

УДК 004.932.72:004.89

 10.25209/2079-3316-2025-16-1-3-44

Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках

Игорь Викторович **Винокуров**✉

Финансовый Университет при Правительстве Российской Федерации, Москва, Россия

✉ igvvinokurov@fa.ru

Аннотация. Массовое появление незаконных и незарегистрированных в Едином Государственном Реестре Недвижимости (ЕГРН) объектов недвижимости осложняет кадастровый учёт для многих субъектов территориального и административного уровня. Традиционные методы выявления объектов подобных типов, основанные на ручном анализе геопространственных данных, трудоёмки и требуют значительного времени.

Для повышения эффективности этого процесса предлагается автоматизировать обнаружение объектов на аэрофотоснимках путём решения задачи инстанс-сегментации с использованием модели глубокого обучения Mask R-CNN. В статье описана подготовка набора данных для этой модели, исследованы основные метрики качества и проанализированы полученные результаты. Показана эффективность модели Mask R-CNN при обнаружении объектов недвижимости, не имеющих регистрации в ЕГРН.

Ключевые слова и фразы: Кадастровый учёт, анализ аэрофотоснимков, инстанс-сегментация, Mask R-CNN, PyTorch

Для цитирования: Винокуров И. В. *Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках* // Программные системы: теория и приложения. 2025. Т. 16. № 1(64). С. 3–44. https://psta.psir.ru/read/psta2025_1_3-44.pdf

Введение

Коттеджные поселения, дачные кооперативы и садовые товарищества обладают высокой плотностью застройки, осложняющей автоматическое обнаружение на аэрофотоснимках незаконно построенных объектов недвижимости и объектов, не зарегистрированных в ЕГРН.

Для реализации аэрофотосъёмки в ППК «Роскадастр» используются квадрокоптеры. На полученных фотографиях ищутся объекты недвижимости, подлежащие **обязательной регистрации**. К таким объектам относятся:

- дачные дома, предназначенные для постоянного проживания,
- объекты, стоящие на фундаменте (бани, террасы, летние кухни),
- строения на участках для индивидуального жилищного строительства.

Ручной поиск ограничивает эффективность и точность этого процесса. С развитием моделей глубокого обучения, таких как RetinaNet [1], SOLO [2], YOLACT [3], Mask R-CNN [4] и других, появляется возможность значительно сократить время и повысить качество анализа изображений за счёт автоматического детектирования объектов. Перечисленные выше модели позволяют не только идентифицировать объекты на изображении, но и точно выделять их границы. Последнее особенно важно для анализа плотных застроек, где объекты могут быть частично перекрыты или находиться в близком соседстве друг с другом. Использование инстанс-сегментации, реализуемой этими моделями, наилучшим образом подходит для анализа геопространственных данных, поскольку наличие масок и ограничивающих рамок вокруг найденных экземпляров объектов позволяет не только обнаружить объекты, но и оценить их размеры. Это помогает заметить изменения размеров в результате перестройки объекта.

В работе приведены результаты исследования применимости модели Mask R-CNN (Mask Region-based Convolutional Neural Network) для обнаружения объектов недвижимости в коттеджных поселениях, дачных кооперативах и садовых товариществах. Особенностью предлагаемого подхода является обработка аэрофотоснимка с целью выявления ограничивающих рамок обнаруженных объектов, сопоставленных с определёнными географическими координатами, и сравнение их наличия или значимого изменения размеров с результатами аналогичной обработки аэрофотоснимков за прошлый период, как правило год. Такой подход позволяет выявлять изменения в застройке и постройки, не имеющие официальной регистрации.

1. Обзор работ по распознаванию объектов на аэрофотоснимках с использованием Mask R-CNN

Mask R-CNN является расширением модели Faster R-CNN, предназначенной для решения задач детектирования объектов и сегментации изображений [5]. Отличительной особенностью Mask R-CNN от Faster R-CNN является создание точных масок для каждого из обнаруженных на изображении объектов. Модель состоит из двух основных компонентов – обнаружение объектов и их сегментация [4].

Для извлечения признаков из входного изображения Mask R-CNN использует архитектуру ResNet [6] или другую свёрточную нейронную сеть. Эти признаки затем передаются в сеть предположения региона (RPN, Region Proposal Network) [4, 6], которая генерирует области интереса (RoI, Region of Interest) – предполагаемые местоположения объектов. Все полученные RoI передаются в основной классификатор, который определяет класс объекта и уточняет его границы.

В отличие от Faster R-CNN, в модели Mask R-CNN реализована ещё одна ветвь, отвечающая за создание масок сегментации. Эта ветвь использует небольшую свёрточную сеть с целью генерации бинарных масок для всех объектов в RoI. Размер маски соответствует размеру RoI.

Для обучения модели используется функция потерь, которая объединяет потери от классификации, регрессии координат и сегментации, что позволяет модели одновременно оптимизировать все три задачи [4].

Mask R-CNN хорошо справляется с перекрывающимися объектами и сложными фонами благодаря способности генерировать точные маски. Однако её производительность может зависеть от качества обучающего набора данных и выбранной архитектуры модели. В целом, Mask R-CNN является сильным инструментом для задач детекции и сегментации объектов в изображениях, обеспечивая высокую точность и гибкость в применении.

В настоящее время существует достаточно большое количество работ по распознаванию объектов различных типов на фотографиях и аэрофотоснимках с использованием этой модели, что свидетельствует об эффективности её применения для решения практических задач.

В работе [4] представлена основная архитектура Mask R-CNN и её применение к различным задачам сегментации, включая инстанс-сегментацию объектов.

В [7] авторы применяют Mask R-CNN для обнаружения зданий на спутниковых снимках высокого разрешения. Показывается высокая

точность сегментации и преимущества использования модели Mask R-CNN для решения подобных задач.

Обнаружение зданий на фотографиях из зон стихийных бедствий с использованием модели Mask R-CNN и предварительной обработки набора данных, с целью повышения эффективности его использования, приведено в [8]. В работе показаны преимущества этой модели перед другими моделями, реализующими детектирование объектов на фотографиях.

