)	0,592	0,582	0,581
Модель YOLOv8l-seg / Model YOLOv8l-seg			
Ограничивающие рамки (Box) / Bounding Boxes (Box)	0,596	0,578	0,586
Маски сегментации (Mask) / Segmentation Masks (Mask)	0,599	0,585	0,584
Модель YOLOv8x-seg / Model YOLOv8x-seg			
Ограничивающие рамки (Box) / Bounding Boxes (Box)	0,593	0,577	0,594
Маски сегментации (Mask) / Segmentation Masks (Mask)	0,602	0,575	0,591

Для сравнения точности распознавания цветков яблони моделью YOLOv8x-seg при использовании ограничивающих рамок и масок сегментации был проведен расчет суммарной площади, занимаемой распознанными цветками яблони на изображениях тестовой выборки. Разработан скрипт Python в среде PyCharm CE с использованием библиотек OpenCV (cv2) и NumPy (np), который позволяет автоматизировать анализ изображений (рис. 10).

Скрипт преобразует изображения в цветовую модель HSV (Hue, Saturation, Value). Используя заданные диапазоны цветов, скрипт позволяет определить пиксели, занятые цветками, находит их контуры, рассчитывает площади и добавляет найденные параметры на маску, после чего маска применяется к исходному изображению для визуализации и анализа результатов. В скрипт добавлено исключение площади текстовых меток объектов (текст с наименованием классов и пороговых значений) и линий из общей площади ограничивающих

рамок. В результате проведенных исследований определено, что среднее количество пикселей, распознанных с помощью модели YOLOv8x-seg в ограничивающих рамках тестового набора данных, превышает среднее количество пикселей сегментов цветков яблони на 37,2 %. Маски сегментации позволяют определить класс, к которому принадлежит каждый пиксель изображения, более детально учитывая форму и точно определяя границы цветков яблони, в то время как ограничивающие рамки определяют только прямоугольную область вокруг цветков яблони с избыточной площадью.

Предлагаемый метод сегментации, основанный на алгоритме машинного обучения, обеспечивает возможность морфологического анализа изображений и выявления маркерных признаков объектов с высоким пространственным разрешением. Средняя абсолютная ошибка моделей YOLOv8-seg (n, s, m, l, x) при подсчете количества цветков яблони на изоб-

ражениях тестовой выборки не превышает 9 %. Оценка интенсивности цветения с помощью алгоритмов машинного обучения основывается на следующих критериях: площадь цветков, измеряемая с помощью сегментации на изображениях; количество цветков на дереве, опре-

деляемое путём подсчёта сегментированных объектов; плотность цветков, оцениваемая как доля цветков на единицу площади кроны дерева; динамика цветения, определяемая на основе данных, собранных в разные временные периоды.

