

УДК: 004.93'12

Сверточные нейронные сети семейства YOLO для мобильных систем компьютерного зрения

С. Г. Небаба^а, Н. Г. Марков

Национальный исследовательский Томский политехнический университет,
Россия, 634050, г. Томск, пр. Ленина, д. 30

E-mail: ^а stepanlfx@tpu.ru

Получено 17.10.2023, после доработки — 19.02.2024.

Принято к публикации 19.02.2024.

Работа посвящена анализу известных классов моделей сверточных нейронных сетей и исследованию выбранных из них перспективных моделей для детектирования летающих объектов на изображениях. Под детектированием объектов (англ. — Object Detection) здесь понимаются обнаружение, локализация в пространстве и классификация летающих объектов. Комплексное исследование выбранных перспективных моделей сверточных нейронных сетей проводится с целью выявления наиболее эффективных из них для создания мобильных систем компьютерного зрения реального времени. Показано, что наиболее приемлемыми для детектирования летающих объектов на изображениях с учетом сформулированных требований к мобильным системам компьютерного зрения реального времени и, соответственно, к лежащим в их основе моделям сверточных нейронных сетей являются модели семейства YOLO, причем наиболее перспективными следует считать пять моделей из этого семейства: YOLOv4, YOLOv4-Tiny, YOLOv4-CSP, YOLOv7 и YOLOv7-Tiny. Для обучения, валидации и комплексного исследования этих моделей разработан соответствующий набор данных. Каждое размеченное изображение из набора данных включает от одного до нескольких летающих объектов четырех классов: «птица», «беспилотный летательный аппарат самолетного типа», «беспилотный летательный аппарат вертолетного типа» и «неизвестный объект» (объекты в воздушном пространстве, не входящие в первые три класса). Исследования показали, что все модели сверточных нейронных сетей по скорости детектирования объектов на изображении (по скорости вычисления модели) значительно превышают заданное пороговое значение, однако только модели YOLOv4-CSP и YOLOv7, причем только частично, удовлетворяют требованию по точности детектирования (классификации) летающих объектов. Наиболее сложным для детектирования классом объектов является класс «птица». При этом выявлено, что наиболее эффективной по точности классификации является модель YOLOv7, модель YOLOv4-CSP на втором месте. Обе модели рекомендованы к использованию в составе мобильной системы компьютерного зрения реального времени при условии увеличения в созданном наборе данных числа изображений с объектами класса «птица» и дообучения этих моделей с тем, чтобы они удовлетворяли требованию по точности детектирования летающих объектов каждого из четырех классов.

Ключевые слова: детектирование летающих объектов на изображениях, сверточная нейронная сеть, YOLO, мобильная система компьютерного зрения

UDC: 004.93'12

Convolutional neural networks of YOLO family for mobile computer vision systems

S. G. Nebaba^a, N. G. Markov

National Research Tomsk Polytechnic University,
30 Lenina ave., Tomsk, 634050, Russia

E-mail: ^a stepanlfx@tpu.ru

Received 17.10.2023, after completion – 19.02.2024.

Accepted for publication 19.02.2024.

The work analyzes known classes of convolutional neural network models and studies selected from them promising models for detecting flying objects in images. Object detection here refers to the detection, localization in space and classification of flying objects. The work conducts a comprehensive study of selected promising convolutional neural network models in order to identify the most effective ones from them for creating mobile real-time computer vision systems. It is shown that the most suitable models for detecting flying objects in images, taking into account the formulated requirements for mobile real-time computer vision systems, are models of the YOLO family, and five models from this family should be considered: YOLOv4, YOLOv4-Tiny, YOLOv4-CSP, YOLOv7 and YOLOv7-Tiny. An appropriate dataset has been developed for training, validation and comprehensive research of these models. Each labeled image of the dataset includes from one to several flying objects of four classes: “bird”, “aircraft-type unmanned aerial vehicle”, “helicopter-type unmanned aerial vehicle”, and “unknown object” (objects in airspace not included in the first three classes). Research has shown that all convolutional neural network models exceed the specified threshold value by the speed of detecting objects in the image, however, only the YOLOv4-CSP and YOLOv7 models partially satisfy the requirements of the accuracy of detection of flying objects. It was shown that most difficult object class to detect is the “bird” class. At the same time, it was revealed that the most effective model is YOLOv7, the YOLOv4-CSP model is in second place. Both models are recommended for use as part of a mobile real-time computer vision system with condition of additional training of these models on increased number of images with objects of the “bird” class so that they satisfy the requirement for the accuracy of detecting flying objects of each four classes.

Keywords: detection of flying objects in images, convolutional neural network, YOLO, mobile computer vision system

Citation: *Computer Research and Modeling*, 2024, vol. 16, no. 3, pp. 615–631 (Russian).

1. Введение

В настоящее время во всем мире интенсивно ведутся исследования и разработки современных систем компьютерного зрения (СКЗ) различного назначения. Такие системы сегодня востребованы для решения многих прикладных задач компьютерного зрения в самых разных сферах человеческой деятельности, таких как управление автономными транспортными средствами, контроль безопасности предприятий и опасных производственных объектов [Zoev, Markov, Ryzhova, 2019], обработка и анализ медицинских изображений и в ряде других областей деятельности [Alzubaidi et al., 2021]. Отдельного внимания заслуживает задача безопасности воздушного пространства, которая требует от СКЗ постоянного мониторинга пространства с целью поиска летающих объектов и обработки больших объемов визуальной информации в режиме реального времени [Клековкин, Марков, 2023]. При этом съемка таких объектов может осуществляться не только в оптическом диапазоне, но и в других диапазонах электромагнитного спектра. В этих случаях даже формирование репрезентативного набора данных в виде совокупности изображений может оказаться далеко не тривиальной задачей [Nebaba, Markov, 2022].

Особое практическое значение среди СКЗ для мониторинга воздушного пространства имеют мобильные (возимые и носимые, включая встраиваемые) СКЗ на основе современных моделей сверточных нейронных сетей (СНС) [Tang, Hua, Wang, 2017]. Вычислительные ресурсы мобильных СКЗ весьма ограничены, поэтому при реализации моделей СНС в таких СКЗ приходится жертвовать либо точностью распознавания летающих объектов на изображениях, либо скоростью вычисления моделей СНС [Николенко, Кадулин, Архангельская, 2018]. Это указывает на актуальность разработки и исследования новых моделей СНС и выбора из них тех, что отвечают требованиям по скорости вычислений и точности распознавания объектов на изображениях, выдвигаемым при проектировании конкретных мобильных СКЗ. Этим требованиям особенно непросто удовлетворить в случае создания мобильных СКЗ, функционирующих в режиме реального времени. Более того, даже изучение отдельных аспектов программной или аппаратной реализации современных моделей СНС в таких мобильных СКЗ может являться крупной исследовательской задачей [Hassan, Rahim, Shin, 2019].

