

**ОБНАРУЖЕНИЕ И КЛАССИФИКАЦИЯ ТРАНСПОРТНЫХ СРЕДСТВ НА СНИМКАХ СВЕРХВЫСОКОГО РАЗРЕШЕНИЯ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ**

Ch. Chen<sup>1,2</sup>, A. A. Минальд<sup>3</sup>, Р. П. Богуш<sup>4</sup>,  
G. Ma<sup>5</sup>, Y. Weichen<sup>5</sup>, С. В. Абламейко<sup>3,6\*</sup>

УДК 535.317.1

<https://doi.org/10.47612/0514-7506-2022-89-2-275-282>

<sup>1</sup> Школа информационных наук и технологий Университета Чжэцзяна Шурен, Ханчжоу, Китай

<sup>2</sup> База международного научно-технического сотрудничества провинции Чжэцзян по обработке и применению изображений дистанционного зондирования, Ханчжоу, Китай

<sup>3</sup> Белорусский государственный университет, Минск, Беларусь; e-mail: [ablameyko@yandex.by](mailto:ablameyko@yandex.by)

<sup>4</sup> Полоцкий государственный университет, Новополоцк, Беларусь

<sup>5</sup> EarthView Image Inc., Хучжоу, Китай

<sup>6</sup> Объединенный институт проблем информатики НАН Беларуси, Минск, Беларусь;  
e-mail: [ablameyko@yandex.by](mailto:ablameyko@yandex.by)

(Поступила 14 февраля 2022)

Предлагается архитектура глубокой нейронной сети, основанная на интеграции сверточной нейронной сети Faster R-CNN с модулем Feature Pyramid Network. На основе данного подхода разработан алгоритм обнаружения и классификации транспортных средств на изображениях и соответствующая модель. Для обучения предложенной модели использована кроссплатформенная среда ML.NET. Представлены результаты сравнения эффективности применения предложенного подхода и сверточных нейронных сетей YOLO v4 и Faster R-CNN. Показано улучшение точности обнаружения и локализации разных типов транспортных средств на снимках сверхвысокого разрешения. Приведены примеры обработки изображений земной поверхности сверхвысокого разрешения и даны соответствующие рекомендации.

**Ключевые слова:** обнаружение и классификация объектов, изображения земной поверхности, изображения сверхвысокого разрешения, нейронная сеть.

The paper proposes a deep neural network architecture based on the integration of the convolutional neural network Faster R-CNN with the Feature Pyramid Network module. Based on this approach, an algorithm for detecting and classifying vehicles in images and a corresponding model have been developed. A cross-platform environment ML.NET was used to train the proposed model. The results of comparing the effectiveness of the proposed approach and convolutional neural networks YOLO v4 and Faster R-CNN are presented. The improvement of the accuracy of detection and localization of different types of vehicles in ultra-high resolutions images is shown. Examples of processing ultra-high resolutions remote sensing images and appropriate recommendations are given.

**Keywords:** object detection and classification, remote sensing images, ultra-high resolutions images, neural network.

**DETECTION AND CLASSIFICATION OF VEHICLES IN ULTRA-HIGH RESOLUTIONS IMAGES USING NEURAL NETWORKS**

Ch. Chen<sup>1,2</sup>, A. A. Minald<sup>3</sup>, R. P. Bohush<sup>4</sup>, G. Ma<sup>5</sup>, Y. Weichen<sup>5</sup>, S. V. Ablameyko<sup>3,6\*</sup> (<sup>1</sup> School of Information Science and Technology at Zhejiang Shuren University, Hangzhou, China; <sup>2</sup> International Science and Technology Cooperation Base of Zhejiang Province: Remote Sensing Image Processing and Application, Hangzhou, China; <sup>3</sup> Belarusian State University, Minsk, Belarus; e-mail: [ablameyko@yandex.by](mailto:ablameyko@yandex.by); <sup>4</sup> Polotsk State University, Novopolotsk, Belarus; <sup>5</sup> EarthView Image Inc., Huzhou, China; <sup>6</sup> United Institute for Informatics Problems, National Academy of Sciences of Belarus, Minsk, Belarus; e-mail: [ablameyko@yandex.by](mailto:ablameyko@yandex.by))

