

ЯРОСЛАВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
им. П.Г. ДЕМИДОВА

На правах рукописи



**Лебедев Антон Александрович**

**ИССЛЕДОВАНИЕ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ  
ОБНАРУЖЕНИЯ ОБЪЕКТОВ НА ВИДЕОИЗОБРАЖЕНИЯХ  
В МЕДИЦИНСКИХ СИСТЕМАХ ПРИКЛАДНОГО  
ТЕЛЕВИДЕНИЯ**

Специальность 2.2.13

Радиотехника, в том числе системы и устройства телевидения

**ДИССЕРТАЦИЯ**

на соискание ученой степени  
кандидата технических наук

Научный руководитель:  
кандидат технических наук, доцент  
Хрящев Владимир Вячеславович

Ярославль – 2022

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	5
ГЛАВА 1 ТЕКУЩЕЕ СОСТОЯНИЕ ОБЛАСТИ ИССЛЕДОВАНИЙ .....	14
1.1 Системы прикладного телевидения .....	14
1.2 Эндоскопия как область медицинского прикладного телевидения.....	18
1.3 Задача обнаружения полипов на эндоскопических изображениях.....	22
1.4 Методы и алгоритмы обнаружения объектов на изображениях .....	26
1.4.1 Метрики оценки качества работы алгоритмов обнаружения объектов на изображениях .....	28
1.4.2 Классические алгоритмы обнаружения объектов на изображениях	31
1.4.3 Двухэтапные алгоритмы обнаружения объектов на изображениях на основе сверточных нейронных сетей.....	35
1.4.4 Одноэтапные алгоритмы обнаружения объектов на изображениях на основе сверточных нейронных сетей.....	38
1.5 Краткие выводы .....	41
ГЛАВА 2 ИССЛЕДОВАНИЕ АЛГОРИТМОВ ОБНАРУЖЕНИЯ ПОЛИПОВ НА ЭНДОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЯХ КИШЕЧНИКА .....	43
2.1 Вводные замечания.....	43
2.2 Алгоритмы обнаружения полипов на основе одноэтапных нейросетевых методов.....	45
2.2.1 Архитектура сверточной нейронной сети SSD.....	47
2.2.2 Архитектура сверточной нейронной сети YOLOv5 .....	49
2.3 Базы эндоскопических изображений .....	51
2.4 Модификация алгоритма аугментации эндоскопических изображений	53
2.5 Тестирование нейросетевых алгоритмов обнаружения полипов.....	60

2.6 Сравнительный анализ нейросетевых алгоритмов обнаружения полипов на базе Kvasir-SEG .....	66
2.7 Краткие выводы .....	69
ГЛАВА 3 РАЗРАБОТКА И АНАЛИЗ АЛГОРИТМА ОБНАРУЖЕНИЯ ПОЛИПОВ НА ЭНДОСКОПИЧЕСКИХ ВИДЕОИЗОБРАЖЕНИЯХ.....	71
3.1 Вводные замечания.....	71
3.2 Особенности алгоритмов обнаружения объектов на видеоизображениях .....	72
3.3 Применение алгоритмов обнаружения полипов для анализа видеоданных колоноскопических исследований.....	75
3.4 Модификация алгоритмов обнаружения полипов для анализа видеопотока в режиме реального времени.....	78
3.5 Разработка алгоритма обнаружения полипов на видеоданных с постобработкой результатов.....	80
3.6 Создание и разметка базы видеоданных колоноскопических исследований в реальной эндоскопической системе прикладного телевидения .....	86
3.6.1. Разработка программно-аппаратного тестового стенда с модулем анализа видеопотока.....	87
3.6.2. Разработка полуавтоматической методики разметки видеоизображений .....	89
3.7 Тестирование итогового нейросетевого алгоритма обнаружения.....	91
3.8 Краткие выводы .....	96
ЗАКЛЮЧЕНИЕ .....	99
ЛИТЕРАТУРА.....	102
ПРИЛОЖЕНИЕ А. Технические характеристики видеопроцессоров эндоскопических систем .....	120

ПРИЛОЖЕНИЕ Б. Свидетельства о государственной регистрации интеллектуальной собственности.....	124
ПРИЛОЖЕНИЕ В. Акты внедрения .....	127

## ВВЕДЕНИЕ

**Актуальность темы.** Цифровая обработка видеоизображений – одна из перспективных и быстро развивающихся областей науки и техники, находящая применение при решении все более широкого круга задач, решаемых в ходе человеческой жизнедеятельности. Одними из наиболее практически востребованных задач в данной области на текущем этапе научно-технического развития являются задачи автоматического обнаружения (детектирования) и классификации объектов на видеоизображениях [15, 23, 32, 51, 57, 115, 132]. Они встречаются в различных сферах применения систем прикладного телевидения (СПТ): в охранных системах для обнаружения движущихся объектов, в промышленном телевидении для аналитики технологических процессов, в медицинских системах для визуализации различных патологий. В большинстве случаев для решения таких задач наряду с алгоритмами цифровой обработки изображений и технического зрения используются методы и алгоритмы машинного обучения [3-6, 8, 11, 25].

Развитие подходов к обнаружению объектов на изображениях можно разделить на два этапа: использование классических алгоритмов машинного обучения (на основе эмпирически выбираемых признаков) и использование методов глубокого обучения (на основе применения сверточных нейронных сетей – СНС). В свою очередь нейросетевые алгоритмы обнаружения объектов разделяются на две основные категории: «одноэтапные алгоритмы обнаружения» и «двухэтапные алгоритмы обнаружения». Первые имеют лучшую производительность, но обычно проигрывают в качестве достигаемых результатов на практике, хотя современные одноэтапные алгоритмы достигают сравнимой с двухэтапными точности обнаружения [56, 72, 108, 127, 146, 147].

Разработка и исследование подобных нейросетевых систем, анализирующих видеопоток в режиме реального времени, представляет собой актуальную теоретическую и практическую проблему в области

построения СПТ. Помимо задачи выбора оптимальной архитектуры нейронной сети и ее обучения исследователи сталкиваются с рядом проблем, касающихся подготовки (разметки) и предобработки входных данных, настройки алгоритмов аугментации (размножения) данных, а также алгоритмов постобработки, учитывающих специфику телевизионного сигнала [7, 43, 88, 140, 156].

