

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования



**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Направление подготовки: 09.06.01 «Информатика и вычислительная техника»

Профиль: 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Инженерная школа информационных технологий и робототехники

Отделение информационных технологий

**Научный доклад об основных результатах подготовленной
научно-квалификационной работы**

Тема научного доклада
Алгоритмическое и программное обеспечение системы компьютерного зрения подвижных робототехнических комплексов

УДК 004.932.2:007.52

Аспирант

Группа	ФИО	Подпись	Дата
А9-39	Ткачёв Семён Александрович		

Руководитель профиля подготовки

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ	Спицын Владимир Григорьевич	д.т.н., профессор		

Руководитель отделения

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Руководитель ОИТ	Шерстнев Владислав Станиславович	к.т.н., доцент		

Научный руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ	Марков Николай Григорьевич	д.т.н., профессор		

1 Актуальность темы НКР

В настоящее время актуальным научным направлением продолжает оставаться создание интеллектуальных систем компьютерного зрения (СКЗ) подвижных (мобильных) робототехнических комплексов различного назначения. Такие подвижные робототехнические комплексы имеют в своем составе встраиваемые СКЗ в виде отдельной подсистемы. В некоторых случаях комплексы периодически подключаются, в основном по радиоканалу, и используют в своей работе относительно автономные возимые СКЗ. Во всех этих случаях обычно речь идет о мобильных СКЗ, обладающих весьма ограниченными вычислительными ресурсами.

Сегодня наиболее перспективным направлением является реализация мобильных (встраиваемых и автономных) СКЗ на основе сверточных нейронных сетей (СНС). Для этого с учетом ограниченных вычислительных ресурсов мобильных СКЗ необходимо разрабатывать новые модели СНС и алгоритмическое и программное обеспечение СКЗ, которые позволят детектировать (обнаруживать, локализовать в пространстве и классифицировать) объекты на изображениях окружающей среды в реальном времени. Более того, предлагаемые новые модели СНС и алгоритмы функционирования СКЗ подлежат исследованию по точности и скорости детектирования объектов на изображениях с учетом выбранного масштаба реального времени при анализе изображений и возможностей по вычислительным ресурсам СКЗ. Все это указывает на актуальность разработки и исследования новых моделей СНС и алгоритмов обработки данных в мобильных СКЗ, функционирующих в реальном времени в условиях ограниченных вычислительных ресурсов этих систем.

2 Цель и задачи исследований

Известно, что не все современные модели СНС отвечают требованиям по скорости вычислений и точности детектирования объектов на изображениях, выдвигаемым при проектировании конкретных мобильных СКЗ. Этим взаимоисключающим требованиям особенно непросто удовлетворить в случаях создания мобильных СКЗ, функционирующих в режиме реального времени. Действительно, во многих практически важных случаях задачи детектирования объектов на изображениях обычно приходится решать в реальном времени и в условиях ограниченных ресурсов памяти и производительности вычислительных устройств мобильных СКЗ. Особенно эти ограничения приходится учитывать при разработке встраиваемых в робототехнические комплексы СКЗ и в случаях возимых или носимых СКЗ, где существуют еще и дополнительные жесткие ограничения по энергопотреблению. Именно таким основным ограничениям (часто говорят, требованиям) должна удовлетворять мобильная СКЗ, предназначенная для детектирования (обнаружения, локализации и классификации) летающих объектов в воздушном пространстве. Все это позволяет считать, что применение сложных моделей СНС во многих реальных задачах создания мобильных СКЗ весьма затруднительно. В этой связи при разработке мобильных СКЗ реального времени растет интерес к эффективным и компактным моделям СНС. По сути, необходимы модели СНС, позволяющие соблюдать баланс между точностью и скоростью детектирования и в итоге позволяющие мобильным СКЗ функционировать в масштабе реального времени и детектировать объекты на изображениях с приемлемой точностью.

Остановимся на количественных характеристиках этих требований к мобильной СКЗ и, соответственно, к лежащей в ее основе той или иной модели СНС. Точность детектирования (классификации) объектов на изображении вычисляется по общепринятым метрикам. В наших исследованиях использовались метрики AP – площадь под кривой Precision-Recall (AUC-PR), оцененная с использованием порогового значения IoU, и mAP – среднее значение AP по всем классам объектов. Для $AP_{0,5}$ и $mAP_{0,5}$ значение IoU равно 0,5, в то

время как для расчета $AP_{0,5:0,95}$ и $mAP_{0,5:0,95}$ используется интервал значений $IoU \in [0,5; 0,95]$ с шагом 0,05. Пороговые значения метрик $AP_{0,5}$ и $mAP_{0,5}$ должны быть 0,8 и 0,7, соответственно, для оптических и тепловизионных изображений. Требование по точности детектирования удовлетворяется, если превышено соответствующее пороговое значение.