Данная статья посвящена выбору из известных классов моделей СНС перспективных моделей для детектирования летающих объектов на изображениях и комплексному исследованию таких моделей с целью выявления наиболее эффективных из них для создания мобильных СКЗ реального времени. Под детектированием объектов (англ. — Object Detection) здесь и далее в статье понимаются обнаружение, локализация в пространстве и классификация летающих объектов.

2. Мобильные СКЗ на основе моделей СНС для детектирования летающих объектов

Задача объектного детектирования летающих объектов на изображениях является основной задачей, решаемой с помощью СКЗ при мониторинге воздушного пространства. Она сводится к трем подзадачам: обнаружение объектов на изображении, локализация их и определение класса каждого из объектов. Сегодня перспективным направлением при решении такой трехэтапной задачи детектирования является использование в СКЗ современных моделей СНС. Это означает, что при создании СКЗ для мониторинга воздушного пространства необходимо выбрать соответствующую модель СНС, решающую эту крупную задачу, реализовать ее программно или аппаратно в составе СКЗ и исследовать ее с учетом требований, предъявляемых к СКЗ и, соответственно, к реализуемой модели СНС. На изображениях и видео, получаемых при мониторинге воздушного пространства с помощью соответствующей аппаратуры в составе таких СКЗ и поступающих на вход реализованной в СКЗ модели СНС, могут появляться один и более летающих объектов, причем в случае нескольких объектов они могут принадлежать к разным

классам. Это необходимо учитывать при начальном выборе модели СНС из набора потенциально перспективных для решения задачи детектирования моделей СНС. Исходные изображения и видео в некоторых СКЗ перед решением задачи детектирования предварительно обрабатываются; в частности, из видео через определенные равные промежутки времени извлекаются отдельные изображения.

Известно, что противоречащим друг другу требованиям к высокой скорости вычислений моделей СНС и высокой точности распознавания объектов на изображениях с их помощью особенно непросто удовлетворить в случаях создания мобильных СКЗ, функционирующих в режиме реального времени [Hassan, Rahim, Shin, 2019; Zoev, Markov, Ryzhova, 2019]. Именно к таким СКЗ относятся мобильные СКЗ, предназначенные для детектирования летающих объектов в воздушном пространстве. Мобильные СКЗ реального времени должны не только обнаруживать и классифицировать летающие объекты, но и отслеживать их перемещение в пространстве (трекинг объектов), а также визуализировать их траектории. Эта вторая крупная задача, решаемая с помощью СКЗ, сводится к вычислению через заданные промежутки времени модели СНС при неоднократном решении подзадачи локализации объектов и в итоге к получению и нанесению на анализируемом изображении координат летающих объектов. При этом требуемый масштаб реального времени при функционировании мобильной СКЗ определяется скоростью перемещения таких объектов в воздушном пространстве и, соответственно, на анализируемом изображении. Как показано в работах [Markov, Zoev, Mytsko, 2022; Zoev, Markov, Ryzhova, 2019], основные вычислительные ресурсы мобильной СКЗ реального времени используются для вычисления модели СНС, программно или аппаратно реализованной в этой СКЗ. Это еще раз подчеркивает актуальность создания и использования в мобильных СКЗ реального времени наиболее эффективных моделей СНС, отвечающих основным требованиям к таким СКЗ.

Рассмотрим количественные характеристики требований к мобильной СКЗ и, соответственно, к лежащей в ее основе модели СНС. Точность детектирования (классификации) объектов на изображении обычно оценивается с использованием общепринятых метрик. Часто в исследованиях применяется метрика AP (англ. — Average Precision, средняя точность) — площадь под кривой точности/полноты (англ. — AUC, area under the precision recall (PR) curve), оцененная с использованием порогового значения IoU, метрики степени пересечения между двумя ограничивающими рамками (англ. — Intersection over Union) [Intersection over Union . . . , 2023]. Также используется метрика mAP (англ. — mean Average Precision) — среднее значение AP по всем классам объектов. Для метрик $AP_{0,5}$ и $mAP_{0,5}$ значение IoU равно 0,5. В соответствии с работой [Intersection over Union . . . , 2023] будем считать, что точность детектирования (классификации) летающих объектов на изображениях с помощью модели СНС высока (требование к точности детектирования выполняется), если значение метрики $AP_{0,5}$ для каждого класса объектов и усредненной метрики $mAP_{0,5}$ по всем классам превышает пороговое значение 0,8.

Другой количественной характеристикой мобильной СКЗ является скорость детектирования объектов на изображении (скорость вычисления модели СНС), измеряемая как число анализируемых изображений в секунду в FPS (англ. — Frames Per Second, кадров в секунду). Значение FPS, позволяющее считать, что мобильная СКЗ удовлетворяет требованию по скорости детектирования летающих объектов с помощью программной или аппаратной реализации обученной модели СНС, положим равным или более 25, исходя из минимальных технических возможностей подсистемы кодирования аналогового видеосигнала в системе PAL (англ. — Phase Alternating Line, построчное изменение фазы). Из соображений совместимости с этой подсистемой кодирования видеосигнала в большинстве практических применений до сих пор наиболее активно используется запись видеосигнала с частотой кадров в секунду, равной 25. Следует отметить, что пороговое значение FPS определяется масштабом реального времени, в котором работает мобильная СКЗ, и может быть выше 25 кадров в секунду. В свою очередь, масштаб реального

времени функционирования СКЗ определяет, на какой скорости перемещения можно обнаружить летающие объекты. Для многих мобильных СКЗ реального времени пороговые значения FPS зависят от особенностей решаемых прикладных задач, но чаще всего находятся в интервале 25–30 [Hassan, Rahim, Shin, 2019; Zoev, Markov, Ryzhova, 2019].

Третья характеристика — алгоритмическая эффективность модели СНС. При ее оценке рассчитываются два показателя: вычислительная сложность и размер модели СНС. Вычислительная сложность модели СНС в GFLOPs (англ. — Giga-FLoating-point OPerations per Second, миллиардов операций с плавающей точкой в секунду) — это количество необходимых операций умножения, сложения и сравнения над числами с плавающей точкой. В контексте задачи обучения и исследования моделей СНС для мобильных СКЗ используется расчет GFLOPs для 16-битных чисел (половинная точность) на графическом процессоре видеокарты NVIDIA Quadro RTX 6000 с архитектурой Turing, что позволяет сократить количество требуемых вычислительных ресурсов для обучения и использования моделей СНС [NVIDIA Docs. train with mixed precision, 2024]. Отметим, что во многих современных системах на кристалле с программируемыми логическими интегральными схемами (ПЛИС), часто применяемых в качестве вычислительного ядра в мобильных СКЗ, также используются операции с 16-битными числами с плавающей точкой. Вычислительная сложность модели СНС определяется числом операций умножения, сложения и сравнения над числами с плавающей точкой при выполнении процедур свертки и подвыборки во всех сверточных слоях и слоях подвыборки модели. Другой показатель — размер модели (компактность архитектуры) СНС в МБ (мегабайт) — это объем памяти вычислительного устройства мобильной СКЗ, необходимый для хранения весовых коэффициентов сверточных слоев и промежуточных буферов при вычислении модели СНС. В соответствии с результатами ряда исследователей [Hassan, Rahim, Shin, 2019; Zoev, Markov, Ryzhova, 2019], пороговые значения вычислительной сложности и размера моделей СНС задаются разработчиками мобильных СКЗ. Причем чаще всего они считают, что модель СНС-кандидат удовлетворяет их требованию по алгоритмической эффективности, если она может быть реализована на графическом процессоре среднего класса (например, от компании Nvidia) или на современной системе на кристалле с ПЛИС (например, от компании Xilinx).