**Введение.** Развитие транспортного движения провоцирует высокий спрос на контроль за дорожным движением в городских районах. В настоящее время контроль за дорожным движением осуществляется с помощью множества наземных датчиков, таких как индукционные петли и стационарные видекамеры. Однако датчики подобного рода лишь частично фиксируют поток движения на главных дорогах. Мониторинг транспортного движения на второстепенных небольших дорогах выполняется очень редко. При этом не осуществляется сбор информации о припаркованных на дорогах автомобилях. Следовательно, общесистемные изображения всей дорожной сети должны дополнять эти случайным образом полученные данные.

Один из способов решения данной задачи — контроль движения транспортных средств (ТС) с помощью снимков земной поверхности, получаемых со спутников или беспилотных летательных аппаратов (БПЛА). С технической точки зрения задача сводится к обнаружению и классификации заданных ТС на изображениях. В настоящее время разработано достаточно много подходов к обнаружению объектов на изображениях земной поверхности. В последнее десятилетие для решения данной задачи стали широко применяться нейронные сети.

В работе [1] предлагается усовершенствованный метод обнаружения ТС на основе сверточных нейронных сетей (СНС) YOLO (You Only Look Once) v3, который направлен на повышение точности выделения предполагаемых областей, содержащих объекты. В [2] для повышения точности обнаружения ТС на перекрестках дорог предложено использовать СНС AlexNet и Faster R-CNN. В [3] рассматривается подход, позволяющий незначительно повысить точность обнаружения за счет метода неконтролируемой доменной адаптации. В [4] представлены эффективные методы обнаружения объектов на изображениях высокого разрешения, которые комбинируют две взаимодополняющие стратегии: прогнозирование ограничивающих рамок на основе соседних визуальных признаков и использование функции изображения высокого уровня для управления двухэтапным процессом поиска, который адаптивно фокусируется на областях, содержащих небольшие объекты. Авторы извлекают функции, необходимые для двух стратегий, с помощью предварительно обученной модели DCNN. В [5] предлагается система обнаружения ТС на аэрофотоснимках высокого разрешения. В частности, авторы включают несколько разнообразных и взаимодополняющих дескрипторов изображений, в том числе гистограмму ориентированных градиентов и локальный двоичный шаблон, а для постобработки используют процедуру для устранения повторяющихся обнаружений за счет немаксимального подавления. Алгоритм для точного обнаружения объектов малых размеров на видеоизображениях высокого разрешения предложен в [6]. Выполняется разбиение изображения на блоки, размер которых равен размеру входного слоя СНС, для обнаружения объектов в каждом блоке используется СНС YOLO. Алгоритм позволяет увеличить точность обнаружения объектов небольших размеров на изображениях с большим разрешением и при этом выполнять параллельную обработку на современных вычислительных средствах.

Среди известных моделей СНС, различающихся архитектурой, точностью и быстродействием, для обнаружения объектов на изображениях и видеопоследовательностях следует выделить модификации R-CNN (Region-based Convolutional Neural Networks) [7], ResNet (Residual Networks) [8] и YOLO [9].