В работе [9] решается задача локализации полигонов зданий на спутниковых снимках высокого разрешения. Показывается эффективность применения модели Mask R-CNN для получения реальных границ контуров зданий на фотографиях с высокой плотностью объектов.

Метод обнаружения и сегментации кораблей на уровне пикселей с использованием модели Mask R-CNN предлагается в [10]. Отмечается преимущества использования этой модели и оценивается эффективность предлагаемого метода.

В [11] показывается, что извлечение контуров зданий из спутниковых снимков является сложной задачей из-за различий в масштабах, структурах и типов зданий. Для решения этой задачи предлагается использовать модель Mask R-CNN. Показывается эффективность её использования. Результаты извлечение отдельных зданий из спутниковых снимков предлагается использовать для приложений, автоматизирующих оценку населения, реализующих городское планирование и других.

Детектирование современной городской архитектуры с использованием модели Mask R-CNN, обученной на наборе данных, состоящим из элементов современных архитектурных стилей, приведено в [12]. В результате сравнения полученных результатов показана эффективность использования Mask R-CNN по сравнению с другими моделями.

В работе [13] показывается сложность обнаружения зданий и различного типа построек на спутниковых снимках из-за освещённости, плотности застроек, различных типов рельефов местности и других факторов. Для эффективного решения этой задачи предлагается использовать модель Mask R-CNN и собственный набор данных с усовершенствованной аугментацией изображений.

Применение модели Mask R-CNN для распознавания различных объектов на спутниковых снимках и аэрофотоснимках с целью автоматизации картографирования и поддержания карт местностей в актуальном состоянии описано в работе [14].

В [15] модель Mask R-CNN предлагается использовать для поддержания точности карт местностей с целью эффективного реагирования

на стихийных бедствия и катастрофы. Актуализировать карты местностей предлагается на основе аэрофотоснимков и спутниковых снимков. В работе получены более чем приемлемые результаты для фотографий с различным качеством, разрешениями и цветовыми каналами.

В [16, 17] предлагается гибридный подход к извлечению контуров зданий из спутниковых снимков низкого разрешения с использованием модели Mask R-CNN. Такой подход открывает перспективы для разработки автоматизированных инструментов обработки спутниковых снимков, и их эффективному использованию при мониторинге землепользования и реагирования на стихийные бедствия.

2. Обоснование выбора модели

Для реализации инстанс-сегментации объектов недвижимости на аэрофотоснимках, исходя из проведённого выше обзора, может быть обосновано выбрана модель Mask R-CNN, поскольку она обладает следующими преимуществами.

(1) Точность

- Модель Mask R-CNN демонстрирует более высокую точность детектирования объектов по сравнению с другими моделями [9, 10].
- Двухэтапный подход модели Mask R-CNN (предложение региона и сегментация [4–6]) позволяет более точно локализовать и сегментировать объекты с одновременным формированием их масок.

(2) Универсальность

- Mask R-CNN можно легко адаптировать для решения различных задач, включая обнаружение и контуризацию объектов, их инстанс- и семантическую сегментацию [4–6].
- Модульная архитектура этой модели позволяет легко добавлять или удалять её компоненты по мере необходимости.

(3) Надежность

- Mask R-CNN более устойчива к шумам и искажениям в изображениях [15].
- Двухэтапный подход этой модели помогает уменьшить количество ложных срабатываний и улучшить общую надежность [4–6].

(4) Поддержка объектов разных размеров

- Mask R-CNN может эффективно сегментировать объекты разных размеров, от маленьких до больших [4, 8].

- Механизм предложения региона [4] позволяет этой модели обнаруживать объекты независимо от их типа и размера (например, СОСО [5]).

(5) Меньше проблем с переобучением.

- Mask R-CNN менее склонна к переобучению [11].
- Использование нескольких функций потерь и двухэтапный подход помогают модели лучше обобщать результаты [4–6, 18]

Модели Mask R-CNN присущи и некоторыми недостатками, которые не являются критичными при решении задачи автоматизации обработки изображений, – это относительно невысокая скорость работы и сложность реализации [19].

Помимо рассмотренных выше моделей, для решения задачи инстанс-сегментации может быть использована и достаточно популярная в последнее время модель YOLO [20]. Несмотря на высокую скорость получения результатов, эта модель обладает существенными недостатками – невысоким качеством распознавания групп небольших объектов из-за ограниченного числа кандидатов для ограничивающих рамок (две) и возможностью дублирования ограничивающих рамок для одного и того же объекта [21]. Сравнение результатов распознавания объектов недвижимости, с аналогичными результатами, полученных с использованием последней версии модели YOLO, приведено в п. 6

3. Постановка задачи

Как уже было отмечено выше, фотографирование исследуемых ППК «Роскадстр» территорий осуществляется с использованием квадрокоптеров. Квадрокоптер (на момент написания статьи – это Phantom 4 RTK) реализует движение по заранее заданной траектории. Камера автоматически делает снимки при достижении определенных точек маршрута, формируя либо его полное изображение, либо изображения фрагментов маршрута (в большинстве случаев с перекрытием).

Целью данной работы является исследование применимости модели Mask R-CNN для реализации автоматической или автоматизированной обработки полученных снимков и выявления на них незаконно построенных или незарегистрированных объектов недвижимости. Для достижения поставленной цели в работе решались следующих задач:

- (1) формирование собственного набора данных для обучения модели,
- (2) анализа качества работы модели,
- (3) анализ результатов применения модели для обнаружения объектов незаконного строительства или объектов недвижимости, не зарегистрированных в ЕГРН.

4. Формирование набора данных

4.1. Аннотирование изображений

Основной задачей при формировании набора данных, используемого для обучения и тестирования модели, является отметка или аннотирование исходных изображений, представляющее собой полигональную контуризацию распознаваемых на фотографиях объектов. Точность контуризации определяет точность обнаружения объектов.

В настоящее время существует большое количество ПО, реализующего аннотирование изображений. Для формирования набора данных было выбрано ПО с открытым исходным кодом **LabelMe**.