Рассмотрим постановку задачи, решаемой в данном исследовании. Пусть на основе модели СНС создается мобильная СКЗ реального времени, которая должна позволять детектировать на изображениях летающие объекты следующих четырех классов:

- «птица»;
- «беспилотный летательный аппарат (БПЛА) самолетного типа»;
- «БПЛА вертолетного (мультироторного) типа»;
- «неизвестный объект».

К неизвестным объектам в контексте решаемой задачи будем относить объекты в воздушном пространстве, которые не входят в первые три класса. Среди таких объектов могут быть следующие: воздушные шары, вертолеты, самолеты и т. п.

Ставится задача выбора из известных моделей СНС наиболее перспективных моделей для детектирования летающих объектов на изображениях и комплексного исследования их с целью выявления наиболее эффективных из них для создания мобильных СКЗ реального времени. Такая эффективная модель (модели) СНС должна удовлетворять всем трем сформулированным выше требованиям к мобильным СКЗ реального времени и решать задачи детектирования не только одиночных летающих объектов, но и нескольких объектов на изображении, в том числе принадлежащих к различным классам.

3. Модели СНС для детектирования летающих объектов

Анализ показывает, что не все современные модели СНС отвечают указанным выше требованиям по скорости вычислений и точности детектирования объектов на изображениях. Например, широко используемая при детектировании объектов на изображениях модель R-CNN дает высокую точность детектирования, но имеет низкую скорость анализа изображений [Girshick et al., 2014]. Модель Faster R-CNN [Ren et al., 2015] и ее расширение Mask R-CNN [He et al., 2017] были разработаны для решения проблемы низкой скорости модели R-CNN, однако скорость их вычислений значительно ниже выбранного порогового значения FPS, равного 25. Для решения задачи детектирования летающих объектов рассматривались классическая модель СНС LeNet5 [Zoev, Markov, Ryzhova, 2019] и ее модификация [Клековкин, Марков, 2023], однако результаты их исследований показали низкую точность детектирования (классификации) изображений с несколькими объектами в воздушном пространстве. Это указывает на неприменимость моделей этого класса СНС для решения задачи, поставленной в данной работе.

Проведенный в [Alkentar et al., 2021; Wang, 2021] анализ различных классов моделей СНС и приведенные выше результаты нашего анализа ряда моделей СНС позволяют считать, что наиболее приемлемыми для детектирования летающих объектов на изображениях с учетом указанных выше требований являются модели семейства YOLO (англ. — You Only Look Once, «ты смотришь только один раз») [Redmon et al., 2016]. Модели этого семейства, как и другие модели СНС с одноступенчатой (однофазной) архитектурой, анализируют изображение за один проход (одномоментно выполняются этапы обнаружения, локализации и классификации объектов), что значительно повышает скорость анализа изображения. При этом такие модели СНС обладают более высокими показателями по точности детектирования объектов, чем, например, известная классическая модель SSD [Alkentar et al., 2021].

Часть моделей семейства YOLO имеет сложную архитектуру и поэтому может превысить пороговые значения по вычислительной сложности и по размеру занимаемой моделью СНС памяти. Не подходят для мобильных СКЗ с невысокими вычислительными ресурсами первые модели этого семейства, такие как YOLO, YOLOv2 и YOLOv3 [Redmon, Farhadi, 2018]. Иными словами, для мобильных СКЗ необходимы модели СНС с более компактными архитектурами, включающими в первую очередь небольшое число сверточных слоев. В связи с этим нами предлагается использовать в мобильных СКЗ и исследовать эффективность моделей СНС семейства YOLO с более современными архитектурами, начиная с модели YOLOv4 [Bochkovskiy, Wang, Liao, 2020]. Предлагается исследовать также модели этого семейства, известные как Tiny YOLO [Ma, Chen, Gao, 2017]. Tiny-модели, называемые далее компактными, обычно имеют небольшое число сверточных слоев и поэтому предъявляют значительно меньшие требования к памяти и производительности вычислительного устройства мобильной СКЗ. Однако точность детектирования (классификации) объектов на изображениях с помощью таких моделей обычно ниже, чем у соответствующих им полноразмерных моделей СНС. Практически все модели семейства YOLO имеют свою уменьшенную (компактную) версию.

Точность детектирования объектов на изображениях некоторых моделей СНС из семейства YOLO исследовалась рядом авторов на наборе данных MS COCO [Lin et al., 2014]. Известно, что на данный момент лучшую точность детектирования объектов на изображениях из этого набора данных показывают модели СНС YOLOv7 [Wang, Bochkovskiy, Liao, 2023]. Однако оригинальная модель YOLOv7 может предъявить весьма высокие требования к вычислительным ресурсам при ее потенциальной реализации в мобильной СКЗ реального времени, и поэтому наряду с исследованием модели YOLOv7 имеет смысл рассматривать ее компактную версию YOLOv7-Tiny.

С учетом приведенных выше требований среди моделей СНС семейства YOLO наиболее перспективными при решении поставленной задачи, на наш взгляд, являются приведенные ниже

пять моделей, созданных с использованием фреймворка Darknet [Redmon et al., 2016]. Кратко рассмотрим их особенности и потенциальные возможности по удовлетворению сформулированным требованиям.

1. СНС YOLOv4

В этой модели, в сравнении с более ранними моделями семейства, был добавлен особый слой — шея (англ. — neck), который является связующим между слоями основы и головы модели СНС [Bochkovskiy, Wang, Liao, 2020]. Он служит для сбора карт признаков с разных этапов сети и повышения скорости обработки изображений. Модель YOLOv4 имеет 137 сверточных слоев и поэтому имеет относительно небольшой размер и весьма невысокую вычислительную сложность (небольшое количество операций). То есть она, возможно, отвечает требованию по алгоритмической эффективности. Это позволяет считать ее потенциально отвечающей указанным выше требованиям, предъявляемым к моделям СНС со стороны мобильных СКЗ, и выбрать для дальнейших исследований.

2. СНС YOLOv4-Tiny

Эта сжатая версия модели СНС YOLOv4 содержит 29 сверточных слоев и около 6 миллионов весовых коэффициентов. Изначально авторами модель СНС YOLOv4 уже позиционировалась как модель для решения задачи детектирования объектов, ориентированная на использование в рамках встраиваемых и мобильных СКЗ. Однако для этого было предложено еще больше упростить ее структуру и уменьшить число параметров для повышения быстродействия. Так появилась упрощенная модель СНС YOLOv4-Tiny. Для обнаружения летающих объектов в реальном времени модель YOLOv4-Tiny является лучшим вариантом по сравнению с СНС YOLOv4, поскольку для мобильных СКЗ, функционирующих в реальном времени, малое время детектирования часто важнее, чем точность детектирования объектов.