Изображения, получаемые с серийных спутников IKONOS, QuickBird, Канопус, обычно имеют разрешение 0.5—1.0 м. На таких спутниковых снимках высокого разрешения можно определить ТС, однако при таком разрешении детали изображения различаются недостаточно, чтобы классифицировать ТС с помощью типовых подходов. На изображениях сверхвысокого разрешения (0.1—0.4 м) ТС имеют достаточные размеры, что позволяет разделить их на классы. Однако обнаружение и классификация ТС на снимках густонаселенного города сверхвысокого пространственного разрешения — довольно сложная задача. ТС имеют разную форму и размеры, могут быть затенены, на их распознавание влияют климатические условия и другие факторы. На дорожном полотне помимо ТС имеются другие объекты: дорожные знаки, лужи, тени, отбрасываемые зданиями и деревьями вблизи дороги. В то же время использование этих данных позволяет достаточно точно подсчитать и проанализировать присутствующие ТС, что необходимо для решения многих задач управления городом. Данная задача очень сложная, и к ее решению применялись различные подходы (см., например, [7—9]). Проведенные исследования показывают сложность задачи обнаружения объектов на изображениях сверхвысокого разрешения. Это связано с широтой палитры представления объектов на таких изображениях, отсутствием наборов данных для обучения сетей, что приводит к невысокому качеству обнаружения и распознавания объектов.

В настоящей работе для обнаружения и классификации ТС предлагается использовать сеть Faster R-CNN, дополнив ее модулем FPN. Разработаны решения, позволяющие обнаруживать и качественно классифицировать ТС на снимках сверхвысокого (0.1—0.3 м) разрешения. По сравнению с другими моделями улучшены количественные характеристики данного процесса.

**Архитектура СНС.** В ряду известных моделей СНС для решения поставленной задачи перспективным является класс сетей R-CNN [10]. Первая версия R-CNN включает в себя этапы предварительного выделения регионов на изображении, вычисления признаков с использованием СНС, например AlexNet, применения классификатора для выделенных признаков и корректировки координат регионов на изображении. Для классификации используется, как правило, метод на основе машин опорных векторов (SVM-классификатор). Во многих случаях модель выделяет множество фрагментов на изображении, которые принадлежат одному и тому же объекту, но сдвинуты на несколько пикселей относительно друг друга. Поэтому применяется не максимальное подавление — подход, основанный на выборе фрагмента изображения с наибольшей вероятностью принадлежности к искомому объекту и корректировке его положения на основе анализа других фрагментов с использованием метрики IoU (пересечения над объединением), позволяющей определить перекрытие фрагментов.

Параметр перекрытия фрагментов  $B_{i_0, j_0}$  и  $B_{i_1, j_1}$  вычисляется на основе коэффициента Жаккарда:

$$IoU = \frac{In}{B_{i_0, j_0}(w)B_{i_0, j_0}(h) + B_{i_1, j_1}(w)B_{i_1, j_1}(h) - In},$$

где

$$In = \left( \min(B_{i_0, j_0}(x) + B_{i_0, j_0}(w), B_{i_1, j_1}(x) + B_{i_1, j_1}(w)) - \max(B_{i_0, j_0}(x), B_{i_1, j_1}(x))) \right) \times \\ \times \left( \min(B_{i_0, j_0}(y) + B_{i_0, j_0}(h), B_{i_1, j_1}(y) + B_{i_1, j_1}(h)) - \max(B_{i_0, j_0}(y), B_{i_1, j_1}(y))) \right),$$

$w$  и  $h$  — ширина и высота фрагментов в координатах кадра;  $x$  и  $y$  — координаты левого верхнего угла фрагмента.

Основной недостаток данной модели — выделение множества фрагментов и их обработка, т. е. большие временные затраты, поэтому дальнейшее повышение эффективности модели предложено в Fast R-CNN [11] и Faster R-CNN [12]. СНС Fast R-CNN направлена на уменьшение использования числа регионов на изображении за счет применения слоя объединения регионов интереса (ROI pooling layer). Такой подход предполагает однократную обработку входного изображения и построение матрицы признаков, которая используется для сопоставления с характеристиками множества обнаруженных фрагментов. В данной модели вместо SVM-классификатора используется полносвязный SoftMax слой, что также позволяет снизить время обработки.