Процесс аннотирования предполагает присваивание метки каждому распознаваемому объекту того или иного типа. При аннотировании аэрофотографий метки присваивались следующим 5-ти типам объектов, таблица 1.

Таблица 1. Типы объектов и их метки

Тип (класс) объекта	Имя метки и цвет сегментации класса
Коттеджный или дачный домик	<i>building</i>
Теплица (парник)	<i>greenhouse</i>
Хозяйственная постройка	<i>outbuilding</i>
Транспортное средство	<i>vehicle</i>
Бассейн	<i>swimming</i>

Основным типом объектов является коттеджный или дачный домик. Остальные типы объектов используются с целью предотвращения ложного распознавания последних и с целью возможного продолжения работы в данном направлении.

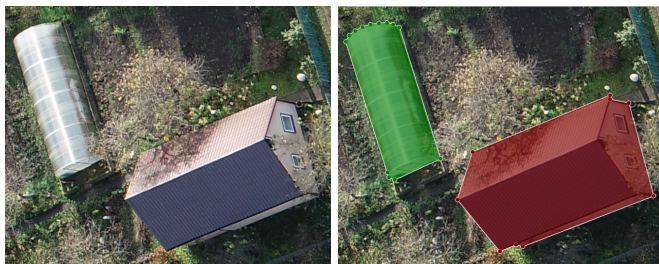
Результаты аннотирования сохранялись в формате JSON, поддерживаемом LabelMe. При формировании набора данных было осуществлено аннотирование 435 полученных с квадрокоптера фотографий. Пример фотографии фрагмента дачного кооператива и реализованная в процессе аннотирования полигональная контуризация 2-х классов объектов – дачного домика и теплицы в LabelMe приведены на рисунке 1, 2а и 2б соответственно.

Сформированный для обучения и тестирования модели набор данных представляет собой совокупность из указанного выше количества аэрофотоснимков и такого же количества JSON-файлов в формате ПО LabelMe, содержащих массивы координат точек полигональной контуризации и имена для объектов 5-ти типов.

Для обучения и тестирования модели было случайным образом выбрано 80% и 20% элементов этого набора данных соответственно.



Рисунок 1. Фотография фрагмента дачного кооператива с квадрокоптера Phantom 4 RTK



(а) Фрагмент изображения с дачным домиком и теплиц

(б) Контуризация объектов 2-х типов

Рисунок 2. Элемент набора данных – изображение объектов и их полигональная контуризация

Диаграмма распределения на аэрофотоснимках объектов, принадлежащих этим классам, приведена на рисунке 3.

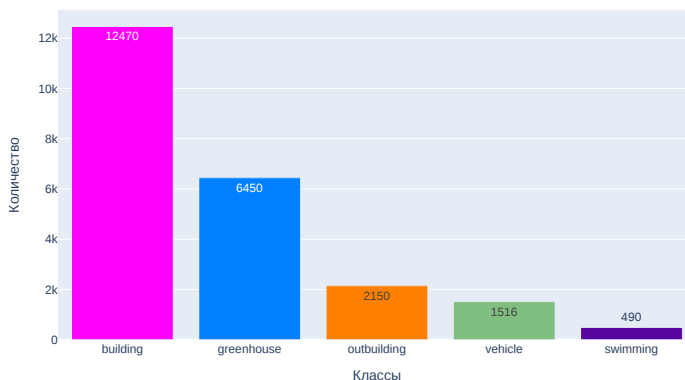


Рисунок 3. Распределение классов на фотографиях набора данных

4.2. Аугментация набора данных

Для повышения обобщающей способности модели сформирован и применён к изображениям из обучающего набора данных конвейер преобразований, реализованных в пакете PyTorch. В самом начале на основе метрики IoU (Intersection over Union) осуществлялось обрезка изображений с сохранением ROI. Затем применялись стохастические преобразования цвета, включающие вариации яркости, контрастности, насыщенности и цветового тона, а также случайное преобразование в градации серого (RandomGrayscale) и выравнивание гистограммы (RandomEqualize).

Дальнейшее увеличения разнообразия заключалось в уменьшении количества цветовых уровней и горизонтального отражения изображений с использованием методов RandomPosterize и RandomHorizontalFlip соответственно. Помимо этого было реализовано масштабирование изображений до максимального размера с сохранением соотношения сторон, дополнения до квадратной формы, и изменении размера до требуемого с использованием антиалиасинга. На заключительном этапе осуществлялось преобразование типа данных в torch.float32 с масштабированием значений пикселей и валидация координат ограничивающих рамок (SanitizeBoundingBoxes).

5. Формирование и исследование модели

5.1. Формирование модели

Mask R-CNN характеризуется относительно высокой сложностью реализации. Как следствие, успешное применение этой модели требует тщательной настройки параметров и архитектуры сети. В работе была

выбрана достаточно быстрая и точная модель обнаружения объектов `maskrcnn_resnet50_fpn_v2` из пакета `torchvision`. Для повышения эффективности извлечения признаков (backbone) в этой модели была выбрана архитектура ResNet-101 с Feature Pyramid Network (FPN) [4]. Общие принципы её формирования и обучения достаточно хорошо описаны в [22] и [23].

Модель предварительно обучена на наборе данных COCO [5]. Для этой модели было экспериментально найдено оптимальное количество эпох дообучения, выбраны оптимизатор `Adam` и шедюлер `OneCycleLR` (механизмы, определяющие уменьшение весов во время обучения и скорость этого процесса), изменено количество выходных каналов и настроено разбиение для якорных боксов – заранее определенных прямоугольных рамок, которые используются для предложения потенциальных RoI при обнаружении объектов на изображении. Код формирования и исследования модели написан с использованием библиотеки машинного обучения PyTorch [24].

Для эффективной реализации инстанс-сегментации объектов различных масштабов, первоначальный свёрточный слой ResNet-101 был модифицирован. Перед слоем с достаточно большим размером ядра (7×7) были добавлены несколько параллельных свёрточных слоёв, имеющих ядра различных размеров, рисунок 4, 5.