Методы цифровой обработки изображений все чаще применяются в медицинских СПТ, используемых для диагностики различных заболеваний. Накопление медицинских изображений и видеоданных в таких системах способствует появлению новых перспективных технологий для анализа и обработки данных с целью частичной автоматизации диагностики и контроля качества диагностических процедур [3-6, 8, 20, 22, 24, 30]. Одной из таких прикладных телевизионных систем является эндоскопическая система, используемая, в частности, для колоноскопических исследований кишечника [9, 88].

Поиск и удаление предраковых аномалий – один из наиболее эффективных способов борьбы с колоректальным раком. При этом особую значимость для анализа видеоизображений имеет выделенный класс объектов – полипы толстого кишечника, так как именно они могут перерасти в рак на последующей стадии. В среднем при обследовании кишечника врачом высокой квалификации пропускается порядка 20% полипов. Таким образом, риск заболевания раком в значительной степени зависит от способности врача-эндоскописта обнаруживать полипы во время колоноскопической процедуры [91]. Недавние исследования показали, что новые эндоскопические устройства и диагностические инструменты повысили вероятность обнаружения полипов [51, 115]. Тем не менее, проблема пропущенных полипов остается по-прежнему актуальной и имеет решающее значение в борьбе с онкологией кишечника [91].

Несмотря на все более совершенное устройство современных эндоскопов, для получаемых с них эндоскопических видеоизображений характерен ряд проблем, усложняющих их анализ [35, 68]:

- в силу конструктивных особенностей в видеоэндоскопах отсутствует автофокусировка, что часто приводит к отсутствию фокусировки на объекте интереса;
- при движении видеоэндоскопа в полости органа часто возникает размытие изображений;
- в ПЗС- или КМОП-матрице присутствует тепловой шум;
- на слизистой часто можно наблюдать блики, вызванные отражением света, испускаемого источником света видеоэндоскопа;
- яркость и контрастность получаемого эндоскопического изображения может сильно меняться в зависимости от геометрии исследуемого органа;
- на практике на выходе видеопроцессора эндоскопической системы обычно используется стандарт передачи видео с чересстрочной разверткой, что приводит к наличию на изображениях специфических искажений.

Таким образом, анализ российской и зарубежной литературы показывает, что актуальной научно-технической задачей является применение алгоритмов цифровой обработки изображений и машинного обучения для автоматизации диагностики в системах эндоскопического прикладного телевидения. Применение такого подхода позволит повысить точность диагностики, уменьшить влияние человеческого фактора на качество исследований, снизить стоимость и временные затраты на их проведение [51, 91, 115].

Важнейший вклад в развитие области обработки цифровой обработки изображений и распознавания образов внесли отечественные ученые В.П. Дворкович, А.В. Дворкович, А.И. Галушкин, Ю.И. Журавлев, В.А. Сойфер, В.Н. Безруков, Ю.В. Визильтер, А.С. Крылов, А.С. Конушин,

Д.С. Ватолин, М.К. Чобану, Ю.С. Бехтин, М.Н. Фаворская, Ю.С. Радченко, Е.В. Медведева, А.Л. Приоров, С.С. Садыков и др. Важнейшие результаты в области обработки цифровых изображений с использованием нейронных сетей получены также следующими зарубежными учеными: Y. LeCun, Y. Bengio, G. Hinton, I. Goodfellow, K. He, R. Girshick, T.-Y. Lin, J. Redmon.

В области применения алгоритмов цифровой обработки для анализа эндоскопических видеоизображений следует отметить работы R. Haidry, M. Hafner, D. Jha, Y. Shin, S. Ali, N.K. Tomar, H. Johansen, D. Johansen, J. Rittscher, Н.А. Обуховой, А.А. Мотыко, С.В. Кашина, Р.О. Куваева, Д.В. Завьялова и др.

**Целью работы** является повышение качества работы систем прикладного телевидения в эндоскопии путем разработки нейросетевых алгоритмов обнаружения объектов на видеоизображениях.

**Для достижения указанной цели в диссертационной работе решаются следующие задачи:**

- разработка и исследование нейросетевых алгоритмов обнаружения объектов на видеоизображениях эндоскопических исследований;
- разработка методики полуавтоматического создания размеченной базы эндоскопических видеоданных для обучения нейросетевых моделей;
- усовершенствование алгоритмов предобработки и постобработки видеоизображений, учитывающих специфику видеоэндоскопических данных;
- тестирование разработанного алгоритма для анализа эндоскопического видеопотока в реальной медицинской системе прикладного телевидения.

**Методы исследования.** При решении поставленных задач использовались современные методы цифровой обработки сигналов и изображений, теория нейронных сетей и глубокого обучения. Для практической реализации алгоритмов применялись современные методы программирования и компьютерного моделирования на языках Python и C++.



**Объектом исследования** являются нейросетевые алгоритмы обнаружения объектов на видеоизображениях в медицинских системах прикладного телевидения.

**Предметом исследования** является разработка, модификация и улучшение качества алгоритмов цифровой обработки видеоизображений и глубокого машинного обучения, применяемых для обнаружения патологических объектов в системах прикладного телевидения в эндоскопии.

### **Научная новизна полученных результатов**

В рамках работы получены следующие новые научные результаты:

1. Предложен нейросетевой алгоритм обнаружения полипов на статических изображениях в базах эндоскопических исследований.
2. Предложен нейросетевой алгоритм обнаружения полипов на видеоизображениях эндоскопических исследований, работающий в режиме реального времени.
3. Разработана методика полуавтоматического создания размеченной базы эндоскопических видеоданных для обучения нейросетевых моделей.
4. Усовершенствованы с учетом специфики видеоэндоскопических данных алгоритмы аугментации и постобработки видеоизображений.

### **Практическая значимость полученных результатов**

1. Разработанный нейросетевой алгоритм позволяет обнаруживать полипы на видеопотоке с эндоскопической системы в режиме реального времени. Он может быть использован в медицинских системах прикладного телевидения для частичной автоматизации диагностики, а также для контроля, обучения и подготовки профильных врачей.
2. Значение метрики средней точности для разработанного алгоритма обнаружения полипов на эндоскопических изображениях с модифицированным алгоритмом аугментации данных составляет 92,5% на общедоступной базе статических изображений Kvasir-SEG,

что позволяет рекомендовать данный алгоритм для построения на его основе модуля анализа видеопотока в эндоскопической системе прикладного телевидения.