Дальнейшее повышение скорости обработки предложено за счет применения архитектуры Faster R-CNN, которая предполагает наличие двух этапов. На первом этапе на изображении определяются области, в которых предположительно расположены объекты, с применением отдельной предобученной глубокой СНС — RPN (Region Proposal Network). СНС RPN использует карту признаков последнего сверточного слоя для предсказания областей интереса, которые содержат искомые объекты, на основе обработки окном фиксированного размера карты признаков. Для каждого положения окна выделяются несколько предполагаемых областей разных размеров, которые называют якорями (традиционно их девять). Для каждой области формируются векторы признаков и выполняется бинарная классификация по принадлежности области к искомому объекту. На втором этапе используется детектор Fast R-CNN, который выполняет поиск объектов в предложенных регионах.

В работе [13] для распознавания радужной оболочки глаза в модель Faster R-CNN добавлен модуль FPN (Feature Pyramid Network), что позволило достаточно качественно распознавать радужную оболочку и периферические области глаза. Данный подход для обнаружения движущихся средств на снимках сверхвысокого разрешения может также дать хорошие результаты. FPN состоит из трех частей [14] — восходящий и нисходящий пути, а также боковые соединения (рис. 1). Восходящий путь описывается иерархической пирамидой, которая представляет собой последовательность сверточных слоев с уменьшающейся размерностью (backbone ConvNet). Верхние слои СНС имеют большее семантическое значение, но меньшее разрешение, для нижних слоев справедливо обратное утверждение. Однако эта часть неустойчива при извлечении признаков в некоторых случаях, например при наличии шума на исходном изображении, который может привести к значительному изменению признаков на верхнем уровне.

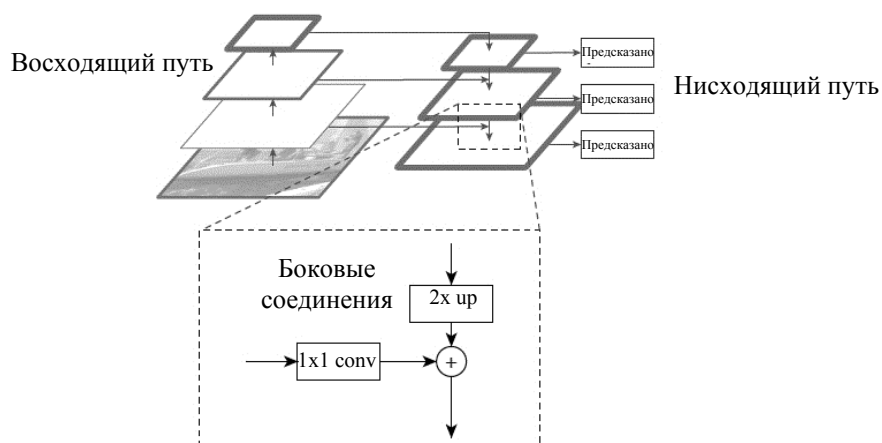


Рис. 1. Структура FPN

Нисходящий путь также описывается пирамидой. Размеры карт признаков верхнего слоя аналогичны размерам карт признаков верхнего слоя восходящего пути и далее увеличиваются по направлению вниз методом ближайшего соседа. Карты признаков соответствующих слоев двух пирамид поэлементно складываются, причем для карты восходящего пути предварительно применяется свертка. Боковые соединения предназначены для решения проблемы затухания признаков при проходе по слоям, включают в себя семантически важную информацию, полученную к концу первой пирамиды, и ранее полученную детальную информацию. Таким образом обеспечивается выделение сильных семантических признаков при различных разрешениях изображения. В архитектуре используется Faster R-CNN в сочетании с FPN, которая заменяет традиционную СНС в Faster R-CNN для извлечения признаков.

**Подготовка данных и обучение СНС.** При реализации модели СНС важен выбор фреймворка, обеспечивающего эффективное машинное обучение. Для поставленной задачи наиболее оптимальным является фреймворк ML.NET, представляющий собой кроссплатформенную среду машинного обучения с открытым исходным кодом для разработчиков .NET, обеспечивающий интеграцию кода, который поддерживает машинное обучение во многих продуктах Microsoft, включая Power BI, Windows Defender и Azure. Данный фреймворк после обучения нейронной сети сохраняет ее структуру в файл формата ONNX — распространенного формата хранения информации о моделях нейронных сетей, который поддерживается программой для визуализации моделей нейронных сетей Netron.