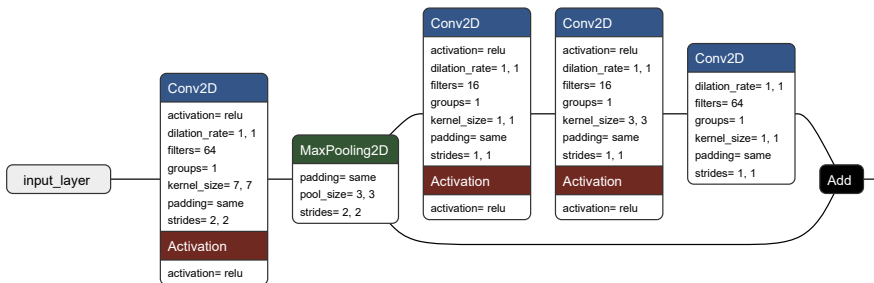


РИСУНОК 4. Начальный свёрточный слой в ResNet-101 и первый residual-блок [22]

Это потенциально позволяет модели одновременно учитывать как контекстуальные особенности (ядро 5×5), так и детальные характеристики объектов (ядра 1×1 и 3×3). Помимо этого, параллельные свёрточные слои позволяют ускорить сходимость модели.

Попытки вариативного изменения residual-блоков этой модели к какому-либо значимому улучшению детектирования объектов не привели.

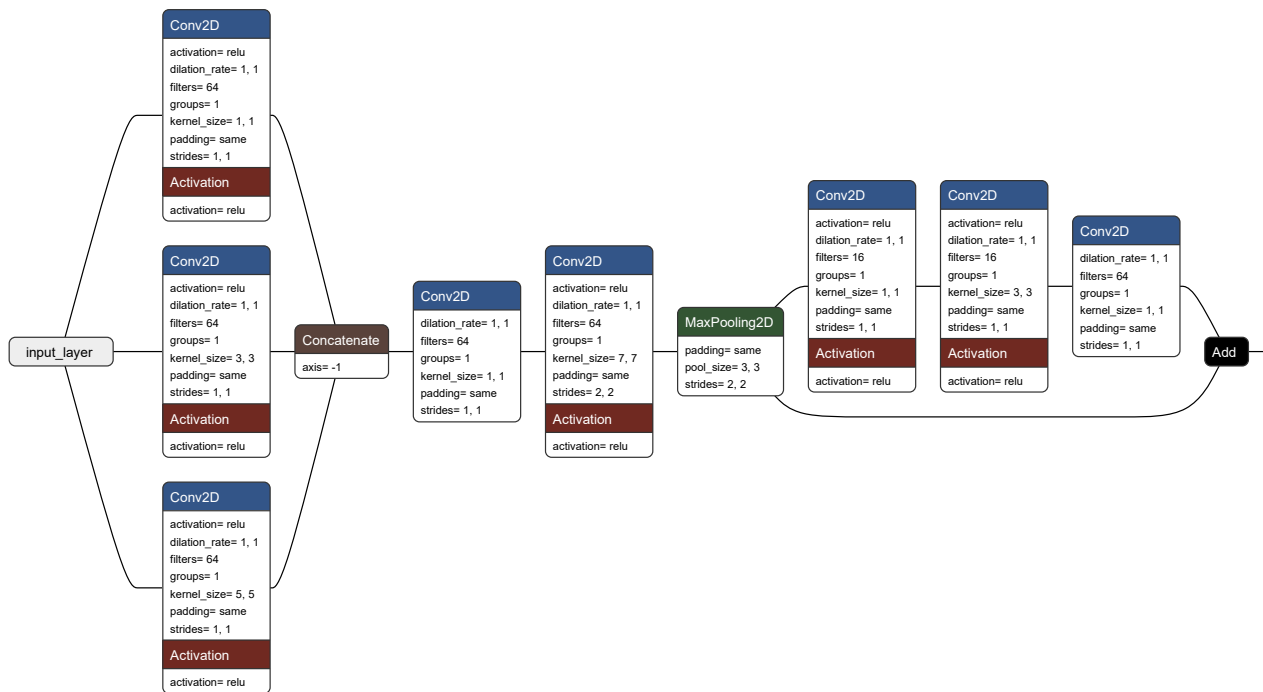


Рисунок 5. Параллельные свёрточные слои в ResNet-101

5.2. Исследование модели

Для анализа точности работы модели в процессе проведения экспериментальных исследований осуществлялось вычисление трёх ключевых метрик – **Loss**, **mAP** и **CS**.

Loss в Mask R-CNN позволяет оценить, насколько точно модель предсказывает классификацию, регрессию координат и сегментацию объектов. Чем ближе значения этой метрик к 0 (0%), тем более точным является результат работы модели. Изменение среднего значения этой метрики для всех 5-ти классов из наборов данных для обучения (train) и тестирования (valid) в зависимости от эпохи обучения приведено на рисунке 6.

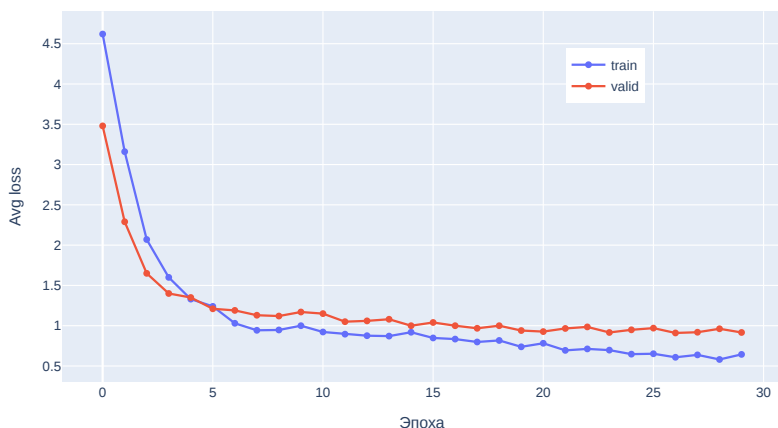


Рисунок 6. Изменение среднего значения функции потерь модели

Метрика **mAP** (mean Average Precision) является средним значением метрики точности по всем классам объектов. Аналогичная метрика **mAP50-95** рассчитывается при различных порогах перекрытия ограничивающих рамок **IoU** (Intersection over Union) – от 50% до 95%. Обе метрики учитывают как точность (**precision**), так и полноту (**recall**) на различных уровнях **IoU**. Чем ближе значения каждой из этих метрик к 1 (100%), тем более точным является результат работы модели. Зависимости метрик **mAP** и **mAP50-95** для всех классов из наборов данных для обучения (train) и тестирования (valid) от эпохи обучения приведена на рисунке 7.