3. Предложена и апробирована на наборе из более чем 31 000 видеокадров методика подготовки базы видеоданных эндоскопических исследований с разметкой областей интереса для обучения нейросетевых моделей, позволяющая сократить соответствующие трудозатраты на 30-40%.
4. Установлено, что применение предложенного алгоритма постобработки позволяет улучшить качество системы обнаружения полипов на видеопотоке, в среднем на 1,5-2 %.

Результаты работы внедрены в клиническую практику эндоскопического отделения Ярославской областной клинической онкологической больницы, а также в перспективные разработки компаний ООО «Точка зрения» (г. Ярославль) и ООО «Смарт анализ» (г. Ярославль). Кроме того, проведено полугодовое тестирование предложенных алгоритмов в рамках специализированного программно-аппаратного-комплекса в ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России (Пироговский центр, г. Москва).

Отдельные результаты диссертационной работы внедрены в учебный процесс ЯрГУ им. П.Г. Демидова в рамках дисциплин «Системы технического зрения» и «Цифровая обработка изображений», а также в научно-исследовательские работы при выполнении гранта РФФИ № 19-37-90153 и гранта победителя программы «Старт-1», договор с Фондом содействия развитию малых форм предприятий в научно-технической сфере № 3183ГС1/48626 от 26.08.2019.

Получены два свидетельства о государственной регистрации программ для ЭВМ (№ 2020618382, № 2022610262) и свидетельство о государственной регистрации базы данных (№ 2022620355).

**Достоверность полученных научных результатов** диссертационной работы обеспечивается корректным использованием математического

аппарата и полученными экспериментальными данными, подтверждающими теоретические гипотезы и согласующимися с результатами, известными из российских и зарубежных научно-технических источников.

**Апробация работы.** Результаты работы обсуждались на следующих научно-технических семинарах и конференциях:

1. Двадцатая – двадцать третья международные конференции «Цифровая обработка сигналов и ее применение», Москва, 2018–2021.
2. Семнадцатая международная научно-техническая конференция «Новые информационные технологии и системы», Пенза, 2020.
3. Четырнадцатая международная научно-технической конференция «Перспективные технологии в средствах передачи информации», Владимир, 2021.
4. Двенадцатая международная конференция по машинному зрению ICMV 2019 (International Conference on Machine Vision), Нидерланды, Амстердам, 2019.
5. 18-й и 19-й международные симпозиумы «IEEE East-West Design & Test Symposium», 2020-2021, Болгария, Варна / Грузия, Батуми.
6. 77-я научно-техническая конференция Санкт-Петербургского НТО РЭС им. А.С. Попова, посвященная Дню радио, Санкт-Петербург, 2022.
7. 77-я всероссийская конференция (с международным участием) «Радиоэлектронные устройства и системы для инфокоммуникационных технологий», Москва, 2022.

**Публикации.** По теме диссертации опубликовано 17 научных работ, из них 4 статьи в журналах из Перечня ВАК, 4 работы, индексируемые в базе SCOPUS, и 9 докладов на научных конференциях различного уровня.

**Структура и объем работы.** Диссертация состоит из введения, трех глав, заключения, списка литературы, содержащего 159 наименований, и трех приложений. Она изложена на 131 страницах машинописного текста, содержит 35 рисунков и 9 таблиц.

## **Основные научные положения и результаты, выносимые на защиту**

- алгоритм обнаружения объектов на статических изображениях эндоскопических исследований на основе сверточной нейронной сети с модифицированным алгоритмом аугментации данных, для которого значение метрики  $AP@[0,25..0,75]$  составило 92,5%, что превосходит известные аналоги на 5% и более;
- алгоритм обнаружения объектов на видеоизображениях эндоскопических исследований с усовершенствованной процедурой постобработки, который позволил улучшить значение метрики  $AP@0,50$  на базе видеоданных ЯОКОБ в среднем на 1,7%;
- методика полуавтоматического создания размеченной базы эндоскопических видеоданных для обучения нейросетевых моделей, позволяющая сократить соответствующие трудозатраты на 30-40%.

### **Благодарности**

Автор выражает благодарность коллективам Центра искусственного интеллекта и цифровой экономики, а также кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, в особенности научному руководителю, кандидату технических наук Хрящеву Владимиру Вячеславовичу, а также доктору технических наук Приорову Андрею Леонидовичу за помощь на всех этапах написания данной работы.

Автор также выражает благодарность за сотрудничество и помощь в работе сотрудникам отделения эндоскопии Ярославской областной клинической онкологической больницы: Кашину Сергею Владимировичу, Куваеву Роману Олеговичу и Завьялову Дмитрию Вячеславовичу. За помощь в экспериментальном тестировании разработанных нейросетевых алгоритмов благодарю сотрудников ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России: Маады Аяса Сергеевича, Субботина Сергея Александровича, Пензина Олега Владимировича.

Кроме того, глубокую благодарность хотелось бы выразить близким – жене и родителям – за возможность заниматься научной деятельностью и поддержку во время написания данной работы.

## ГЛАВА 1

# ТЕКУЩЕЕ СОСТОЯНИЕ ОБЛАСТИ ИССЛЕДОВАНИЙ

### 1.1 Системы прикладного телевидения

Современное телевидение – научно-техническое направление, объединяющее вопросы электрооптического анализа и синтеза как движущихся, так и неподвижных изображений, и передачи их на расстояние по электрическим и волоконно-оптическим линиям связи [6]. Современные телевизионные системы делятся на вещательные и прикладные, отличающиеся характером решаемых задач, а, следовательно, и своими техническими особенностями [3].

Системы прикладного телевидения используются как средство контроля, наблюдения и связи при решении различных практических задач, в которых непосредственное визуальное наблюдение не позволяет получить нужную информацию об объекте наблюдения или же оно в данных условиях трудоемко, опасно или вовсе неосуществимо.

Одним из существенных различий между прикладным и вещательным телевидением является количество потребителей: вещательное телевидение ориентировано на широкое вещание, в то время как СПТ рассчитаны на одного или нескольких потребителей. В связи с указанной особенностью СПТ не требуют такой жесткой стандартизации параметров, как вещательные системы [3, 6, 11].