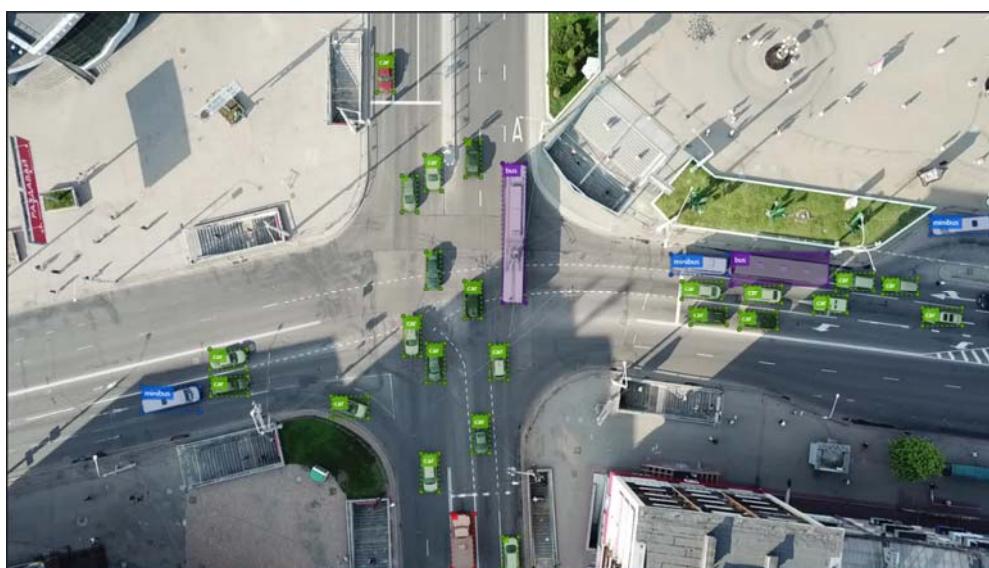


Рис. 2. Размеченное изображение в формате VoTT JSON

Для обучения используемой архитектуры СНС Faster R-CNN с FPN, реализованной на фреймворке ML.NET, использована база данных [15], которая включает в себя размеченные наборы данных, а также изображения, полученные с использованием БПЛА в г. Минске. Следует отметить, что формат разметки данных [15], который подготовлен как YOLO Darknet TXT, изменен в формат VoTT JSON. Пример размеченного изображения приведен на рис. 2.

Для обучения СНС использовано расширение ML.NET Model Builder, которое устанавливается в среду разработки Visual Studio. Реализация обучения СНС включает в себя создание нового проекта типа ML.NET Model Builder, выбор сценария и настройку окружения. Однако фреймворк ML.NET не поддерживает локальный способ обучения СНС для задачи распознавания объектов, поддерживается только обучение в облачной платформе Microsoft Azure. Поэтому для настройки тренировки в Azure ML необходимо сначала создать сценарий обучения. На первом шаге добавления данных используется размеченная база данных. Для этого указывается путь к файлу vottMainObject-export.json, который содержит информацию обо всех изображениях в наборе данных и соответствующей им разметке. После завершения данного этапа выполняется обучение модели. В зависимости от количества тренировочных данных, архитектуры нейронной сети, вычислительных мощностей и других факторов время тренировки колеблется от получаса до нескольких часов.