Метрика **CS** (Confidence Score) представляет собой вероятность того, что объект принадлежит к определенному классу. **CS** является мерой уверенности модели в том, что предсказанный объект действительно

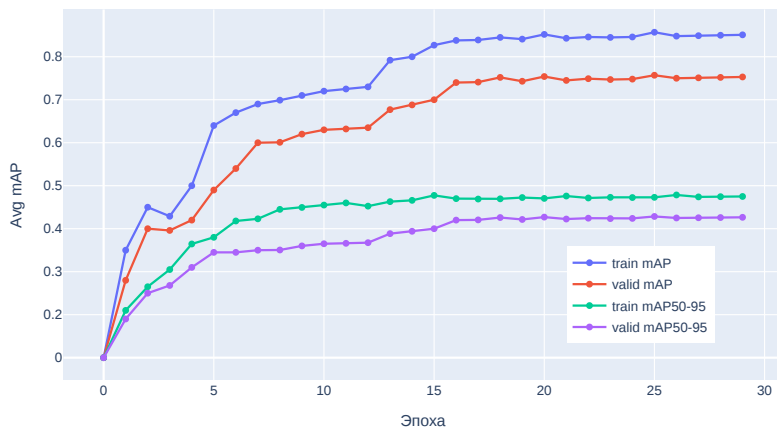


Рисунок 7. Изменение среднего значения метрик mAP и $mAP50-95$ при детектировании ограничивающих рамок

присутствует в ограничивающей рамке. Эта метрика особенно важна для моделей, реализующих детектирование объектов. Значения CS варьируются от 0 до 1 (100%); более высокие значения указывают на большую уверенность модели в правильности своего предсказания. Зависимость этой метрики для всех классов из наборов данных обучения (train) и тестирования (valid) от эпохи обучения показано на рисунке 8.

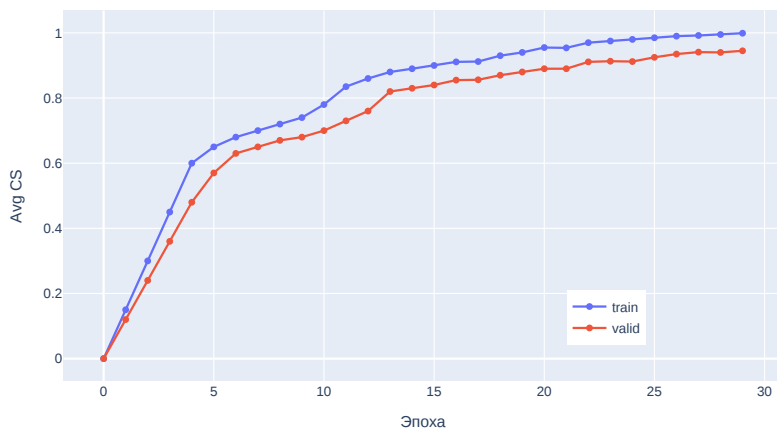


Рисунок 8. Изменение среднего значения метрики CS при детектировании объектов

В таблице 2 приведены значения метрик точности для обучающего и тестового наборов данных на последней эпохе обучения модели, см. рисунки 6–8.

Таблица 2. Значения метрик точности на последней эпохе обучения модели

Набор	Loss	mAP	mAP50-95	CS
train	0.741	0.851	0.475	0.999
valid	0.915	0.753	0.416	0.991

Вычисление значений метрик точности в Google Colab Pro в среде выполнения T4 GPU занимало порядка 80-100 минут. Из приведенных выше графиков и таблицы следует, что обученная на собственном наборе модель Mask R-CNN обладает более чем приемлемой точностью выявления интересующих нас объектов недвижимости на аэрофотоснимках. Несколько примеров детектирования дачных домиков из наборов данных для обучения модели и для её тестирования приведены на рисунках 9 и 10 соответственно.

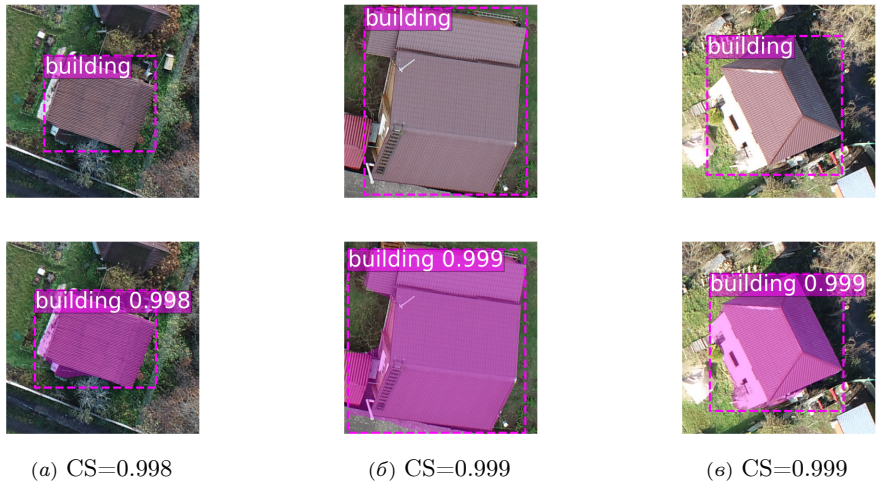


Рисунок 9. Детектирование объектов недвижимости на изображениях из тестового набора данных

При использовании этой модели в 3-5% случаев были выявлены незначительные ошибки, проявляющиеся в виде неверной идентификации объектов или неправильной сегментации границы. Появление этих ошибок связано с плотностью расположения сегментируемых объектов, с недостаточно представительным набором данных, наличием шумов на снимках



Рисунок 10. Детектирование объектов недвижимости на изображениях из тестового набора данных

и недостаточной резкостью области изображения. Матрица ошибок детектирования объектов недвижимости на изображениях из тестового набора данных показана на рисунке 11.