С точки зрения информационных свойств СПТ можно разделить на следующие три категории [11]:

- телевизионные системы, в которых вся задача сводится к воспроизведению ТВ изображения. В данных телевизионных системах ключевая роль отводится оператору. Именно оператор осуществляет анализ передаваемой через систему информации с помощью своей зрительной системы;

- полуавтоматические телевизионные системы, в которых кроме задачи воспроизведения ТВ изображения, решается также задача получения количественных данных об объекте наблюдения на основе входной оптической информации и вывода данных на экране. В данных системах оператор анализирует передаваемую информацию, дополненную дополнительной информацией, сформированной телевизионной системой;
- автоматические телевизионные системы, в которых осуществляется анализ и обработка закодированной в видеосигнале оптической информации. В данных системах отсутствует оператор, а предоставляемый окончательный результат анализа подается на управляющее устройство.

Для решения прикладных задач в полуавтоматических и автоматических системах используются методы цифровой обработки изображений, технического зрения и машинного обучения. При этом задачи, решаемые с применением данных методов, могут быть достаточно разнообразны.

В соответствии с назначением цифровые телевизионные системы можно условно разделить на три категории. К первой категории относятся системы для решения сложных научных проблем. В основе данной категории систем лежит применение ЭВМ общего назначения. Ко второй категории относятся системы для решения широкого круга задач обработки изображений. К третьей категории относятся системы, предназначенные для работы в промышленности и способные обрабатывать относительно простые телевизионные изображения. Промышленные цифровые телевизионные системы часто применяют в качестве систем технического зрения (СТЗ) роботов, которые решают задачи автоматического контроля изделий, совершают сборочные и другие технологические операции [3].

Особое внимание следует уделить разделению цифровых телевизионных систем по времени обработки изображений. В соответствии с

данным критерием выделяют цифровые телевизионные системы, работающие в режиме реального времени и системы с длительной обработкой. В первом случае время обработки одного изображения не должно превышать время длительности одного кадра. Обработка видеопотока с точки зрения оператора происходит без задержек. Во втором случае время обработки одного изображения может занимать существенное время – от нескольких секунд до десятков минут и даже часов (например, при обработке больших спутниковых изображений, в особенности стереопар изображений) в зависимости от сложности алгоритма обработки [3].

Одной из областей применения СПТ является медицина. Современные телевизионные методы используются для осмотра биологических объектов как в видимой области спектра (эндоскопия – метод осмотра полых внутренних органов, микроскопия – исследование поверхности кожного покрова, волос, микропрепаратов и др.), так и в невидимых лучах (рентгеноскопия, ультрафиолетовая и инфракрасная микроскопия, инфракрасная офтальмоскопия и т. п.) [4, 5, 8].

Помимо непосредственного наблюдения и трансляции медицинских исследований удаленным специалистам, современные телевизионные системы позволяют также осуществлять автоматический анализ качественных и количественных характеристик наблюдаемых объектов на основе алгоритмов технического зрения и машинного обучения. Данное направление относится к области измерительного прикладного телевидения и является относительно новым. Необходимость развития данной области не вызывает сомнений: частичная автоматизация процесса медицинской диагностики позволяет не только повысить качество оказания медицинской помощи, но и оптимизировать труд врача, обеспечивая экономию времени, затраченного на рутинные процедуры.

Среди основных задач в области измерительного прикладного телевидения в медицине главную роль играет задача автоматического



обнаружения областей с патологической анатомией на медицинских изображениях и видеоданных.

Одной из областей, перспективных с точки зрения внедрения алгоритмов измерительного прикладного телевидения, является эндоскопия. От результатов проведения эндоскопического исследования зависят дальнейшие действия врачей других специализаций (хирургии, онкологии и др.), а также здоровье пациента. Таким образом, врачебные ошибки специалистов в области эндоскопии, неизбежно возникающие ввиду высокой сложности эндоскопического исследования и по ряду иных причин, имеют значительные последствия для дальнейшего прогноза течения заболевания. Указанные ошибки являются результатом неверной интерпретации качественных и количественных характеристик патологических областей желудочно-кишечного тракта. Их можно разделить на следующие группы:

- пропуск патологии, когда специалист не смог обнаружить область интереса (область с патологической анатомией);
- неверное определение вида патологии, например, специалист определяет раковую опухоль как язву или наоборот и т.д.;
- неверное определение характеристик обнаруженной патологии, например, области расположения или размеров [1, 10, 26, 31, 65, 126].

Одним из наиболее перспективных путей повышения качества эндоскопической диагностики является разработка эффективных алгоритмов для обнаружения и классификации патологий на эндоскопических данных. Такие алгоритмы могут применяться как в режиме реального времени непосредственно в процессе эндоскопического исследования, так и для верификации результатов после него. Внедрение таких подходов в медицинские СПТ позволит повысить эффективность работы врачей. Кроме того, такие алгоритмы могут использоваться в процессе обучения и подготовки новых медицинских специалистов.

Таким образом, на сегодняшний день разработка новых эффективных средств измерительного прикладного телевидения в области эндоскопии

является актуальной научно-технической задачей. Ее решению будет способствовать создание технических и методологических средств, позволяющих разрабатывать современные алгоритмы анализа изображений и видеоданных на основе методов глубокого машинного обучения. Исследованию таких алгоритмов автоматического обнаружения патологий на эндоскопических видеоизображениях толстого кишечника и посвящена настоящая работа.

## **1.2 Эндоскопия как область медицинского прикладного телевидения**

Эндоскопия – малоинвазивная медицинская диагностическая процедура, позволяющая проводить осмотр полых органов через естественные отверстия. При проведении эндоскопии желудочно-кишечного тракта (ЖКТ) в полость органа вводится видеоэндоскоп (Рисунок 1.1). Это высокотехнологичный многофункциональный оптоэлектронный инструмент, позволяющий не только производить малоинвазивный осмотр и диагностику ЖКТ, но и осуществлять взятие биологического материала из тканей подозрительных новообразований для гистологического анализа, выполнять их удаление, останавливать кровотечения и др. [97].



Рисунок 1.1 – Видеоэндоскоп GIF-160 фирмы Olympus

Для возможности выполнения всех вышеуказанных функций дистальная (передняя) часть видеоэндоскопа содержит два мощных ксеноновых источника света, выход инструментального канала и канала подачи воздуха и воды, а также линзу объектива, как показано на Рисунке 1.2. Ручка управления позволяет манипулировать дистальным концом, изгибая его более чем до 200 градусов, для осмотра труднодоступных участков слизистой оболочки ЖКТ.