**Результаты и их обсуждение.** Для оценки качества результатов обнаружения объектов определенного класса базовыми величинами являются: количество правильных обнаружений, т. е. истинно положительные случаи (TP); отрицательных обнаружений как положительных, т. е. ложноположительные случаи (FP); правильных отрицательных обнаружений, т. е. истинно отрицательные случаи (TN); положительных обнаружений как отрицательных, т. е. ложноотрицательные случаи (FN). Количество правильных и ложных обнаружений определяется на основе применения метрики объединения над пересечением для найденных объектов и аннотированных на изображениях базы данных с пороговым уровнем  $T = 50\%$  для принятия решения. Далее с использованием полученных значений вычисляется доля положительных результатов среди всех обнаружений для данного класса объектов, т. е. определяется точность:

$$P = TP / (TP + FP),$$

которая изменяется в диапазоне 0—1, при этом не учитывается ситуация, когда используемый детектор не обнаруживает объекты, т. е. не учитываются пропуски объектов, соответствующих искомому классу.

Анализ таких случаев выполняется на основе вычисления полноты, характеризующей долю правильного обнаружения из всех объектов, которые детектор обнаружил для данного класса:

$$R = TP / (TP + FN).$$

Полнота также может принимать значения от 0 до 1, однако она не учитывает количество FP.

Случай полученной высокой точности с низкой полнотой свидетельствует не только о правильном обнаружении большинства объектов, но и о множестве ложных срабатываний. При высокой полноте и низкой точности ложные срабатывания минимальны, но ряд объектов искомого класса не обнаруживаются детектором. Соответственно, данные метрики являются взаимно дополняющими. Поэтому для оценки эффективности обнаружения объектов одного класса используют метрику — среднюю точность по классу (AP), определяемую на основе точности и полноты. С использованием  $P$  и  $R$  рассчитывается метрика AP для каждого класса для 11 пороговых уровней в диапазоне от 0 до 1 с шагом (0,1) по формуле [16]:

$$AP = \frac{1}{11} \sum_{i=0}^{10} P_{\text{int}}(R_i),$$

где  $P_{\text{int}}(R_i) = \max P(\tilde{R}), \tilde{R} \geq R_i$ .

Предложенный алгоритм обнаруживает несколько классов объектов, поэтому для оценки его эффективности вычисляется метрика mAP, предполагающая усреднение значений метрики AP по классам. Важный критерий при обнаружении объектов — точное определение их местоположения в координатах кадра, называемое точностью локализации объекта. Локализация объектов на статическом изображении позволяет определить не только координаты объектов, но и взаимное их расположение. Для этого рассчитывается погрешность локализации объектов как среднее между координатами обнаруженного объекта и аннотированного в используемой базе данных для всех объектов.



Рис. 3. Результаты обнаружения и классификации транспортных средств на изображениях, снятых под углом  $90^\circ$  (а) и  $70^\circ$  (б) в летний период; снятых под углом  $90^\circ$  в зимний период (в)