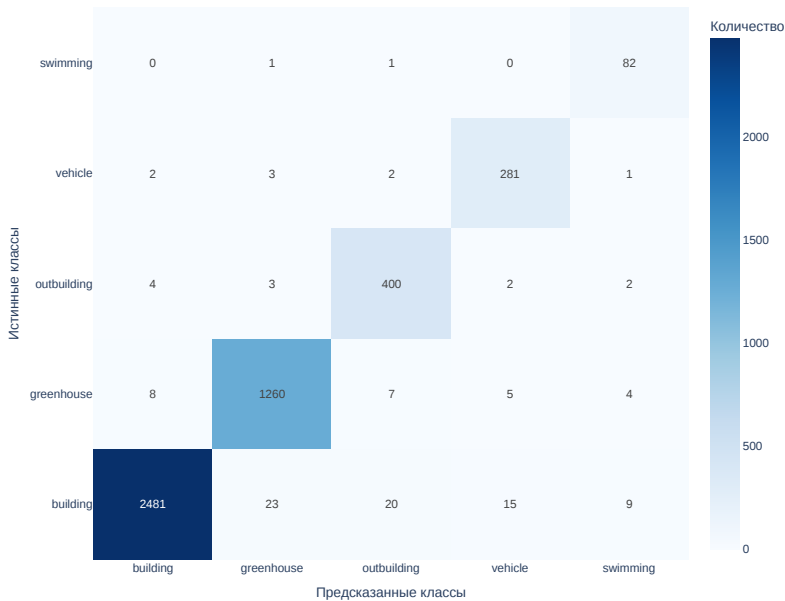


Рисунок 11. Матрица ошибок модели для тестового набора данных

Пример неверной идентификации для нескольких объектов классов *greenhouse*, *outbuilding* и *vehicle* приведён на рисунке 12. Легко заметить, что все объекты класса *building*, из-за достаточно большого их количества в обучающем наборе данных, идентифицируются верно, что является вполне достаточным для поставленной выше цели работы.

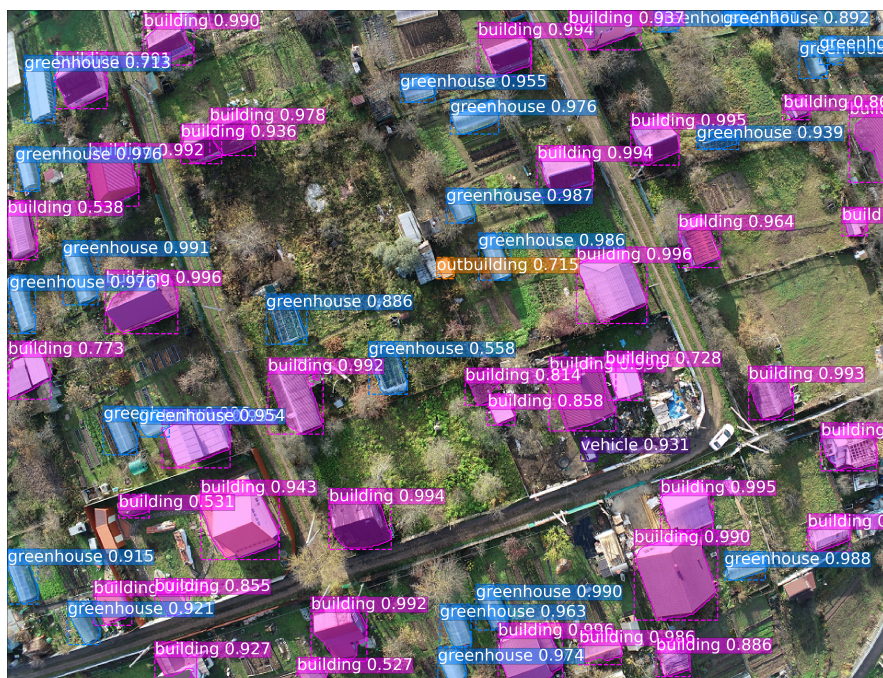


Рисунок 12. Результаты детектирования объектов разных типов на фотографии фрагмента дачного кооператива (см. рисунок 1)

6. Сравнение с моделью YOLO

Модель глубокого обучения YOLO в настоящее время интенсивно развивается и наравне с Mask R-CNN может быть использована для реализации детектирования объектов недвижимости на аэрофотоснимках. Исследования, аналогичные приведённым выше, были проведены и для модели *yolo11n-seg* от Ultralytics. На рисунке 13 приведены зависимости значений метрик точности mAP и mAP50-95 при детектировании ограничивающих рамок объектов недвижимости от эпохи обучения. Из этого рисунка видно, что эти метрики достигают значений, сопоставимых с полученными моделью Mask R-CNN, на большем количестве эпох.

Достаточно часто при анализе полученных результатов были выявлены случаи соотнесения обнаруженного объекта с разными классами. На рисунке 14, отображающем результаты детектирования, аналогичные приведённым на рисунке 12, видно несколько таких случаев (например, с объектами классов *building* и *outbuilding*). При исследовании результа-

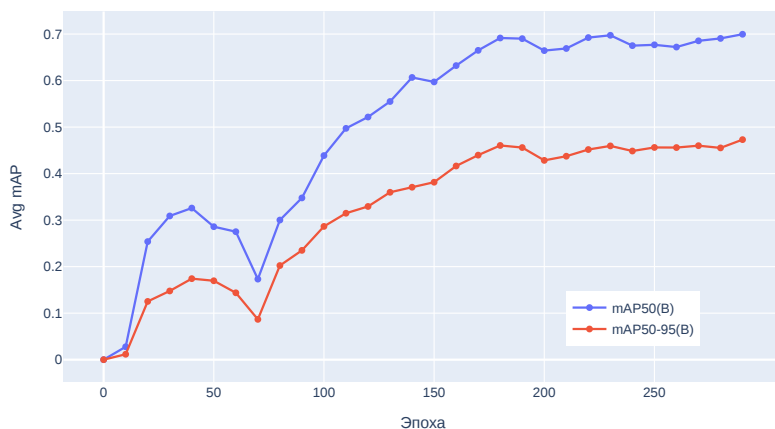


РИСУНОК 13. Изменение среднего значения метрик mAP и mAP50-95 модели YOLO при детектировании ограничивающих рамок