Рисунок 1.2 – Устройство дистальной части видеоэндоскопа

Изображение, получаемое с видеоэндоскопа при проведении исследования, выводится на монитор, расположенный перед врачом-эндоскопистом. Кроме того, данные с видеоэндоскопа могут параллельно транслироваться в учебных целях для большой аудитории молодых врачей и студентов-медиков [13]. В частности, в Ярославской областной клинической онкологической больнице происходит трансляция проводимых исследований в специализированную ординаторскую аудиторию, оснащенную большими ТВ-панелями.

Большинство современных эндоскопических систем поддерживает возможность формирования видеоданных с высоким разрешением, обеспечивающих возможность качественного осмотра органов ЖКТ [2, 29

48]. Однако, несмотря на все более совершенное устройство современных видеоэндоскопов, для получаемых с них эндоскопических изображений характерен ряд проблем, перечисленных во введении данной работы.

При проведении эндоскопии широко используется ряд технологий, направленных на улучшение качества диагностики. Примером является применение узкоспектрального режима NBI (Narrow Band Imaging) – позволяющего производить осмотр тканей в узком спектре света [137]. При использовании такого режима работы ткани освещаются светом с двумя длинами волн видимого диапазона: синий свет с длиной волны 415 нм проникает только в поверхностные слои слизистой оболочки и хорошо поглощается гемоглобином в поверхностных капиллярах, а зеленый свет с длиной волны 540 нм способен проникать чуть глубже и поглощаться более глубокими сосудистыми структурами (Рисунок 1.3).

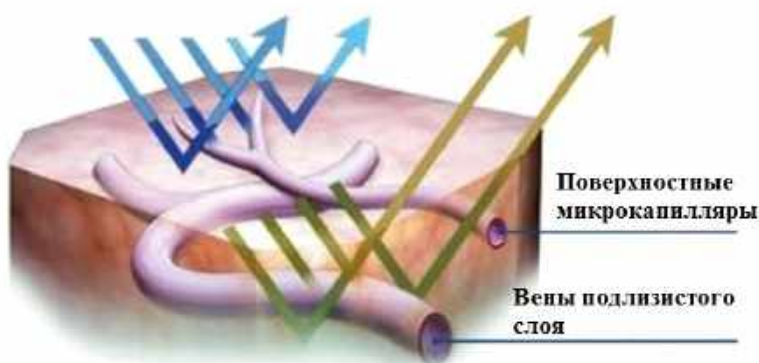
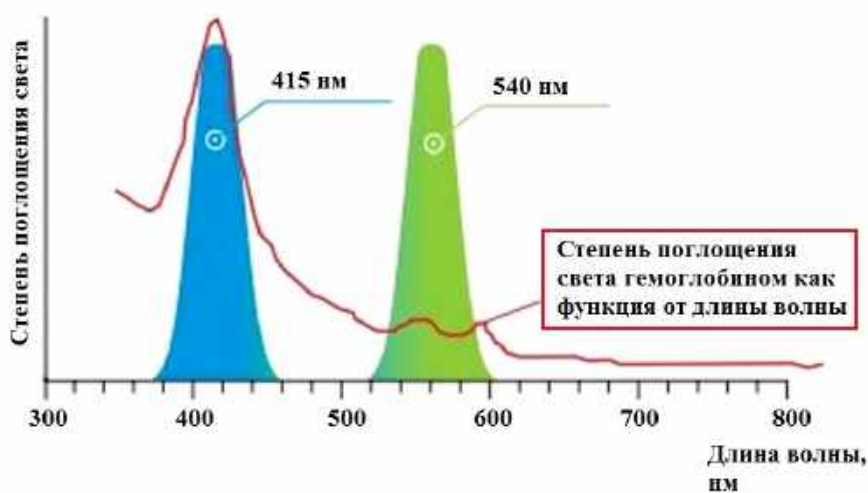
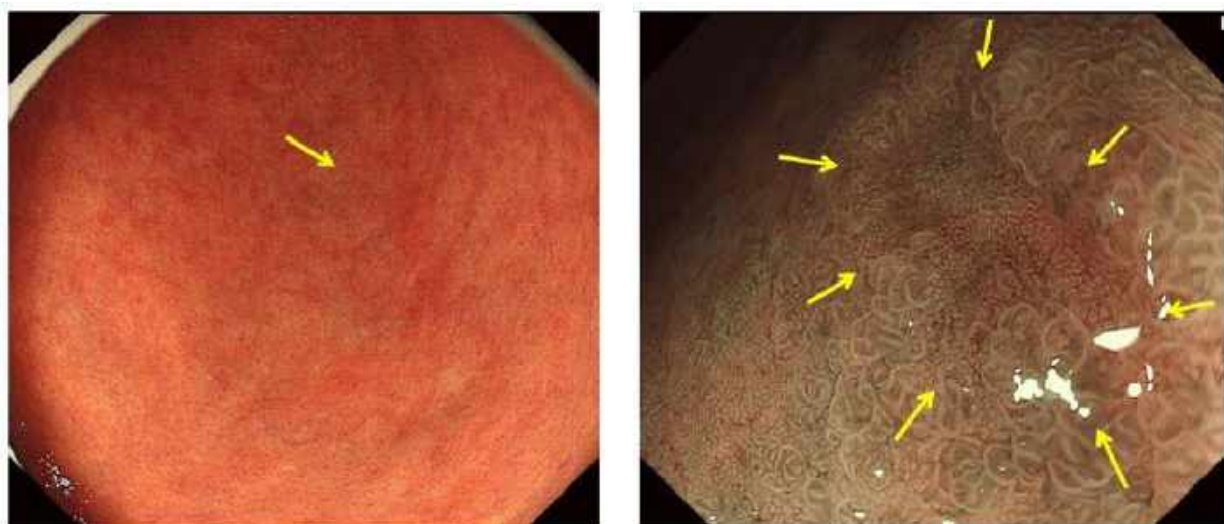


Рисунок 1.3 – Режим узкоспектральной визуализации NBI

Использование режима NBI значительно повышает возможности визуализации при проведении эндоскопии, позволяя максимально выделять микрокапиллярный рисунок, по которому врач может судить о наличии патологических изменений слизистой оболочки (Рисунок 1.4).