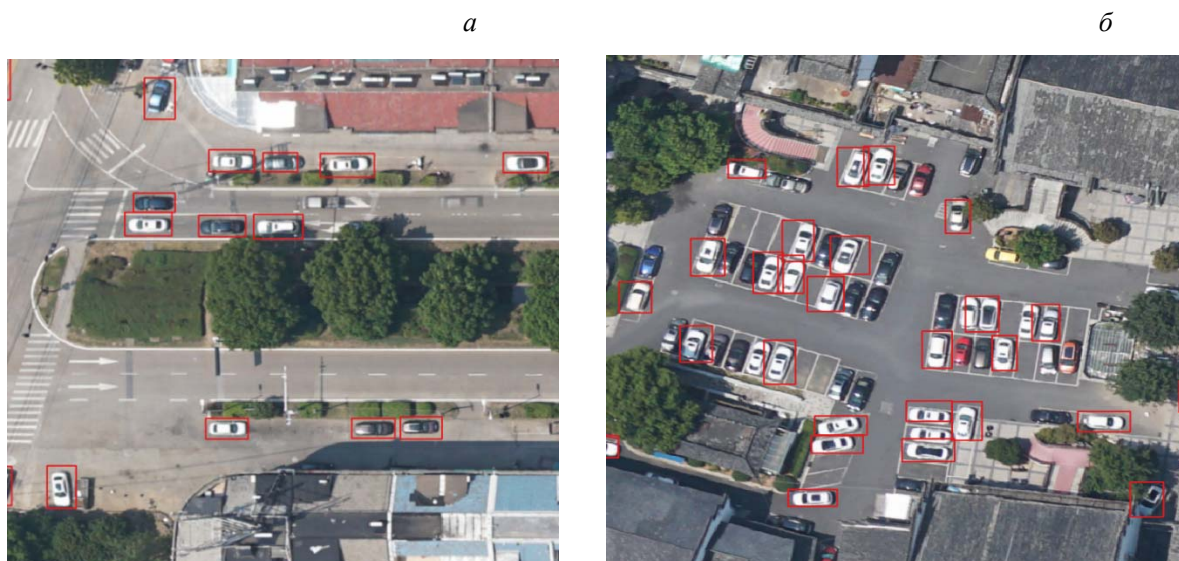
Тестирование результативности работы предложенного алгоритма выполнено на изображениях, полученных с разных БПЛА с разрешением 0.1—0.3 м. Результаты эффективности обнаружения и локализации объектов предлагаемого подхода с СНС YOLOv4 [17] и Faster R-CNN представлены в табл. 1. Точность обнаружения для используемого подхода на основе Faster R-CNN с FPN выше, значение метрики mAP больше, чем для СНС YOLO v4 и Faster R-CNN при обнаружении ТС на снимках сверхвысокого разрешения. Обеспечивается также повышение точности локализации объектов.

**Т а б л и ц а 1.** Сравнение эффективности обнаружения объектов

Метрика	YOLOv4	Faster R-CNN	Faster R-CNN с FPN
mAP	0.44917	0.62702	0.66534
Точность обнаружения, %	89	93	95
Точность классификации, %	99	99	99
Погрешность локализации, пиксель	12.3	8.4	7.5
Время обработки, мс	107	834	853

На рис. 3 представлены результаты обнаружения и классификации движущихся ТС на снимках сверхвысокого (0.2—0.4 м) разрешения с БПЛА. Полученные результаты подтверждают эффективность предложенного подхода, так как значение метрики точности mAP по сравнению с другими СНС выше, а средняя погрешность обнаружения ниже. К достоинствам также следует отнести правильное обнаружение объектов на изображениях, снятых под небольшим углом, и четкое разделение по классам легковой автомобиль/автобус/грузовик. Если движущиеся ТС являются темными на особенно темных участках изображений, то их трудно различать и нельзя исключать небольшие ошибки в данном подходе. Нейронная сеть может ошибиться в граничных условиях: если объект не полностью попадает на изображение или похож по цвету на фон, например мокрый асфальт или тень.

При очень высоком (0.1 м) разрешении автомобили скорее всего являются самыми маленькими объектами, которые можно обнаружить. В этом пространственном масштабе внешний вид объектов сильно зависит от спутниковой оптики и последующей обработки необработанных изображений. Из-за этого применение обученных моделей к обнаружению других объектов даже при сохранении разрешения представляет собой открытый вопрос, который требует дальнейшего исследования. Недостаток предложенного подхода — более низкая скорость обработки по сравнению с другими подходами, однако с учетом непрерывного совершенствования вычислительной техники этот недостаток несущественный.



**Рис. 4.** Результаты обнаружения легковых автомобилей на изображении с разрешением 0.1 м (а); белых автомобилей (б)

**Заключение.** Для повышения эффективности обнаружения и классификация объектов на снимках сверхвысокого разрешения, получаемых с беспилотных летательных аппаратов, предложено интегрировать сверточную нейронную сеть Faster R-CNN с модулем Feature Pyramid Network. На основе данного подхода разработан алгоритм, который реализован на языке программирования C++. Для обучения предложенной модели использована кроссплатформенная среда ML.NET. На используемой базе данных Faster R-CNN с модулем Feature Pyramid Network можно получить значение метрики mAP = 0.66534; средняя погрешность локализации объектов 7.5 пикселей. Сравнение эффективности предложенного подхода и сверточных нейронных сетей YOLO v4 и Faster R-CNN показывает повышение метрики mAP и уменьшение погрешности локализации для разных типов транспортных средств на снимках сверхвысокого разрешения. Представленная модель перспективна для решения практических задач по обнаружению транспортных средств на изображениях сверхвысокого разрешения.