Архитектура модели YOLOv4-Tiny имеет несколько изменений по сравнению с архитектурой исходной модели YOLOv4, что позволяет ей достичь высокой скорости вычислений. Так, помимо снижения числа сверточных слоев, стало меньше полей привязки для прогнозирования. В качестве слоя neck в архитектуре СНС YOLOv4-Tiny используется сочетание модулей SPP (англ. — Spartial Pyramid Pooling, объединение пространственных пирамид) и PANet (англ. — Path Aggregation Network, сеть агрегации путей) [Chen, Zhong, Zhang, 2021]. Модуль SPP объединяет карты признаков разного масштаба, выполняя несколько операций субдискретизации с различным размером окна (1×1 , 5×5 , 9×9 и 13×13) и используется для расширения области восприятия и выделения наиболее значимых семантических признаков. При этом модуль SPP практически не влияет на скорость выполнения модели СНС. Вместо модуля FPN (англ. — Feature Pyramid Network, функциональная пирамидальная сеть), использовавшегося в YOLOv3, в YOLOv4-Tiny для агрегации признаков применяется модуль PANet. В модели СНС YOLOv4-Tiny также используются модули SAM (англ. — Segment Anything Model, сегмент «любая модель») [Sindhu, 2021]. В модифицированном модуле PANet сложение карт признаков при пробросе заменено конкатенацией, а также модифицирован модуль SAM.

3. СНС YOLOv4-CSP

СНС YOLOv4-CSP — модификация модели YOLOv4, содержащая 142 сверточных слоя. Эта модель отличается от базовой модели YOLOv4 наличием механизма проброса признаков (англ. — Cross Stage Partial, CSP, частичный проброс признаков) между слоями для снижения сложности модели СНС [Wang, Bochkovskiy, Liao, 2021]. Механизм проброса признаков в общем виде реализуется следующим образом: выход из слоя BottleneckCSP («бутылочное горлышко», или узкое место CSP) конкатенируется с выходом другого сверточного слоя. В модели YOLOv4 реализован тип слоев C3. В отличие от слоев BottleneckCSP слои C3 содержат на один

сверточный слой меньше и не содержат нормализации. Для получения модели СНС YOLOv4-CSP была модифицирована архитектура модели YOLOv4 и применены вместо слоев C3 слои BottleneckCSP. Наличие дополнительных соединений наборов карт признаков и дополнительного слоя BottleneckCSP у новой модели должно повысить детальность результирующей карты классификации и, соответственно, повысить точность детектирования объектов малых размеров на изображениях. Последнее немаловажно в случае летающих объектов малых размеров.

4. СНС YOLOv7

Это наиболее актуальная на данный момент модель СНС семейства YOLO, содержащая 132 сверточных слоя [Wang, Bochkovskiy, Liao, 2023]. Отличается от предыдущих моделей СНС этого семейства рядом оптимизаций (таких как расширенное эффективное объединение слоев, англ. — Extended Efficient Layer Aggregation Networks, E-ELAN), что позволяет добиться большего быстродействия при вычислении модели при той же точности классификации объектов на изображении. Среди основных изменений, сделанных в этой модели, отметим следующие.

а. Расширенная агрегация эффективных уровней

Эффективность сверточных слоев моделей семейства YOLO в backbone-части (магистральной, или основной, части) имеет важное значение для скорости их вычислений. Авторы модели YOLOv7 начали ее модификацию с достижения максимальной эффективности уровней с помощью Cross Stage Partial Networks. При этом они основывались на исследованиях, проведенных по этой теме, с учетом объема памяти, необходимого для хранения слоев в памяти, а также расстояния, которое требуется градиенту для обратного распространения ошибки по слоям. Чем короче градиент, тем быстрее сеть может обучаться. Последний уровень агрегации — это E-ELAN, расширенная версия вычислительного блока ELAN.

б. Методы масштабирования модели

Модели обнаружения объектов обычно выпускаются в виде серии моделей для разного размера входа сети, так как для разных задач требуются разные уровни точности и скорости вывода. Как правило, модели обнаружения объектов учитывают глубину сети, ширину сети и разрешение, на котором обучается сеть. В модели YOLOv7 согласованно масштабируются глубина и ширина сети, объединяя слои вместе. Исследования абляции показывают, что этот метод поддерживает оптимальную архитектуру модели при масштабировании для разных размеров объектов.

в. Вспомогательный блок head

Head-часть (головная часть) известных моделей семейства YOLO делает окончательные прогнозы для них, но, поскольку она находится далеко от входа, может быть выгодно добавить вспомогательную head-часть в модель, которая находится где-то посередине. Вспомогательный блок head обучается не так эффективно, как основной, потому что между ним и предсказанием меньше слоев, поэтому авторы модели YOLOv7 экспериментируют с различными уровнями наблюдения для этого блока.

г. Планирование повторной параметризации

Методы повторной параметризации включают усреднение набора весов модели СНС для создания модели, более устойчивой к общим шаблонам, которые она пытается смоделировать. В исследованиях в последнее время основное внимание уделялось повторной параметризации (перепараметризации) на уровне модулей, когда часть моделей СНС имеет свои собственные стратегии перепараметризации. Поэтому авторы используют пути распространения градиентного потока, чтобы увидеть, какие модули в модели YOLOv7 должны использовать стратегии повторной параметризации, а какие нет.

5. СНС YOLOv7-Tiny

Эта облегченная версия модели СНС YOLOv7 содержит 87 сверточных слоев, около 6 миллионов параметров и ряд оптимизаций. В отличие от других версий модели YOLOv7 в ней используется функция активации Leaky ReLU (англ. — Leaky Rectified Linear Unit, выпрямленный линейный блок с уклоном) [Maas, Hannun, Ng, 2013], в то время как в других версиях модели СНС YOLOv7 используется функция активации SiLU (англ. — Sigmoid-weighted Linear Unit, сигмовидно-взвешенный линейный блок) [Elfwing, Uchibe, Doya, 2018]. Несмотря на то что функция SiLU обладает рядом преимуществ перед функцией Leaky ReLU, например учетом нелинейной составляющей функции, она сложнее в вычислении, поэтому при проектировании компактных моделей СНС часто отказываются от ее использования в пользу функции Leaky ReLU.

Выбранные пять моделей СНС из семейства YOLO подлежат дальнейшему исследованию. В таблице 1 приведены некоторые характеристики этих моделей СНС, полученные для изображений из набора данных MS COCO в работах [Wang, Bochkovskiy, Liao, 2023; Wang, Bochkovskiy, Liao, 2021]. Из таблицы 1 следует, что наименьшее число весовых коэффициентов и, соответственно, наименьшие размеры имеют компактные модели СНС YOLOv4-Tiny и YOLOv7-Tiny. Эти же модели обладают малой вычислительной сложностью. Все это позволяет считать, что они со значительным превышением удовлетворяют требованию максимальной алгоритмической эффективности и подходят для реализации в мобильных СКЗ с весьма малыми вычислительными ресурсами. Полноразмерные модели СНС также удовлетворяют этому требованию, и их несложно будет реализовать в мобильных СКЗ с более высокими, но практически приемлемыми вычислительными ресурсами. Все это дает основание не проводить дополнительные исследования выбранных пяти моделей СНС на предмет соответствия требованию их максимальной алгоритмической эффективности.