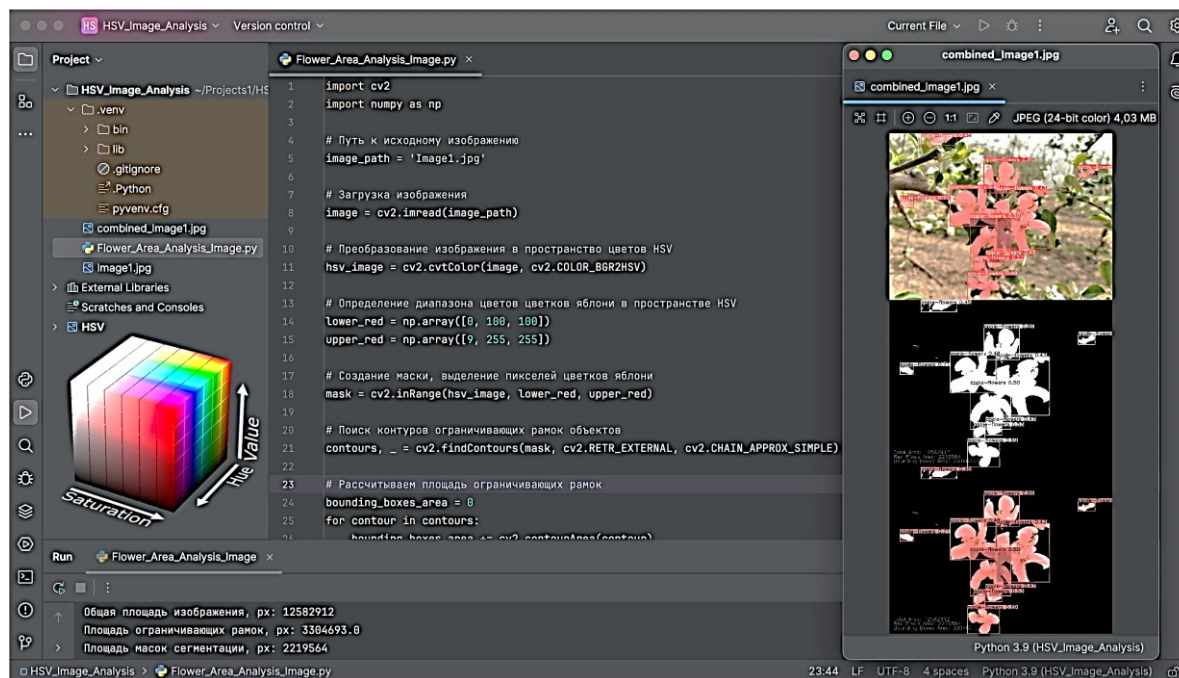


Рис. 10. Результаты сравнения точности распознавания суммарной площади цветков яблони с помощью ограничивающих рамок и масок сегментации, скрипт Python в среде PyCharm CE /

Fig. 10. The results of comparing the accuracy of recognizing the total area of apple blossom flowers using bounding boxes and segmentation masks were obtained using a Python script in the PyCharm CE environment

Результаты исследований могут быть использованы для разработки роботизированных систем, способных к автономной оценке интенсивности цветения в промышленных садах. Для этого роботизированные платформы оснащаются RGB-камерами, которые с помощью компьютерного зрения и машинного обучения сегментируют и классифицируют цветки на деревьях [22]. Дальнейшее развитие метода может включать в себя работу с более разнообразными наборами данных, а также улучшение производительности моделей. Методы фотограмметрии в сочетании с координатными данными о расположении деревьев могут быть использованы для определения принадлежности цветков конкретным деревьям для массового фенотипирования растений в режиме реального времени. Съёмка с нескольких ракурсов позволит строить 3D-модели деревьев и привязывать к ним распознанные цветки, создавая детализированные модели сада [23]. В дальнейших исследованиях для определения общего количества цветков необходимо использовать алгоритмы трекинга (Deep SORT, SORT, BYTETrack, FairMOT и др.), которые учиты-

вают частичное перекрытие цветков на изображениях с разных ракурсов. Инкрементальное расширение датасета, нормализация и автоматическая коррекция цвета изображений позволят алгоритмам адаптироваться к различным условиям освещения, обеспечивая более точное распознавание, классификацию и сегментацию.

Заключение. В результате проведенных исследований определены гиперпараметры алгоритма машинного обучения модели YOLOv8-seg для распознавания, классификации и сегментации изображений цветков яблони. Установлено, что модели YOLOv8x-seg (метрика mAP50 = 0,591) и YOLOv8l-seg (метрика mAP50 = 0,584) показали более высокую производительность при распознавании цветков яблони. Скорость обработки кадров (Frame Rate, FR) моделями сверточных нейронных сетей составила от 10,27 (модель YOLOv8x-seg) до 57,32 кадра/с (модель YOLOv8n-seg). Средняя абсолютная ошибка моделей при распознавании цветков яблони и подсчете их количества на тестовой выборке не превышает 9 %.