Второй количественной характеристикой мобильной СКЗ является скорость детектирования объектов на изображении (скорость вычислений модели СНС), измеряемая как число анализируемых изображений в секунду в FPS (англ. Frames Per Second). Значение FPS, позволяющее считать, что мобильная СКЗ удовлетворяет требованию по скорости детектирования объектов с помощью модели СНС, должно быть не менее 25. Третья количественная характеристика – алгоритмическая эффективность модели СНС, определяемая как количество ресурсов вычислительного устройства, необходимое для достижения определенных показателей модели СНС в решении поставленной задачи. Она рассчитывается на основе вычислительной сложности и размера модели СНС. Вычислительная сложность модели СНС, в FLOPs – это количество операций умножения, сложения и сравнения над 16-битными числами с плавающей точкой. Размер модели (компактность архитектуры) СНС, в МБ – это объем памяти вычислительного устройства СКЗ, необходимый для хранения весовых коэффициентов сверточных фильтров и промежуточных буферов при вычислении модели СНС. Пороговые значения вычислительной сложности и размера модели СНС, позволяющие считать, что мобильная СКЗ (и, соответственно, модель СНС в ее основе) удовлетворяет требованиям разработчиков ее вычислительного устройства, задаются в техническом задании на создание мобильной СКЗ.

Анализ различных классов моделей СНС позволяет считать, что наиболее приемлемыми для детектирования объектов на изображениях являются модели класса YOLO (англ. You Only Look Once). Модели СНС класса YOLO, как и другие модели СНС с одноступенчатой архитектурой, обрабатывают изображение за один проход, что значительно повышает скорость анализа

изображения. Среди моделей СНС класса YOLO перспективной является модель YOLOv4. Это одноступенчатая (часто говорят, одноэтапная) модель обнаружения и классификации объектов, созданная на основе предыдущих оригинальных моделей YOLO. Ряд авторов показывает, что по критерию точности предпочтение следует отдать модели СНС YOLOv4. Более того, она является весьма компактной моделью, поскольку содержит малое число сверточных слоев и поэтому занимает малый объем памяти вычислительного устройства мобильной СКЗ.

Пусть на основе выбранной для реализации и исследований модели СНС из класса YOLO проектируется мобильная СКЗ, которая позволяет детектировать на оптических изображениях летающие объекты следующих четырех классов:

- Птица;
- Беспилотный летательный аппарат (БПЛА) самолетного типа;
- БПЛА вертолетного типа;
- Неизвестный объект.

Кроме того, эта мобильная СКЗ должна позволять детектировать на тепловизионных изображениях летающие объекты следующих двух классов:

- БПЛА вертолетного типа;
- Планер.

Целью данной работы является разработка на основе базовой модели YOLOv4 класса YOLO новой модели YOLOv4-CSP и исследование ее с помощью разработанного алгоритмического и программного обеспечения для последующего включения такой модели в мобильную СКЗ реального времени с целью детектирования летающих объектов указанных четырех классов на оптических изображениях и двух указанных классов на тепловизионных (инфракрасных) изображениях.

Для реализации поставленной цели необходимо решить *следующие задачи*.

1. Сформировать два набора RGB-изображений и тепловизионных изображений летающих объектов. Причем на изображениях первого набора содержатся объекты четырех классов: «Птица, «БПЛА самолетного типа», «БПЛА вертолетного типа» и «Неизвестный объект». Более того, на каждом изображении первого набора содержится один и более объектов перечисленных классов и объекты будут иметь разные размеры (разномасштабные объекты). Изображения второго набора содержат также разномасштабные объекты, но только двух классов : «БПЛА вертолетного типа» и «Планер».

2. Для автоматизации процессов создания датасетов необходимо разработать соответствующее алгоритмическое и программное обеспечение.

3. Разметить и аннотировать объекты указанных классов на RGB-изображениях и тепловизионных изображениях из созданных наборов и сформировать в итоге два датасета (первый датасет сформировать на оптических изображениях первого набора, а второй датасет - на тепловизионных изображениях второго набора) для обучения, валидации и исследования новой модели CHC YOLOv4-CSP из семейства YOLO.

4. Провести разработку путем модификации выбранной в результате анализа базовой модели CHC YOLOv4 и полученной таким образом модели YOLOv4-CSP. Провести обучение последней с использованием обучающих и валидационных выборок из подготовленных датасетов.

5. Исследовать эффективность обученной модели CHC YOLOv4-CSP с использованием тестовых выборок каждого из датасетов.

6. Провести анализ всех результатов исследований модели CHC на предмет соответствия ранее сформированным требованиям к мобильной СКЗ и, соответственно, к этой модели CHC.

3 Краткое содержание работы

В первой главе вводится понятие мобильных СКЗ. По ряду критериев анализируются различные современные модели СНС, в том числе из семейства YOLO, которые, в принципе, можно использовать для создания мобильных СКЗ. В итоге выбирается наиболее перспективная для создания мобильной СКЗ модель СНС YOLOv4 из этого семейства. Формулируются цель работы и решаемые для ее реализации задачи.

Во второй главе описано решение задачи создания двух датасетов для обучения, валидации и исследования моделей СНС. В качестве исходных изображений при формировании датасетов используются оптические изображения летающих объектов четырех классов и тепловизионные изображения летающих объектов двух классов.