В то же время ни одна из пяти моделей СНС на наборе данных MS COCO не дает точность детектирования (классификации) объектов, превышающую заданное пороговое значение 0,8. Более того, для сравнения выбранных нами моделей СНС из семейства YOLO по точности детектирования объектов необходимо решать конкретную прикладную задачу детектирования летающих объектов заданных выше классов и исследование проводить на наборе данных, сформированном из изображений таких объектов. При этом следует создавать набор данных и обучать модели СНС таким образом, чтобы они давали точность детектирования, превышающую этот порог. Также необходимы исследования эффективности моделей СНС по скорости детектирования летающих объектов на изображении, то есть следует оценить скорость вычисления каждой из пяти моделей СНС. Последнее крайне актуально при проектировании мобильных СКЗ реального времени.

Таблица 1. Основные характеристики современных моделей СНС семейства YOLO

Модель СНС	Число весовых коэффициентов модели СНС, млн	GFLOPs	mAP _{0,5} (на валидационной выборке)
YOLOv4	64,4	142,8	0,682
YOLOv4-CSP	52,9	120,4	0,686
YOLOv4-Tiny	6,1	6,9	0,421
YOLOv7	36,9	104,7	0,697
YOLOv7-Tiny	6,2	3,5	0,473

Итак, результаты выбора наиболее перспективных моделей СНС из семейства YOLO и анализ их характеристик, приведенных в таблице 1, получены на первом этапе решения поставленной задачи. На втором этапе решения задачи необходимо провести обучение, валидацию и комплексное исследование выбранных пяти моделей СНС из семейства YOLO при детектировании

летающих объектов четырех классов на изображениях. Для этого сначала следует сформировать соответствующий набор данных. Исследования должны позволить выбрать наиболее эффективную модель (модели) для создания мобильных СКЗ реального времени. Обучение, валидацию и исследование этих пяти моделей следует провести с использованием фреймворка Darknet, на котором изначально и разрабатывались выбранные модели СНС семейства YOLO [Redmon et al., 2016].

4. Формирование набора данных

Для обучения и валидации выбранных моделей СНС семейства YOLO, а также для их исследования (тестирования) был подготовлен набор данных с изображениями четырех классов летающих объектов: «птица», «БПЛА квадрокоптер» (мультироторный), «БПЛА самолетного типа» и «неизвестный объект». Он создавался на основе большого количества оптических изображений летающих объектов выбранных классов. Размер каждого изображения — 416×416 пикселей. Формирование набора данных происходило за счет объединения размеченных другими авторами наборов изображений (сторонних наборов данных) и ручного сбора и разметки изображений, удовлетворяющих требованиям, изложенным выше. Изображения в наборе данных сформированы таким образом, чтобы на них присутствовали как одиночные объекты, так и несколько объектов, причем не только одного класса. Объем полученного набора данных оказался недостаточно большим для получения полноценных обучающих, валидационных и тестовых выборок. Поэтому для формирования полноценного по объему набора данных использовался подход, суть которого заключается в проведении аугментации уже размеченных изображений, входящих в набор данных. При этом применялись различные известные методы аугментации (незначительные изменения масштаба, отражения по вертикальной оси и так далее). Для изображений с летающими объектами малых размеров (размер объекта — не более 32×32 пикселей) выполнялся процесс копирования и вставки таких объектов на другие изображения набора данных в случайные места таким образом, чтобы они не перекрывали уже имеющиеся на изображении объекты.

Полученный набор данных включает 6230 изображений, содержащих 33 810 объектов, и столько же файлов аннотаций. Каждое изображение набора данных содержит один или более объектов одного или разных классов. В качестве примера на рис. 1 представлены четыре изображения из набора данных.

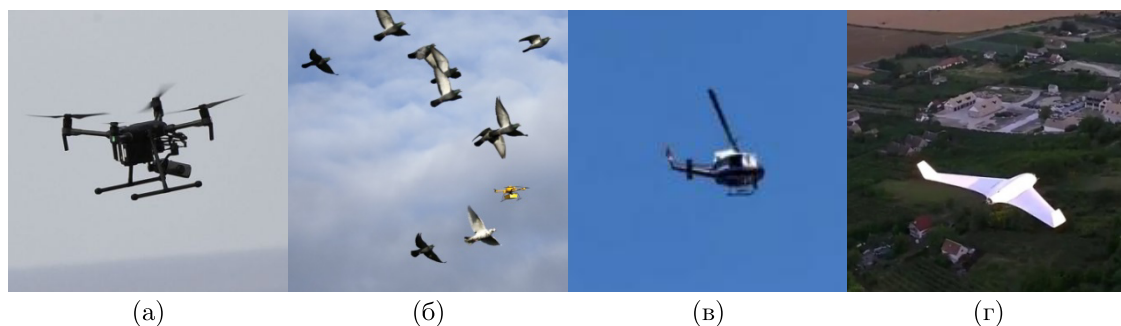


Рис. 1. Примеры изображений из набора данных с разным числом и классом объектов: а) один объект класса «БПЛА квадрокоптер»; б) несколько объектов класса «птица» и один объект класса «БПЛА квадрокоптер»; в) один объект класса «неизвестный объект» (вертолет, не относится к первым трем классам объектов); г) один объект класса «БПЛА самолетного типа»

Количественные характеристики изображений по классам объектов и по числу объектов на изображениях в этом наборе данных представлены ниже.

Общее число объектов в наборе данных по классам:

- «птица» — 5642;
- «БПЛА квадрокоптер» — 4845;
- «БПЛА самолетного типа» — 3514;
- «неизвестный объект» — 19 809.

Количество изображений в наборе данных по классам объектов:

- «птица» — 1912;
- «БПЛА квадрокоптер» — 2676;
- «БПЛА самолетного типа» — 2212;
- «неизвестный объект» — 4161.

Набор данных был разделен на следующие выборки:

- обучающая — 4358 изображений (70 % объема набора данных);
- проверочная (валидационная) — 1249 изображений (20 %);
- тестовая — 623 изображения (10 %).

Количество экземпляров объектов в наборе данных больше в связи с тем, что на одном изображении может быть несколько объектов одного или нескольких классов. На рис. 2 представлены столбчатые диаграммы распределения числа изображений в наборе данных по количеству объектов конкретного класса на одном изображении. По оси Y приведено количество изображений каждого класса в наборе данных с соответствующим числом объектов, а по оси X — число объектов на отдельном изображении.