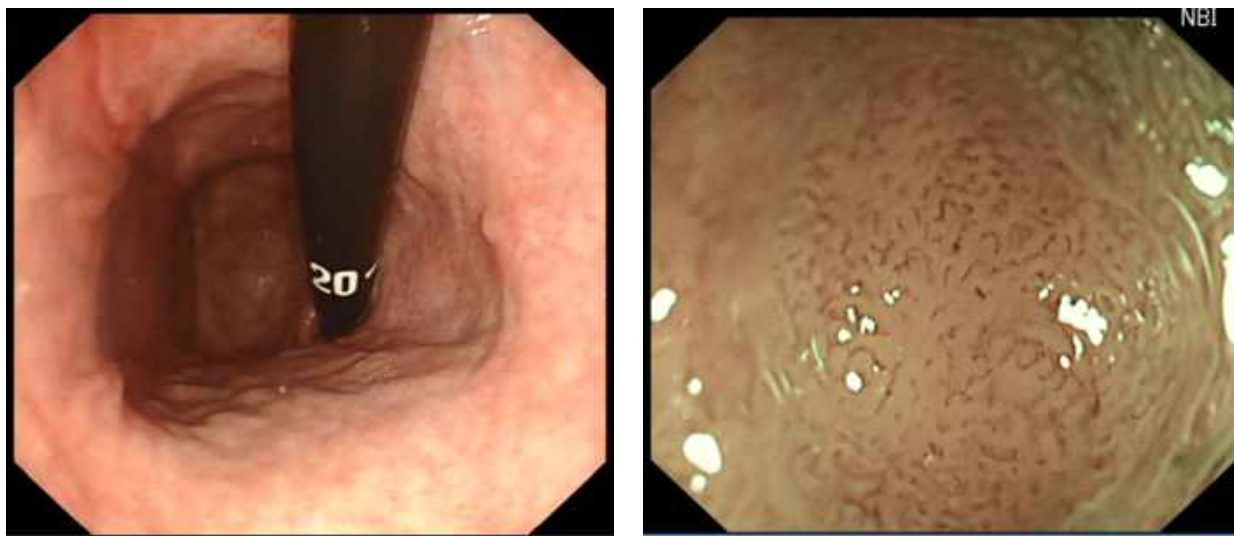
а) осмотр в белом свете

б) осмотр в режиме NBI

Рисунок 1.4 – Использование режима NBI для диагностики патологии по микрососудистому рисунку (стрелки указывают на область с патологией)

Другим примером технологий, широко используемых в эндоскопии, является режим с оптическим увеличением, достигающим 150 крат. При таком увеличении хорошо визуализируется микрососудистый рисунок и микроархитектоника поверхности слизистой оболочки, что приближает данную методику к оптической микроскопии [14, 29]. Особое значение данный режим имеет при осмотре пациентов с множественными участками, подозрительными на наличие патологии. В таких случаях невозможно осуществить отбор тканей путем биопсии из всех патологических очагов ввиду травматичности данной процедуры, и использование эндоскопии с увеличением позволяет снизить количество биопсий на 80-90%. При этом дополнительное использование режима NBI способствует лучшей

визуализации. Пример изображений в режиме с увеличением приведен на Рисунке 1.5.



а) осмотр без увеличения

б) осмотр с увеличением в 115 крат в режиме NBI

Рисунок 1.5 – Использование оптического увеличения

Таким образом, современный видеоэндоскоп представляет собой сложнейшее высокотехнологичное устройство, позволяющее получать изображения органов, выполнять забор биологического материала и осуществлять хирургические вмешательства, что делает его незаменимым инструментом в диагностике заболеваний ЖКТ.

### **1.3 Задача обнаружения полипов на эндоскопических изображениях**

Среди видов эндоскопии ЖКТ можно выделить два основных: гастроскопия и колоноскопия. При гастроскопии осуществляется визуальный осмотр, диагностика и лечение слизистой оболочки пищевода, желудка и двенадцатиперстной кишки. При колоноскопии врач осматривает и оценивает состояние внутренней поверхности толстой кишки при помощи эндоскопа. Колоноскопия даёт возможность визуально диагностировать

такие заболевания, как язвы, полипы и др., а также провести биопсию и удалить некоторые поражения. Колоноскопическое исследование сегодня – один из наиболее эффективных путей борьбы с колоректальным раком (КРР).

Колоректальный рак – группа злокачественных новообразований тканей эпителия толстого кишечника. В настоящее время колоректальный рак является одним из наиболее распространенных видов рака. Более того, в большинстве развитых стран количество случаев заболевания колоректальным раком увеличивается год от года, что связано с изменением типа питания и снижением физической активности населения. Согласно данным медицинских исследований, в 95% случаев рак толстой кишки развивается из полипов. Практически у трети пациентов заболевание диагностируется в запущенной стадии, что практически не оставляет шансов на благоприятный исход [9, 12, 104].

Поиск и удаление предраковых аномалий – один из наиболее эффективных способов борьбы с КРР. При этом особую значимость имеют полипы толстого кишечника, так как именно они могут перерасти в КРР на поздней стадии. Таким образом, раннее обнаружение полипов имеет решающее значение в борьбе с онкологией толстого кишечника.

Наряду с ведением здорового образа жизни профилактикой КРР является регулярный скрининг (обследование) толстой кишки. Различные исследования показывают, что скрининг в масштабах всего населения позволяет снизить частоту КРР или же улучшить прогноз, если заболевание уже развилось [27, 81]. Колоноскопия считается лучшим диагностическим инструментом для исследования толстой кишки, в том числе для раннего выявления и удаления полипов.

Полипы – аномальные разрастания ткани, выступающие над слизистой оболочкой. Они могут возникать в любом месте желудочно-кишечного тракта, но в основном обнаруживаются в колоректальной области и часто считаются предшественниками КРР [82, 87]. В толстом кишечнике выделяют несколько типов классификации полипов [33, 34]: по количеству, по



патоморфологическим данным, по особенностям прикрепления к стенке кишки (Рисунок 1.6).

