**К вопросу сравнения эффективности нейронных сетей YOLOv8 и Mask R-CNN**

**в задачах сегментации виноградников**

*Целью данного исследования является сравнение эффективности двух передовых моделей машинного обучения, YOLOv8 и Mask R-CNN, в контексте сегментации виноградников на изображениях. Для достижения этой цели были решены следующие задачи: адаптация иерархии датасетов, обучение моделей, оценка метрик Precision, Recall и других метрик качества на валидационных изображениях, а также анализ и сравнение полученных результатов.*

***Ключевые слова:*** *Mask R-CNN, YOLOv8, искусственный интеллект, компьютерное зрение, обнаружение изображений, сегментация изображений, информационные технологии, precision, recall.*

**Введение**

В современном сельском хозяйстве использование передовых технологий играет ключевую роль в повышении эффективности производства и оптимизации использования земельных ресурсов. Одной из таких инноваций является применение нейронных сетей для анализа земельных участков с целью определения и оптимизации посадки виноградников.

YOLOv8 представляет собой последнее поколение моделей YOLO (You Only Look Once), известных своей высокой скоростью работы и способностью обнаруживать объекты в реальном времени. Mask R-CNN, в свою очередь, является основанным на сверточных нейронных сетях методом, который зарекомендовал себя в задачах сегментации объектов с высокой точностью и детализацией.

Сегментация виноградников имеет свои особенности, связанные с необходимостью точного выделения контуров растений и оценки их площади. Поэтому выбор наилучшей модели для этой задачи играет ключевую роль в точности и надежности полученных результатов.

В данной статье представлен анализ эффективности YOLOv8 и Mask R-CNN в задаче сегментации виноградников, с фокусом на их способности точно выделять и классифицировать виноградники на изображениях с различным разрешением и условиями освещения.

1. **Формулировка цели исследования**

Цель проводимого исследования – это определение лучшей из моделей, основанных на архитектурах Mask R-CNN и YOLOv8, по метрикам Precision, Recall и mAP.

Для достижения данной цели были решены следующие задачи:

1. Адаптация иерархии датасетов под требования представленных моделей;
2. Обучение выбранных моделей машинного обучения на одном исходном наборе данных;
3. Получение метрик Precision, Recall и mAP на валидационных изображениях
4. Выявление модели машинного обучения, которая будет иметь лучшие показатели значения полученных метрик.
5. **Методы решения задачи исследования**

Mask R-CNN — это мощная сверточная нейронная сеть, разработанная для задач сегментации и детекции объектов на изображениях. Она использовалась в данном исследовании для сегментации и детекции объектов. В задачах компьютерного зрения Mask R-CNN является зарекомендовавшей себя архитектурой, применение которой даёт представление о целесообразности применения нейронных сетей в поставленной задаче. Данная сеть в своей основе имеет сверточную нейронную сеть, архитектура которой была изменена и расширена для работы с высокоточным определением объектов и их сегментацией.

Основной особенностью Mask R-CNN является разделение всей архитектуры на несколько основных частей: сеть-основа (backbone), сеть предложений регионов (RPN), голову детектора объектов и голову сегментации. Сеть-основа представляет собой мощную сверточную архитектуру, такую как ResNet, которая извлекает признаки из изображения. В процессе обработки изображение проходит через несколько слоев сверток и преобразований, что позволяет сети выделить ключевые признаки, необходимые для дальнейшей детекции и сегментации объектов.

Сеть предложений регионов (RPN) отвечает за генерацию предложений регионов, которые содержат объекты. Это промежуточный этап, который помогает сети сосредоточиться на наиболее значимых частях изображения. Данная сеть производит множество прямоугольных регионов, среди которых затем выбираются те, которые, с наибольшей вероятностью, содержат объекты.

Следующим этапом является применение RoIAlign, который используется для точного выравнивания предложений регионов на карте признаков. Этот метод избегает проблем, связанных с интерполяцией при использовании RoIPool, обеспечивая точные и стабильные маски объектов.

Голова детектора объектов предсказывает классы объектов и точные координаты bounding box для каждого обнаруженного объекта на изображении. Это позволяет не только определить, где находится объект, но и к какому классу он принадлежит.

Основная особенность Mask R-CNN заключается в наличии дополнительной головы для предсказания масок сегментации объектов. Эта часть архитектуры позволяет точно выделить контуры объектов на изображении, что особенно важно для задач с высокой детализацией.