5. Обучение, валидация и исследование выбранных моделей СНС

Для обучения, валидации и тестирования выбранных пяти моделей СНС из семейства YOLO использовался фреймворк Darknet [Redmon et al., 2016]. В качестве основы при обучении моделей были взяты весовые коэффициенты предобученных моделей семейства YOLO, описанных в работах [Wang, Bochkovskiy, Liao, 2021; Wang, Bochkovskiy, Liao, 2023; Wang, 2021]. При обучении, валидации и тестировании моделей СНС использовался компьютер с характеристиками: процессор Intel Core i9-11900KF, 64 ГБ ОЗУ, видеокарта NVIDIA Quadro RTX 6000 с объемом видеопамяти 24 ГБ. Обучение моделей проводилось со следующими параметрами:

- количество эпох (epoch) — 200;
- доверительный порог, conf_thres — 0,25;
- порог IoU, iou_thres — 0,5;
- размер входного изображения — 416 × 416 пикселей.

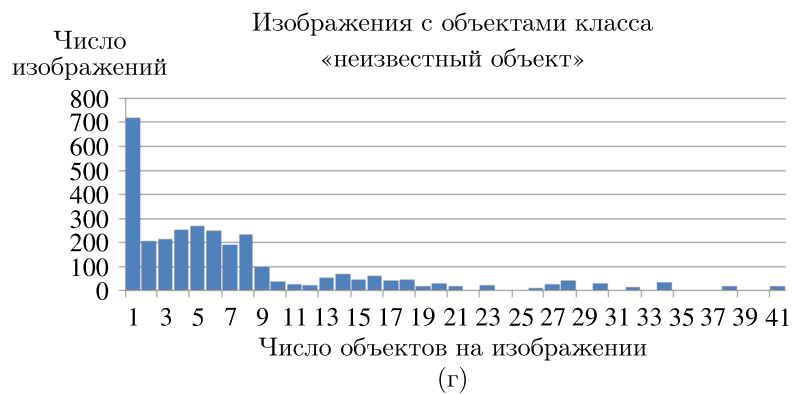
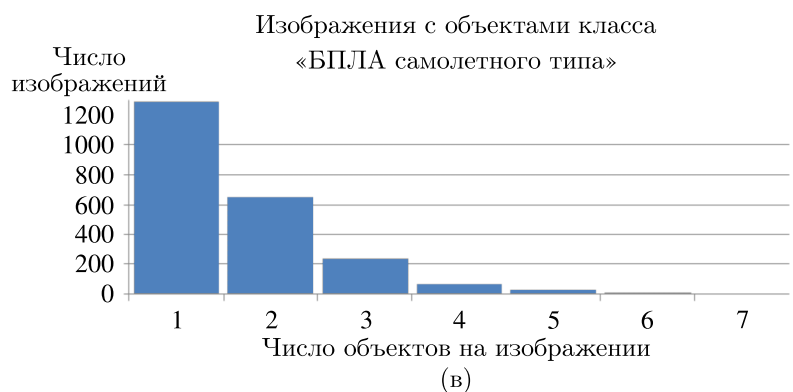
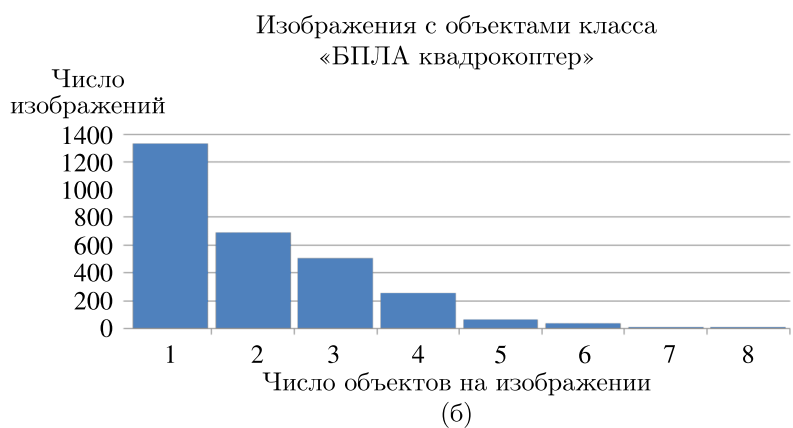


Рис. 2. Распределение числа изображений в наборе данных по количеству объектов конкретного класса на отдельном изображении: (а) для объектов класса «птица»; (б) для объектов класса «БПЛА квадрокоптер»; (в) для объектов класса «БПЛА самолетного типа»; (г) для объектов класса «неизвестный объект»

Таблица 2. Точность детектирования (классификации) объектов на изображениях валидационной выборки

Модель СНС	«Птица», $AP_{0,5}$	«БПЛА квадрокоптер», $AP_{0,5}$	«БПЛА самолетного типа», $AP_{0,5}$	«Неизвестный объект», $AP_{0,5}$	$mAP_{0,5}$
YOLOv4	0,679	0,785	0,810	0,727	0,750
YOLOv4-Tiny	0,584	0,776	0,826	0,666	0,713
YOLOv4-CSP	0,726	0,877	0,897	0,848	0,837
YOLOv7	0,767	0,888	0,903	0,847	0,861
YOLOv7-Tiny	0,685	0,828	0,868	0,781	0,791

Результаты обучения, поверенные на валидационной выборке объемом 1249 изображений, в виде значений метрики $AP_{0,5}$ по каждому классу летающих объектов и усредненной метрики $mAP_{0,5}$ по всем классам, представлены в таблице 2. Из нее следует, что наиболее высокую точность детектирования по метрикам $AP_{0,5}$ и $mAP_{0,5}$ показывают модели СНС YOLOv7 и YOLOv4-CSP.

Комплексные исследования пяти моделей СНС проводились в виде двух серий экспериментов с обученными и валидированными моделями СНС. Для этого использовалась тестовая выборка из набора данных объемом 623 изображения. В первой серии исследовалась точность детектирования (классификации) летающих объектов на изображениях с помощью каждой из моделей СНС, а во второй серии — скорость вычисления каждой из них.

В первой серии экспериментов результаты, полученные на тестовой выборке, сравнивались с результатами, полученными на валидационной выборке, а также с исходной предпосылкой о том, что точность детектирования по метрикам $AP_{0,5}$ по каждому классу объектов и усредненной метрики $mAP_{0,5}$ должна превышать значение 0,8. Результаты первой серии экспериментов в виде значений метрики $AP_{0,5}$ по каждому классу объектов и усредненной метрики $mAP_{0,5}$ по всем классам на тестовой выборке представлены в таблице 3.

Таблица 3. Точность детектирования (классификации) объектов на изображениях тестовой выборки

Модель СНС	«Птица», $AP_{0,5}$	«БПЛА квадрокоптер», $AP_{0,5}$	«БПЛА самолетного типа», $AP_{0,5}$	«Неизвестный объект», $AP_{0,5}$	$mAP_{0,5}$
YOLOv4	0,630	0,759	0,785	0,691	0,716
YOLOv4-Tiny	0,564	0,754	0,799	0,650	0,691
YOLOv4-CSP	0,724	0,845	0,897	0,858	0,831
YOLOv7	0,784	0,881	0,896	0,877	0,859
YOLOv7-Tiny	0,649	0,804	0,856	0,777	0,771

Из таблиц 2 и 3 следует, что значения метрики $mAP_{0,5}$ для всех моделей СНС на валидационной выборке выше, чем для тестовой, что указывает на правильность подготовки набора данных и обучения моделей СНС.