тов, полученных с использованием Mask R-CNN ни одного аналогичного случая выявлено не было. Кроме этого, при анализе достаточно большого количества результатов детектирования, было замечено более точное (до 10%-15%, см. рисунки 12 и 14) определение масок объектов моделью Mask R-CNN по сравнению с YOLO. Неправильное (неточное) определение маски объекта может быть критичным для ИС ППК «Роскадастр».

В целом, модель YOLO от Ultralytics оставила благоприятное впечатление простотой использования и наличием готового функционала для проведения экспериментов и анализ полученных результатов. В ряде случаев, в зависимости от специфики решаемой задачи и набора данных, использование моделей YOLO позволяет получить немного лучший результат, по сравнению с Mask R-CNN [25].

7. Практическая реализация результатов

Результаты, полученные в ходе работ по повышению эффективности обнаружения коттеджных и дачных домиков на аэрофотоснимках с использованием модели Mask R-CNN, были реализованы в бета-версии одной из подсистем информационной системы (ИС) ППК «Роскадастр». Основная цель этой подсистемы заключается в определении наличия регистрации объектов строительства в ЕГРН. В этой подсистеме реализуется процесс, включающий ортотрансформирование, создание цифровой модели местности (ЦММ), получение точек привязки и расчёт координат объектов недвижимости [26].



Рисунок 14. Результаты детектирования объектов разных типов с использованием YOLO

Фотография, сделанная с квадрокоптера, разбивается на некоторое количество фрагментов (тайлов), которые анализируются на предмет наличия или отсутствия незарегистрированных объектов. Размер тайла в пикселях определяется масштабом изображения и детектирующими способностями модели. Разбиение фотографии на фрагменты обосновывается её большим размером, который может составлять нескольких сотен тысяч пикселей и размера, достигающим 1Т. Пример обнаружения незарегистрированного дачного домика на одном из фрагментов аэрофотоснимка приведён на рисунке 15.

Заключение

В данной работе была проведена оценка возможности использования модели Mask R-CNN с изменённым backbone для обнаружения и сегментации на аэрофотоснимках коттеджных и дачных домиков. Для обучения модели был создан собственный набор данных, включающий в себя изображения объектов и их аннотации. Проведённые исследования показали, что модель Mask R-CNN успешно справляется с задачей

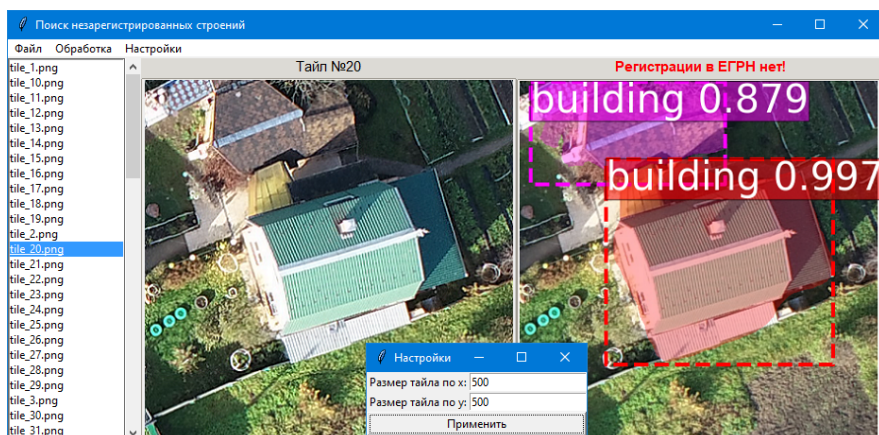


Рисунок 15. Пример выявления незарегистрированного дачного домика на фрагменте аэрофотоснимка

инстанс-сегментации, демонстрируя приемлемую точность по метрикам Loss, mAP, mAP50-95 и CS.

Разработанная бета-версия подсистемы ИС ППК «Роскадастр» на базе Mask R-CNN использует эту модель для распознавания объектов на изображениях, полученных с квадрокоптера, что существенно сокращает время и ресурсы, затрачиваемые на выявление нелегальной и незарегистрированной застройки, и повышает эффективность работы по контролю за использованием земельных ресурсов.

Дальнейшие исследования по применению модели Mask R-CNN будут направлены на повышение её точности в результате расширения и оптимизации обучающего набора данных, а также на интеграцию этой модели в существующие системы мониторинга и контроля за использованием земельных ресурсов. Кроме того, перспективным направлением является разработка алгоритмов автоматической верификации результатов работы модели, основанных на сравнении с данными ЕГРН.

Список использованных источников

- [1] Lin T.-Y., Goyal P., Girshick R. B., He K., Dollár P. *Focal loss for dense object detection.* – Computing Research Repository (CoRR). – 2017. – 10 pp. arXiv 1708.02002 24
- [2] Wang X., Kong T., Shen Ch., Jiang Y., Li L. *SOLO: Segmenting objects by locations.* – Computing Research Repository (CoRR). – 2019. – 19 pp. arXiv 1912.04488 24
- [3] Duda K., Ivanov A. *On decidability of amenability in computable groups* // Archive for Mathematical Logic. – 2022. – Vol. 61. – Pp. 891–902. 24