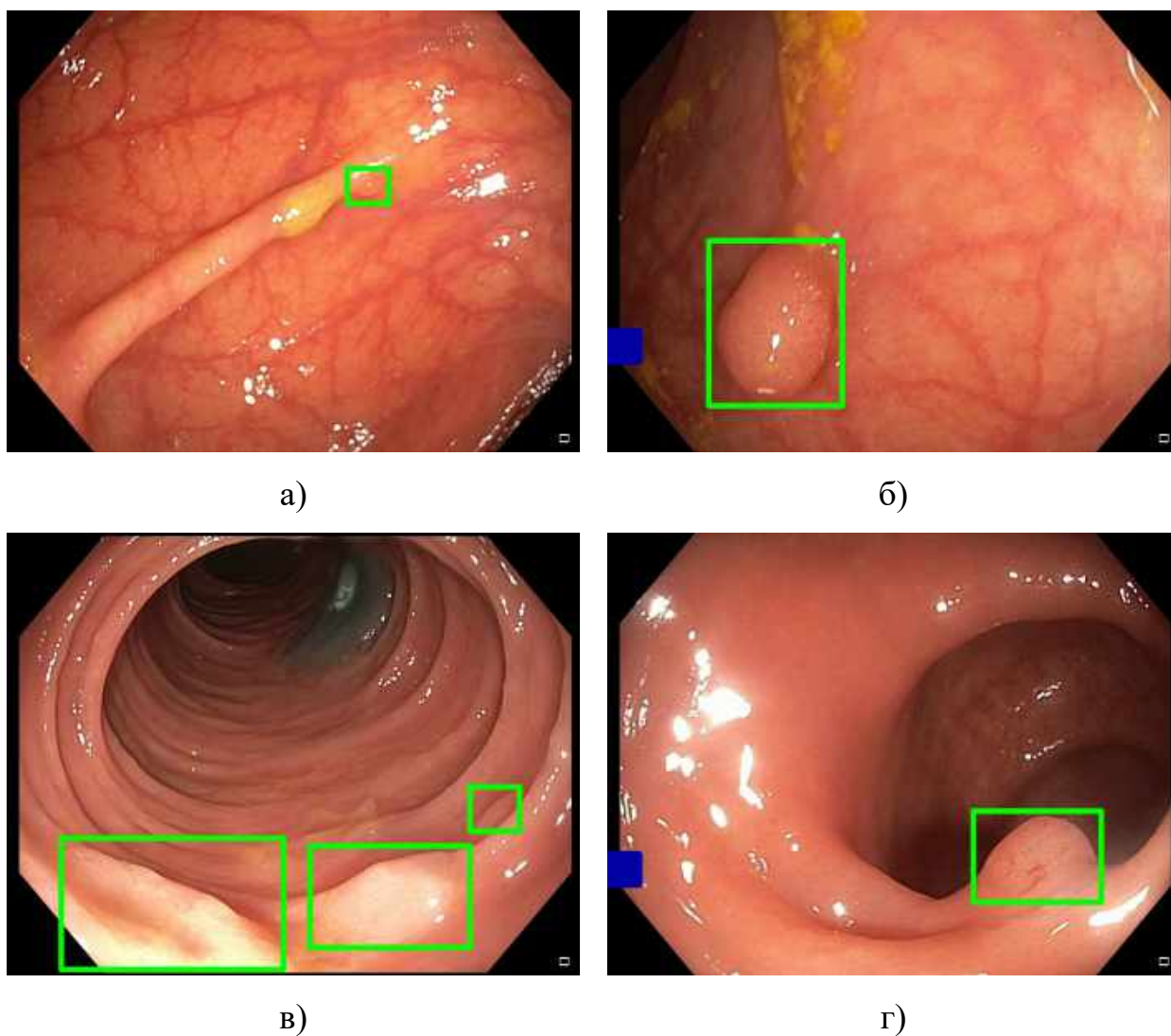


Рисунок 1.6 – Примеры изображений полипов

Размер колоректальных полипов может существенно различаться: от совсем небольших до весьма развитых ( $\geq 10$  мм) [102]. Обычно с помощью эндоскопического оборудования можно обнаружить и удалить относительно крупные полипы. Небольшие колоректальные полипы ( $\leq 5$  мм) являются весьма сложными объектами для обнаружения [125].

Колоноскопия – дорогостоящая и трудоемкая инвазивная процедура, которая требует качественной подготовки кишечника, а также наполнения кишечника воздухом во время обследования [99]. Тем не менее, в среднем



при обследовании не выявляется порядка 20% полипов. Таким образом, риск заболевания раком зависит от способности врача-эндоскописта обнаруживать полипы [91]. Недавние исследования показали, что новые эндоскопические устройства и диагностические инструменты улучшили частоту обнаружения полипов [51, 115]. Однако проблема пропущенных полипов остается по-прежнему актуальной.

Видеозаписи колоноскопических исследований, произведенных в клинических центрах, содержат в себе значительный объем данных. Однако собранные данные используются неэффективно, так как они являются весьма трудоемкими для обработки врачами-эндоскопистами [132]. Повторный просмотр видеозаписей исследований чаще всего не осуществляется, что значительно повышает риск пропуска патологии на ранней стадии. Автоматическая подборка и разметка видеоданных – необходимое условие для создания надежных систем компьютерной диагностики, которые позволят повысить качество клинической эндоскопии [57]. Отсутствие общедоступных наборов данных [140] является критическим узким местом в этой области, тормозящим процесс разработки алгоритмов автоматического анализа эндоскопических данных.

Создание баз эндоскопических видеоданных является весьма трудоемкой задачей и требует специальных знаний в предметной области. Как правило, при колоноскопии более мелкие полипы или плоские полипы, которые обычно упускаются из виду во время процедуры, бывает трудно обнаружить даже во время ручной разметки. Другой проблемой является высокая вариабельность изображений полипов: патологии бывают разных размеров, форм, текстур, цветов и ориентации [87]. Поэтому при разработке автоматизированных систем критически важно, чтобы все проблемы и нюансы, характерные для колоноскопии, принимались во внимание.

Современные системы автоматического обнаружения и сегментации полипов, основанные на глубоком машинном обучении, должны обладать высокой производительностью как в задаче анализа статических

изображений, так и при анализе видеоданных, полученных при проведении колоноскопического исследования [18, 88, 100, 103, 149]. Идеальная автоматическая система для обнаружения, локализации и сегментации полипов должна обеспечивать:

1. стабильную производительность и повышенную устойчивость к вариабельности изображений полипов, полученных от разных пациентов;
2. возможность работы в реальном времени, необходимую для клинического применения;
3. простоту в использовании, в том числе клинически интерпретируемые выходные данные.