Третья глава посвящена особенностям разработки, обучения и валидации новой модели СНС YOLOv4-CSP на основе анализа моделей СНС семейства YOLO и выбранной для модификации модели СНС YOLOv4.

В четвертой главе приводятся и анализируются результаты исследований эффективности обученной и валидированной модели СНС YOLOv4-CSP при использовании тестовых выборок первого и второго датасетов.

4 Основные полученные результаты и сделанные выводы

Во время выполнения НКР были получены следующие результаты.

1. Проанализированы модели СНС для детектирования летающих объектов на изображениях с целью создания мобильных СКЗ. Выявлено, что наиболее перспективными для реализации этой цели являются модели СНС семейства YOLO. Более того, по ряду причин предпочтение следует отдать модели YOLOv4.

2. Наш опыт показал, что объем каждого из изначально полученных датасетов, созданных за счет объединения размеченных другими авторами наборов изображений (сторонних датасетов) и ручного сбора и разметки изображений, обычно невелик. Поэтому для формирования полноценных по объемам датасетов необходимо использовать следующий известный подход. Суть его в проведении аугментации уже размеченных изображений, входящих в датасет. При этом применялись различные известные методы аугментации (незначительные изменения масштаба, отражения по вертикальной оси и так далее), а для изображений с летающими объектами малых размеров (размер не более 32×32 пикселей) предложен оригинальный способ аугментации. Он заключается в том, что выполняется процесс копирования и вставки на другие изображения датасета малоразмерных объектов с применением к ним трансформаций (небольшое изменение размера, вращение) в случайные места таким образом, чтобы они не перекрывали уже имеющиеся на изображении объекты.

3. Учитывая особенности разметки изображений в инструментальной среде Roboflow, описанные выше подходы к формированию датасетов и оригинальный способ аугментации, был разработан алгоритм создания датасетов.

4. Для проведения описанной выше предварительной подготовки изображений и реализации предложенного алгоритма разработано соответствующее программное обеспечение на языке программирования Python

3 в среде Anaconda с использованием следующих библиотек: OpenCV, Numpy, Pandas, Shapely, Scikit Learn, YAML и стандартных библиотек Python.

5. С помощью разработанного алгоритмического и программного обеспечения созданы два датасета. Основной датасет, включающий оптические изображения, имеет 6230 изображений с 33810 летающими объектами четырех классов на них и столько же файлов аннотаций. Второй датасет, включающий 2790 тепловизионных изображений, имеет 5580 объектов двух классов и столько же файлов аннотаций.

6. Получены следующие результаты исследования разработанной модели СНС YOLOv4-CSP на тестовых выборках первого и второго датасетов.

а. Показано, что точность детектирования объектов на изображениях выше при использовании функции активации Mish для различных размеров входных изображений. Эта функция является гладкой, немонотонной, ограниченной снизу и неограниченной сверху, поэтому эти ее свойства позволяют получать наилучшие результаты по сравнению с функцией активации Leaky ReLU.

б. Точность детектирования объектов на изображениях с помощью этой модели СНС выше на входных изображениях большего размера. Это имеет место для каждой из выборок датасетов. Так, например, при изменении размера изображения с 416×416 на 608×608 пикселей на тестовой выборке первого датасета точность детектирования объектов по метрике $mAP_{0.5}$ выросла на 3,3 %, а по метрике $mAP_{0.5:0.95}$ – на 5,7 %.

с. Точность детектирования объектов ниже на изображениях тестовой выборки, чем на изображениях обучающей и проверочной (валидационной) выборок каждого из датасетов. Так, при использовании тестовой выборки с изображениями с размером 608×608 пикселей по сравнению с обучающей выборкой первого датасета точность детектирования объектов по метрике $mAP_{0.5}$ снизилась на 7 %, а по метрике $mAP_{0.5:0.95}$ – на 4,1 %. Точность детектирования объектов класса «Птица» по метрике $AP_{0.5}$ хуже, чем для объектов других классов. Это объясняется меньшим числом изображений

объектов этого класса в первом датасете, а также малыми размерами объектов на этих изображениях. По метрике $AP_{0.5:0.95}$ точность ниже при детектировании объектов класса «Неизвестный объект».

d. Скорость детектирования объектов (скорость вычисления модели СНС) выше на изображениях меньшего размера. Так, при изменении размера изображения первого датасета с 416×416 на 608×608 пикселей время обработки одного изображения увеличилось на 2,8 мс.

е. Точность детектирования объектов на оптических изображениях выше на 7 и 10 % (по метрикам $mAP_{0.5:0.95}$ и $mAP_{0.5}$ соответственно), чем на тепловизионных изображениях.

7. Итоговый анализ всех этих результатов исследований разработанной модели YOLOv4-CSP на соответствие требованиям, предъявляемым к моделям СНС со стороны проектируемых мобильных СНС, позволяет сделать вывод, что эта модель СНС соответствует сформированным требованиям.