References

1. Pflanz M., Gebbers R., Zude M. Influence of tree-adapted flower thinning on apple yield and fruit quality considering cultivars with different predisposition in fructification. *Acta Hortic.* 2016;1130:605–611. DOI: <https://doi.org/10.17660/ActaHortic.2016.1130.90>
2. Link H. Significance of flower and fruit thinning on fruit quality. *Plant growth regulation.* 2000;31(1):17–26. DOI: <https://doi.org/10.1023/A:1006334110068>
3. Bhattarai U., Bhushal S., Majeed Y., Karkee M. Automatic Blossom Detection in Apple Trees Using Deep Learning. *IFAC-PapersOnLine* 2020;53(2):15810–15815. DOI: <https://doi.org/10.1016/j.ifacol.2020.12.216>
4. Lei X., Yuan Q., Xyu T., Qi Y., Zeng J., Huang K., et al. Technologies and Equipment of Mechanized Blossom Thinning in Orchards: A Review. *Agronomy.* 2023;13(11):2753. DOI: <https://doi.org/10.3390/agronomy13112753>
5. Robinson T. L., Gonzalez L., Cheng L., Ziang Y., Peck G., Arnoldussen B., et al. Studies in precision crop load management of apple. *Acta Hortic.* 2023;1366:219–226. DOI: <https://doi.org/10.17660/ActaHortic.2023.1366.25>
6. Ahmed D., Sapkota R., Churuvija M., Karkee M. Machine Vision-Based Crop-Load Estimation Using YOLOv8. 2023. pp. 1–22 DOI: <https://doi.org/10.48550/arXiv.2304.13282>
7. Anthony B., Serra S., Musacchi S. Optimizing Crop Load for New Apple Cultivar: “WA38”. *Agronomy.* 2019;9(2):107. DOI: <https://doi.org/10.3390/agronomy9020107>
8. Sun K., Wang X., Liu S., Liu C. Apple, Peach, and Pear Flower Detection Using Semantic Segmentation Network and Shape Constraint Level Set. *Computers and Electronics in Agriculture.* 2021;185:106150. DOI: <https://doi.org/10.1016/j.compag.2021.106150>
9. Chen Z., Su R., Wang Y., Chen G., Wang Z., Yin P., Wang J. Automatic Estimation of Apple Orchard Blooming Levels Using the Improved YOLOv5. *Agronomy.* 2022;12(10):2483. DOI: <https://doi.org/10.3390/agronomy12102483>
10. Aggelopoulou A. D., Bochtis D., Fountas S., Swain K. C., Gemtos T. A., Nanos G. D. Yield prediction in apple orchards based on image processing. *Precision Agriculture.* 2011;12(3):448–456. DOI: <https://doi.org/10.1007/s11119-010-9187-0>
11. Horton R., Cano E., Bulanon D., Fallahi E. Peach Flower Monitoring Using Aerial Multispectral Imaging. *Journal of Imaging.* 2017;3(1):2. DOI: <https://doi.org/10.3390/jimaging3010002>
12. Shurygin B., Smirnov I., Chilikin A., Khort D., Kutyrev A., Zhukovskaya S., Solovchenko A. Mutual Augmentation of Spectral Sensing and Machine Learning for Non-Invasive Detection of Apple Fruit Damages. *Horticulturae.* 2022;8(12):1111. DOI: <https://doi.org/10.3390/horticulturae8121111>
13. Wang X. A., Tang J., Whitty M. DeepPhenology: Estimation of Apple Flower Phenology Distributions Based on Deep Learning. *Computers and Electronics in Agriculture.* 2021;185:106123. DOI: <https://doi.org/10.1016/j.compag.2021.106123>
14. Zhou X., Sun G., Xu N., Zhang X., Cai J., Yuan Y., Huang Y. A Method of Modern Standardized Apple Orchard Flowering Monitoring Based on S-YOLO. *Agriculture.* 2023;13(2):380. DOI: <https://doi.org/10.3390/agriculture13020380>
15. Yuan W., Hua W., Heinemann P. H., He L. UAV Photogrammetry-Based Apple Orchard Blossom Density Estimation and Mapping. *Horticulturae.* 2023;9(2):266. DOI: <https://doi.org/10.3390/horticulturae9020266>