- [4] He K., Gkioxari G., P. Dollár, Girshick R. B. *Mask R-CNN*.– Computing Research Repository (CoRR).– 2017.– 12 pp. arXiv  1703.06870  ↑^{24, 25, 27, 28, 32}
- [5] Ren S., He K., Girshick R. B., Sun J. *Faster R-CNN: Towards real-time object detection with region proposal networks*.– Computing Research Repository (CoRR).– 2015.– 14 pp. arXiv  1506.01497  ↑^{25, 27, 28, 32}
- [6] He K., Zhang X., Ren S., Sun J. *Identity mappings in deep residual networks*.– Computing Research Repository (CoRR).– 2016.– 15 pp. arXiv  1603.05027  ↑^{25, 27, 28}
- [7] Lin T.-Y., Maire M., Belongie S. J., Bourdev L. D., Girshick R. B., Hays J., Perona P., Ramanan D., P. Dollár, Zitnick C. L. *Microsoft COCO: Common objects in context*.– Computing Research Repository (CoRR).– 2014.– 15 pp. arXiv  1405.0312  ↑²⁵
- [8] Xu Y., Wu L., Xie Z., Chen Z. *Building extraction in very high resolution remote sensing imagery using deep learning and guided filters* // Remote. Sens.– 2018.– Vol. **10**.– No. 1.– id. 144.– 18 pp.  ↑^{26, 27}
- [9] Han Q., Yin Q., Zheng X., Chen Z. *Remote sensing image building detection method based on Mask R-CNN* // Complex Intell. Syst.– 2022.– Vol. **8**.– Pp. 1847–1855.  ↑^{26, 27}
- [10] Zhao K., Kang J., Jung J., Sohn G. *Building extraction from satellite images using Mask R-CNN with building boundary regularization* // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (18–22 June 2018, Salt Lake City, UT, USA).– IEEE.– ISBN 9781538661017.– id. 242.– 4 pp.  ↑^{26, 27}
- [11] Nie X., Duan M., Ding H., Hu B., Wong E. K. *Attention Mask R-CNN for ship detection and segmentation from remote sensing images* // IEEE Access.– 2020.– Vol. **8**.– Pp. 9325–9334.  ↑^{26, 28}
- [12] Jenila Vincent M., Varalakshmi P. *Extraction of building footprint using MASK-RCNN for high resolution aerial imagery* // Environmental Research Communications.– 2024.– Vol. **6**.– No. 7.– id. 075015.– 17 pp.  ↑²⁶
- [13] Zhu X., Hu L., Wang J. *Urban modern architecture recognition based on Mask-RCNN and ECA attention mechanism*, Fifth International Conference on Geoscience and Remote Sensing Mapping (ICGRSM 2023) (13–15 October 2023, Lianyungang, China), Proc. SPIE.– vol. **12980**.– 2024.– ISBN 9781510672789.– id. 129801D.  ↑²⁶
- [14] Raghavan R., Chander Verma D., Pandey D., Anand R., Kumar Pandey B., Singh H. *Optimized building extraction from high-resolution satellite imagery using deep learning* // Multimedia Tools and Applications.– 2022.– Vol. **81**.– No. 29.– Pp. 42309–42323.  ↑²⁶
- [15] Ulanov D., Syrov A. *Building footprint extraction based on RGBD satellite imagery*, CS230 Deep Learning (Winter 2020, Stanford University, CA).– 2020.– 11 pp.  ↑^{26, 27}
- [16] Solanki A., Singh R. K., Demeneze B. *Aerial pictures semantic segmentation applying deep learning* // International Journal of Trendy Research in Engineering and Technology.– 2021.– Vol. **5**.– No. 1.– Pp. 42–48.  ↑²⁷
- [17] NourEldeen A., Wahed M. E. *Enhanced building footprint extraction from satellite imagery using Mask R-CNN and PointRend* // Bulletin of Electrical Engineering and Informatics.– 2024.– Vol. **5**.– No. 13.– Pp. 3601–3608.  ↑²⁷

- [18] He K., Zhang X., Ren S., Sun J. *Deep residual learning for image recognition.*– Computing Research Repository (CoRR).– 2015.– 12 pp. arXiv:1512.03385 doi ↑28
- [19] Mills Ch. J. *PyTorch Mask R-CNN tutorial.*– GitHub repository.– 2023. URL ↑28
- [20] Redmon J., Divvala S., Girshick R. B., Farhadi A. *You Only Look Once: Unified, real-time object detection.*– Computing Research Repository (CoRR).– 2015.– 10 pp. arXiv:1506.02640 doi ↑28
- [21] Khanam R., Hussain M. *YOLOv11: An overview of the key architectural enhancements.*– 2024.– 9 pp. arXiv:2410.17725 doi ↑28
- [22] Lin T. -Y., P. Dollár, Girshick R. B., He K., Hariharan B., Belongie S. J. *Feature pyramid networks for object detection.*– Computing Research Repository (CoRR).– 2016.– 10 pp. arXiv:1612.03144 doi ↑32
- [23] Waleed A. *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow.*– GitHub repository.– 2017. URL ↑32
- [24] Stevens E., Antiga L., Viehmann T. *Deep Learning with PyTorch.*– New York: Manning Publications.– 2020.– ISBN 9781617295263.– 520 pp. ↑32
- [25] Sapkota R., Dawood A., Karkee M. *Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments* // Artificial Intelligence in Agriculture.– 2024.– Vol. 13.– No. 1.– Pp. 84–99. doi ↑39
- [26] Булавицкий В. Ф. *Применение беспилотных летательных аппаратов для оперативного получения аэрофотоснимков местности* // Электронное научное издание «Учёные заметки ТОГУ».– 2013.– Т. 4.– № 4.– С. 1747–1755. ✳ ↑39

Поступила в редакцию 21.10.2024;
одобрена после рецензирования 24.12.2024;
принята к публикации 11.01.2025;
опубликована онлайн 31.01.2025.

Рекомендовал к публикации

к.т.н. В. П. Фраленко

Информация об авторе:



Игорь Викторович Винокуров

Кандидат технических наук (PhD), ассоциированный профессор в Финансовом Университете при Правительстве Российской Федерации. Область научных интересов: информационные системы, информационные технологии, технологии обработки данных



0000-0001-8697-1032

e-mail: igvvinokurov@fa.ru

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*