Результаты второй серии экспериментов по оценке скорости вычисления каждой из исследуемых моделей СНС представлены в таблице 4. Во втором столбце приведены значения времени обработки одного изображения каждой из моделей СНС (англ. — inference), в третьем столбце — значения общего времени на анализ одного изображения с помощью каждой из моделей СНС (включая этапы предобработки и постобработки изображений), а в четвертом столбце — значения FPS, рассчитанные по значениям общего времени из третьего столбца.

Таблица 4. Результаты оценки скорости вычисления моделей СНС

Модель СНС	Обработка изображения, мс	Общее время, мс	FPS
YOLOv4	30,2	33,5	29,9
YOLOv4-Tiny	5,3	5,4	185,2
YOLOv4-CSP	21,4	22,9	43,7
YOLOv7	20,8	22,1	45,2
YOLOv7-Tiny	10,8	11,0	90,9

6. Обсуждение полученных результатов

Из таблиц 2 и 3 следует, что средняя точность детектирования (классификации) по усредненной по всем классам летающих объектов метрике $mAP_{0,5}$ для моделей СНС YOLOv4-CSP и YOLOv7 превышает пороговое значение 0,8. Аналогичный результат для этих же моделей СНС имеет место по метрике $AP_{0,5}$ для объектов каждого из классов, кроме класса «птица». Невысокий результат для объектов класса «птица» объясняется меньшим числом изображений с объектами этого класса, чем число изображений объектов других классов в обучающей выборке. Более того, на этих изображениях находится большое число птиц малых размеров, что затрудняет их детектирование. Все это означает, что модели СНС YOLOv4-CSP и YOLOv7 частично удовлетворяют требованию по точности детектирования (классификации), предъявляемому к моделям СНС для мобильных СКЗ. Модель СНС YOLOv7-Tiny достаточно близка к пороговому значению по точности детектирования по метрике $mAP_{0,5}$ и не превышает пороговое значение по метрике $AP_{0,5}$ для объектов классов «птица» и «неизвестный объект». Модели СНС YOLOv4 и YOLOv4-Tiny как по метрике $mAP_{0,5}$, так и по метрике $AP_{0,5}$ для объектов каждого из классов показывают более низкую точность детектирования, чем у упомянутых трех моделей.

Из таблицы 4 следует, что все модели СНС по скорости детектирования объектов на изображении (по скорости вычисления модели) значительно превышают заданное пороговое значение FPS, равное 25. Наиболее значительное превышение порога характерно для компактных моделей. Это позволяет считать, что все исследуемые модели СНС удовлетворяют требованию к ним по скорости вычислений и могут быть использованы в мобильных СКЗ реального времени.

Учитывая все анализируемые результаты исследований пяти перспективных моделей СНС семейства YOLO, а также показанную выше возможность этих моделей удовлетворить требованию по алгоритмической эффективности при создании мобильных СКЗ, можно сформулировать два следующих вывода. Во-первых, наиболее эффективной моделью СНС среди пяти рассмотренных является модель YOLOv7, второе место занимает модель YOLOv4-CSP. Во-вторых, обе эти модели СНС следует рекомендовать к реализации в составе мобильной СКЗ реального времени, используемой для решения задачи детектирования летающих объектов четырех классов. Однако дополнительно следует увеличить в созданном наборе данных число размеченных изображений с объектами класса «птица» и дообучить эти модели, чтобы полностью удовлетворить требованию по точности детектирования летающих объектов всех четырех классов. Отметим, что полученные результаты комплексных исследований пяти моделей СНС семейства YOLO являются новыми.

В работах [Zoev et al., 2024; Markov, Zoev, Mytsko, 2022; Zoev, Markov, Ryzhova, 2019] приведены результаты исследований мобильных СКЗ при аппаратной реализации в них моделей СНС семейства YOLO и класса U-Net в различных системах на кристалле с ПЛИС компаний Altera и Xilinx. Из сравнения этих результатов с приведенными здесь результатами комплексных исследований моделей YOLOv7 и YOLOv4-CSP можно сделать вывод о принципиальной возможности использования этих моделей для аппаратной реализации в мобильной СКЗ реального времени для детектирования летающих объектов. Однако в качестве ее вычислительного

ядра должна быть система на кристалле с ПЛИС с характеристиками не хуже, чем у системы Zynq 7000 (Kintex FPGA) компании Xilinx. При этом такая СКЗ должна будет удовлетворять всем сформулированным выше требованиям. Конечно, окончательный вывод о возможности использования каждой из этих моделей СНС в той или иной конкретной мобильной СКЗ реального времени должен быть сделан только по результатам ее тестирования, что является предметом дальнейших исследований.

7. Заключение

Сегодня особое практическое значение имеют мобильные СКЗ на основе современных моделей СНС. При создании таких СКЗ актуальны разработка и исследование новых моделей СНС, отвечающих жестким требованиям как по скорости их вычисления, так и по точности детектирования (классификации) объектов на изображениях в условиях ограниченных вычислительных ресурсов, доступных в мобильных СКЗ. При этом одной из важных решаемых задач является выбор наиболее перспективных моделей СНС для детектирования летающих объектов на изображениях и комплексное исследование их с целью выявления наиболее эффективных для создания мобильных СКЗ реального времени.

Краткий анализ различных классов моделей СНС позволил считать, что наиболее приемлемыми для детектирования летающих объектов на изображениях с учетом сформулированных требований к мобильным СКЗ реального времени и, соответственно, к лежащим в их основе моделям СНС являются модели семейства YOLO. Анализ потенциальных возможностей моделей из этого семейства для удовлетворения этим требованиям выявил пять наиболее перспективных моделей СНС: YOLOv4, YOLOv4-Tiny, YOLOv4-CSP, YOLOv7 и YOLOv7-Tiny. Отметим, что результаты анализа основных классов моделей СНС и выбора этих пяти моделей для детектирования летающих объектов на изображениях с учетом сформулированных требований являются новыми. Для обучения, валидации и комплексного исследования этих моделей СНС для детектирования летающих объектов четырех классов на изображениях был разработан соответствующий набор данных. По итогам исследований обученных и валидированных моделей СНС на тестовой выборке этого набора получены следующие новые результаты и сделаны выводы.