16. Farjon G., Krikeb O., Hillel A. B., Alchanatis V. Detection and counting of flowers on apple trees for better chemical thinning decisions. *Precision Agriculture.* 2020;21(3):503–521. DOI: <https://doi.org/10.1007/s11119-019-09679-1>
17. Terven J., Córdova-Esparza D.-M., Romero-González J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction.* 2023;5(4):1680–1716. DOI: <https://doi.org/10.3390/make5040083>
18. Wang J., Gao Z., Zhang Y., Zhou J., Wu J., Li P. Real-Time Detection and Location of Potted Flowers Based on a ZED Camera and a YOLO V4-Tiny Deep Learning Algorithm. *Horticulturae.* 2022;8(1):21. DOI: <https://doi.org/10.3390/horticulturae8010021>
19. Wu D. H., Lv S. C., Jiang M., Song H. B. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Computers and Electronics in Agriculture.* 2020;178:105742. DOI: <https://doi.org/10.1016/j.compag.2020.105742>
20. Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., et al. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 2014. pp. 740–755. DOI: <https://doi.org/10.48550/arXiv.1405.0312>
21. Maxwell A. E., Warner T. A., Guillén L. A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies - Part I: Literature Review. *Remote Sensing.* 2021;13(13):2450. DOI: <https://doi.org/10.3390/rs13132450>
22. Хорт Д. О., Кутырёв А. И., Смирнов И. Г., Моисеев Г. В., Соловьев В. И. Управление движением сельскохозяйственной автономной роботизированной платформы. *Сельскохозяйственные машины и технологии.* 2023;17(1):25–34. DOI: <https://doi.org/10.22314/2073-7599-2023-17-1-25-34> EDN: UXTFRN
23. Khort D. O., Kutyrev A. I., Smirnov I. G., Moiseev G. V., Soloviev V. I. Agricultural Autonomous Robotic Platform Motion Control. *Sel'skokhozyaystvennyye mashiny i tekhnologii = Agricultural Machinery and Technologies.* 2023;17(1):25–34. (In Russ.). DOI: <https://doi.org/10.22314/2073-7599-2023-17-1-25-34>
24. Хорт Д. О., Кутырёв А. И., Смирнов И. Г., Воронков И. В. Разработка системы автоматизированного управления агротехнологиями в садоводстве. *Сельскохозяйственные машины и технологии.* 2021;15(2):61–68. DOI: <https://doi.org/10.22314/2073-7599-2021-15-2-61-68> EDN: AQNSNT
25. Khort D. O., Kutyrev A. I., Smirnov I. G., Voronkov I. V. Development of an Automated Management System for Agricultural Technologies in Horticulture. *Sel'skokhozyaystvennyye mashiny i tekhnologii = Agricultural Machinery and Technologies.* 2021;15(2):61–68. (In Russ.). DOI: <https://doi.org/10.22314/2073-7599-2021-15-2-61-68>

Сведения об авторах

✉ Кутырёв Алексей Игоревич, кандидат техн. наук, заведующий лабораторией интеллектуальных цифровых систем мониторинга, диагностики и управления процессами в сельскохозяйственном производстве, ведущий научный сотрудник, ФГБНУ «Федеральный научный агроинженерный центр ВИМ», 1-й Институтский проезд, д.5, г. Москва, Российская Федерация, e-mail: vim@vim.ru, ORCID: <https://orcid.org/0000-0001-7643-775X>, e-mail: alexeykutyrev@gmail.com

Information about the authors

✉ Alexey I. Kutyrev, PhD in Engineering, Head of the Laboratory of intelligent digital systems for monitoring, diagnostics and process management in agricultural production, leading researcher, Federal Scientific Agroengineering Center VIM, 1st Institute passage, 5, Moscow, Russian Federation, e-mail: vim@vim.ru, ORCID: <https://orcid.org/0000-0001-7643-775X>, e-mail: alexeykutyrev@gmail.com

✉ – Для контактов / Corresponding author