Работа выполнена при частичной финансовой поддержке Национальной программы набора иностранных экспертов (гранты № G2021016001L, G2021016002L) и Базовой программы исследований общественного благосостояния провинции Чжэцзян (грант № LGJ19F020002).

- [1] **Yuntao Li, Zhihuan Wu, Lei Li, Daoning Yang, Hongfeng Pang.** *J. Appl. Remote Sens.*, **15**, N 2 (2021) 026505, doi: 10.1117/1.JRS.15.026505
- [2] **Q. Tan, J. Ling, J. Hu, X. Qin, J. Hu.** *IEEE Access*, **8** (2020) 153394—153402, doi: 10.1109/ACCESS.2020.3017894
- [3] **Y. Koga, H. Miyazaki, R. Shibasaki.** *Remote Sens.*, **12** (2020) 575, doi: 10.3390/rs12030575
- [4] **L. Yongxi, T. Javidi.** *Proc. 53rd Annual Allerton Conf. Communication, Control, and Computing* (2015) 1091—1098, doi: 10.1109/ALLERTON.2015.7447130
- [5] **W. Shao, W. Yang, G. Liu, J. Liu.** *Proc. IEEE Int. Geoscience and Remote Sensing Symposium* (2012) 4379—4382, doi: 10.1109/IGARSS.2012.6350403
- [6] **D. Vorobjov, I. Zakharova, R. Bohush, S. Ablameyko.** *Advances in Neural Networks-ISNN2018. Lecture Notes in Computer Science*, **10878** (2018) 503—510, doi: 10.1007/978-3-319-92537-0
- [7] **И. Н. Пугачев, Г. Я. Маркелов, В. С. Тормозов.** *Вестн. ТОГУ*, **45**, № 2 (2017) 13—20
- [8] **Xungen Li, Feifei Men, Shuaishuai Lv, Xiao Jiang, Mian Pan, Qi Ma, Haibin Yu.** *ISPRS Int. J. Geo-Information*, **10** (2021) 549, doi: 10.3390/ijgi10080549
- [9] **A. Froidevaux, A. Julier, A. Lifschitz, M. Pham, R. Dambreville, S. Lefèvre, P. Lassalle, T. Huynh.** *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, Sept. (2020) 256—259, doi: 10.1109/IGARSS39084.2020.9323827
- [10] **R. Girshick, J. Donahue, T. Darrell, J. Malik.** *IEEE Transact. Pattern Analysis and Machine Intelligence*, **38** (2016) 142—158
- [11] **R. Girshick.** *Proc. IEEE Int. Conf. Computer Vision. Santiago, Chile* (2015) 1440—1448
- [12] **S. Ren, K. He, R. Girshick, J. Sun.** *IEEE Transact. Pattern Analysis and Machine Intelligence*, **39** (2015) 1137—1149
- [13] **D. R. Lucio, R. Laroca, L. A. Zanlorensi, G. Moreira, D. Menotti.** *Region Detection Using Coarse Annotations*, doi: 10.1109/SIBGRAPI.2019.00032, arXiv:1908.00069
- [14] **T.-Yi Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie.** *Feature Pyramid Networks for Object Detection* arXiv:1612.03144v2 [cs.CV] 19 Apr 2017
- [15] <https://github.com/jekhor/aerial-cars-dataset>
- [16] **M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, A. Zisserman.** *Int. J. Computer Vision*, **88** (2010) 303—338, doi: 10.1007/s11263-009-0275-4
- [17] <https://arxiv.org/pdf/2004.10934.pdf>