Только модели СНС YOLOv4-CSP и YOLOv7 частично удовлетворяют требованию по точности детектирования (классификации) летающих объектов. При этом у них точность детектирования по метрике $mAP_{0,5}$ превышает пороговое значение 0,8, а по метрике $AP_{0,5}$ превышает это пороговое значение для объектов каждого из классов, кроме класса «птица». Все исследуемые модели СНС по скорости детектирования объектов на изображении (по скорости вычисления модели) значительно превышают заданное пороговое значение FPS, равное 25. Это позволяет считать, что все они удовлетворяют требованию по скорости вычислений и могут быть использованы в мобильных СКЗ реального времени. Учитывая все эти результаты, можно сделать следующие выводы. Наиболее эффективной моделью СНС среди пяти исследуемых является модель YOLOv7, а на втором месте находится модель YOLOv4-CSP. Обе модели СНС можно рекомендовать к реализации в составе мобильной СКЗ реального времени, используемой для решения задачи детектирования летающих объектов четырех классов, при условии увеличения в созданном наборе данных числа размеченных изображений с объектами класса «птица» и дообучения этих моделей, с тем чтобы они полностью удовлетворяли требованию по точности детектирования таких объектов.

Список литературы (References)

- Клековкин В. А., Марков Н. Г. Модель сверточной нейронной сети LeNet5 для обнаружения и классификации объектов воздушного пространства на изображениях // Известия Томского политехнического университета. Промышленная кибернетика. — 2023. — Т. 1, № 2. — С. 11–16.

- Klekovkin V.A., Markov N.G.* Model' svertochnoi neuronnoi seti LeNet5 dlya obnaruzheniya i klassifikatsii ob'ektov vozdušnogo prostranstva na izobrazheniyakh [Model of convolutional neural network LeNet5 for detection and classification of air space objects in images] // Bulletin of the Tomsk Polytechnic University. Industrial cybernetics. — 2023. — Vol. 1, No. 2. — P. 11–16 (in Russian).
- Николенко С. И., Кадурич А. А., Архангельская Е. О.* Глубокое обучение. Погружение в мир нейронных сетей. — СПб.: Питер, 2018.
- Nikolenko S.I., Kadurin A.A., Arkhangelskaya E.V.* Glubokoe obuchenie. Pogruzhenie v mir nejronnyh setej [Deep learning. Dive into the world of neural networks]. — Saint Petersburg: Piter, 2018 (in Russian).
- Alkentar S.M., Alsahwa B., Assalem A., Karakolla D.* Practical comparison of the accuracy and speed of YOLO, SSD and Faster RCNN for drone detection // Journal of Engineering. — 2021. — Vol. 27, No. 8. — P. 19–31.
- Alzubaidi L., Zhang J., Humaidi A.J., Al-Dujaili A., Duan Y., Al-Shamma O., Santamaría J., Fadhel M.A., Al-Amidie M., Farhan L.* Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions // Journal of big Data. — 2021. — Vol. 8, No. 1. — P. 53.
- Bochkovskiy A., Wang C.Y., Liao H.Y.M.* YOLOv4: Optimal speed and accuracy of object detection // arXiv preprint. — 2020. — arXiv:2004.10934
- Chen W., Zhong X., Zhang J.* Optimization research and defect object detection of aeroengine blade boss based on YOLOv4 // Journal of Physics: Conference Series. — 2021. — Vol. 1746. — ID: 012076.
- Elfwing S., Uchibe E., Doya K.* Sigmoid-weighted linear units for neural network function approximation in reinforcement learning // Neural networks. — 2018. — Vol. 107. — P. 3–11.
- Girshick R., Donahue J., Darrell T., Malik J.* Rich feature hierarchies for accurate object detection and semantic segmentation // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2014. — P. 580–587.
- Hassan S.A., Rahim T., Shin S.Y.* Real-time uav detection based on deep learning network // 2019 International Conference on Information and Communication Technology Convergence (ICTC), IEEE. — 2019. — P. 630–632.
- He K., Gkioxari G., Dollár P., Girshick R.* Mask R-CNN // Proceedings of the IEEE international conference on computer vision. — 2017. — P. 2980–2988.
- Intersection over Union (IoU) in object detection & segmentation. — [Electronic resource]. — <https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation> (accessed: 17.10.2023).
- Lin T.-Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., Perona P., Ramanan D., Zitnick C.L., Dollár P.* Microsoft COCO: common objects in context // Proceedings of Computer Vision – ECCV. — 2014. — Vol. 13. — P. 740–755.
- Ma J., Chen L., Gao Z.* Hardware implementation and optimization of tiny-YOLO network // Digital TV and Wireless Multimedia Communication: 14th International Forum, IFTC. — 2017. — Vol. 14. — P. 224–234.
- Maas A.L., Hannun A.Y., Ng A.Y.* Rectifier nonlinearities improve neural network acoustic models // Proceedings of the 30th International Conference on Machine Learning. — 2013. — Vol. 30, No. 1. — ID: 16489696.
- Markov N.G., Zhev I.V., Mytsko E.A.* FPGA hardware implementation of the YOLO subclass convolutional neural network model in computer vision systems // Proceedings of the 2022 International Siberian Conference on Control and Communication (SIBCON 2022), IEEE. — 2022. — P. 1–5.
- Nebaba S.G., Markov N.G.* Effectiveness of moving objects detecting and tracking in airspace by images in near-infrared // Light & Engineering. — 2022. — Vol. 30. — P. 62–69.
- NVIDIA Docs. train with mixed precision. — [Electronic resource]. — <https://docs.nvidia.com/deep-learning/performance/pdf/Training-Mixed-Precision-User-Guide.pdf> (accessed: 27.01.2024).

- Redmon J., Divvala S., Girshick R., Farhadi A.* You only look once: Unified, real-time object detection // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2016. — P. 779–788.
- Redmon J., Farhadi A.* Yolov3: An incremental improvement // arXiv preprint. — 2018. — arXiv:1804.02767
- Ren S., He K., Girshick R., Sun J.* Faster R-CNN: Towards real-time object detection with region proposal networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2015. — Vol. 39, No. 6. — P. 1137–1149.
- Sindhu V.S.* Vehicle identification from traffic video surveillance using YOLOv4 // 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). — 2021. — P. 1768–1775.
- Tang W., Hua G., Wang L.* How to train a compact binary neural network with high accuracy? // Proceedings of the AAAI conference on artificial intelligence. — 2017. — Vol. 31, No. 1. — P. 18–27.
- Wang C. Y., Bochkovskiy A., Liao H. Y. M.* Scaled-YOLOv4: Scaling cross stage partial network // arXiv preprint. — 2021. — arXiv:2011.08036
- Wang C. Y., Bochkovskiy A., Liao H. Y. M.* YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2023. — P. 7464–7475.
- Wang S.* Research towards YOLO-series algorithms: Comparison and analysis of object detection models for real-time UAV applications // Journal of Physics: Conference Series. — 2021. — Vol. 1948, No. 1. — ID: 012021.
- Zoev I. V., Markov N. G., Ryzhova S. E.* Intelligent computer vision system for unmanned aerial vehicles for monitoring technological objects of oil and gas industry // Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering. — 2019. — Vol. 330, No. 11. — P. 34–49.
- Zoev I., Maslov K., Markov N., Mytsko E.* Analysis of hardware-implemented U-Net-like convolutional neural networks // Communications in Computer and Information Science. Tools and Methods of Program Analysis. — 2024. — Vol. 1559. — P. 52–63.