В данном исследовании была использована классическая архитектура Mask R-CNN. Так как Mask R-CNN была изначально создана для решения задачи сегментации, был доработан вывод обработанных изображений. Модификация заключается в переборе сегментированных объектов на изображении и создании детекционной рамки, размер которой соответствует крайним точкам сегмента объекта.

Рис. 1. Архитектура Mask R-CNN

YOLOv8 — это новейшая современная модель YOLO, которая находит применения в задачах обнаружения объектов, классификации изображений и сегментации объектов. YOLOv8 является улучшенной версией модели YOLOv5, которая включает в себя многочисленные архитектурные изменения и улучшения по сравнению с предыдущей моделью [3]. YOLOv8 достигает высокой точности при работе с COCO (крупномасштабный набор данных для обнаружения, сегментации и субтитров) [4].

YOLOv8 является новейшим представителем семейства моделей YOLO. Особенность работы данных моделей заключается в разовом проходе изображения через операции, заложенные в устройство моделей, что позволяет получать результат за меньшее время, чем у классических CNN-моделей. Архитектура YOLOv8 разделена на 2 части: backbone и head.

Backbone — это часть нейронной сети, которая обычно состоит из серии сверточных слоев, предназначенных для идентификации и извлечения визуальных элементов, таких как края, формы и текстуры из изображений. Затем эти функции передаются на последующие уровни для дальнейшей обработки и анализа.

Head (predicted head) — это часть архитектуры, состоящая из блоков сегментации и детектирования объектов на карте признаков, созданной с помощью backbone.

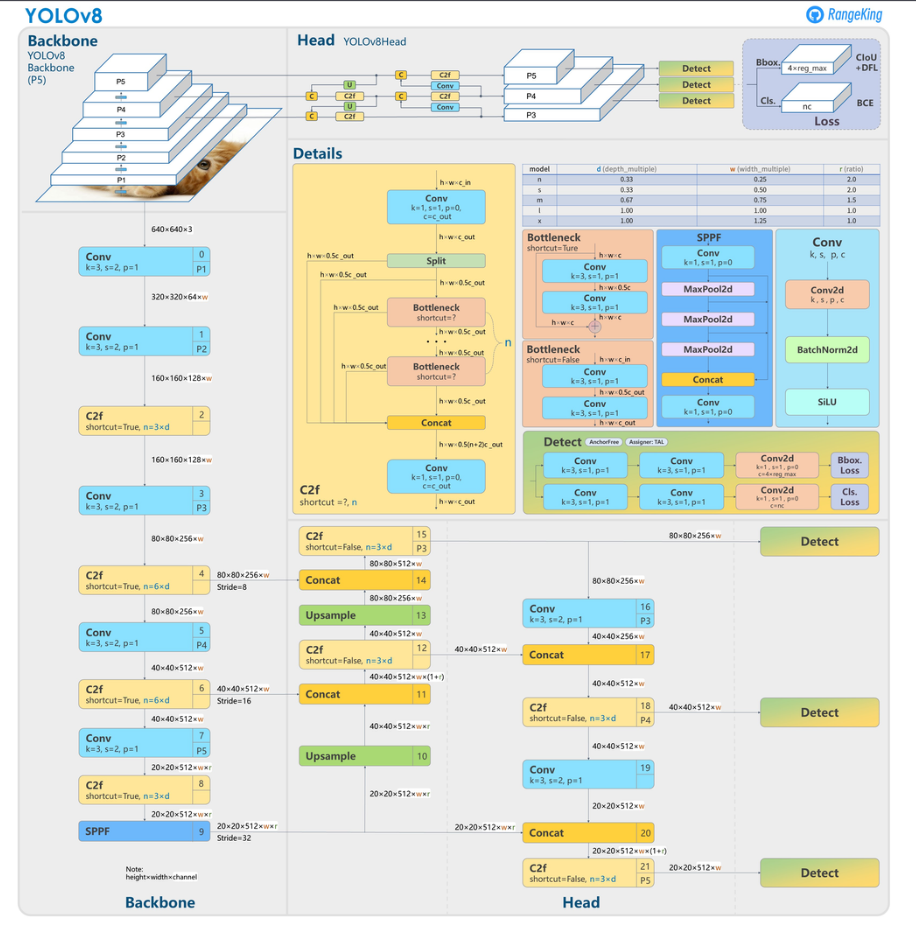


Рис. 2. Архитектура YOLOv8

1. **Результаты работы**