Выполнение каждого из этих пунктов требует больших ресурсов и огромных затрат. Определённый прогресс в этой научно-технической области связан с появлением начиная с 2010 г. алгоритмов глубокого машинного обучения, позволяющих решать задачи анализа изображений на новом, прежде недостижимом уровне.

#### **1.4 Методы и алгоритмы обнаружения объектов на изображениях**

Обнаружение объектов – важная задача в областях цифровой обработки изображений и технического зрения, которая связана с обнаружением визуальных объектов определенного класса (таких, например, как люди, животные или автомобили) на изображениях. Целью процесса обнаружения объектов является разработка вычислительных методов и моделей, которые извлекают из изображений информацию о том, где и какие объекты находятся. Данная информация часто служит основой для решения других проблем технического зрения, например, в автоматических системах распознавания лиц. Как правило, для работы алгоритма идентификации по лицу на первом этапе необходимо найти лица на изображении. Обнаружение объектов также лежит в основе многих других задач технического зрения, таких как сегментация обнаруженных объектов

[54, 77-79], синтез текстового семантического описания изображения [94, 154, 155], трекинг (сопровождение) объектов [92] и т. д. Решение таких задач в настоящее время широко используется во многих практических приложениях, таких как автономное вождение, зрение роботов, видеонаблюдение, медицина и т. д.

Развитие подходов к обнаружению объектов можно разделить на два исторических периода: период классических алгоритмов (на основе вручную спроектированных признаков) до 2014 г. и период обнаружения объектов на основе методов глубокого обучения (на основе применения сверточных нейронных сетей (СНС) после 2014 г. На Рисунке 1.7 приведена история развития подходов к обнаружению объектов с основными ключевыми алгоритмами, здесь отмечены следующие алгоритмы: VJ [146, 147], HOG [56], DPM [62-64], R-CNN [72], SPPNet [80], Fast R-CNN [69], Faster R-CNN [130], YOLO [127], SSD [111], Pyramid Networks [107], Retina -Net [108].

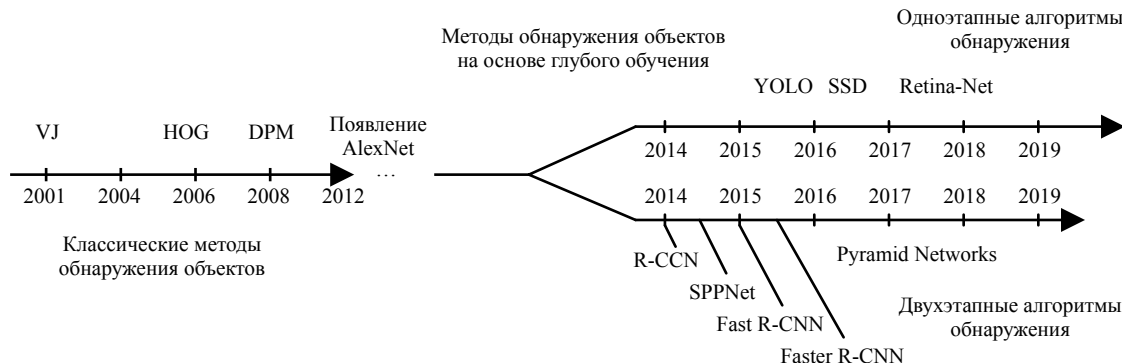


Рисунок 1.7 – Линия времени с ключевыми методами и алгоритмами обнаружения объектов

Прежде чем описать подробнее сами методы обнаружения объектов на изображениях, рассмотрим подходы к оцениванию качества работы алгоритмов данного класса.

### 1.4.1 Метрики оценки качества работы алгоритмов обнаружения объектов на изображениях

Вопрос оценки эффективности алгоритмов обнаружения объектов не является простым, и ответы на него модифицировались на разных этапах развития соответствующих технологий. На ранних этапах в научном сообществе не было единого подхода и общепринятых критериев оценки эффективности обнаружения. Например, в ранних исследованиях по обнаружению пешеходов [56] в качестве метрик обычно использовались коэффициент промахов ( $MR$ , выражение 1.1) и доля ложных срабатываний на окно ( $FPPW$ , выражение 1.2), а также кривая зависимости коэффициента промахов от доли ложных срабатываний на окно как комплексная оценка эффективности алгоритма обнаружения объектов. Однако оценка ложных срабатываний для каждого окна может иметь недостатки и в некоторых случаях не может корректно оценить эффективность алгоритма в целом [59]. В 2009 г. предложен новый протокол тестирования алгоритмов обнаружения пешеходов от Калифорнийского технологического института [58, 59], и с тех пор метрика доли ложных срабатываний на окно изменилась на долю ложных срабатываний на изображение ( $FPPI$ , выражение 1.3)

$$MR = \frac{FN}{TP + FN}, \quad (1.1)$$

где  $FN$  – число ложноотрицательных результатов, а  $TP$  – число истинно положительных результатов;

$$FPPW = \frac{FP}{W}, \quad (1.2)$$

где  $FP$  – число ложноположительных результатов, а  $W$  – суммарное количество проанализированных алгоритмом окон (многие алгоритмы того времени использовали метод скользящего окна для обнаружения объектов);

$$FPPI = \frac{FP}{I}, \quad (1.3)$$

где  $I$  – число проанализированных алгоритмом изображений.

В последние годы наиболее часто используемой метрикой для алгоритмов обнаружения объектов является средняя точность, которая первоначально была представлена в исследовании VOC2007 [60]. Она определяется как усреднение множества значений точности обнаружения, определяемых выражением (1.4), соответствующих различным значениям полноты, рассчитываемым по выражению (1.5), и обычно оценивается для каждого класса объектов независимо.

Точность (Precision,  $P$ ):

$$P = \frac{TP}{TP + FP}. \quad (1.4)$$

Полнота или чувствительность (Recall,  $R$ ):

$$R = \frac{TP}{TP + FN}. \quad (1.5)$$

Средняя точность (Average Precision,  $AP$ ):

$$AP = \frac{\sum_{r=1}^N P_r}{N}, \quad (1.6)$$

где  $P_r$  – значение точности, соответствующее определенному значению полноты  $r$ , а  $N$  – множество различных значений полноты. Оно получается путем варьирования порога оценки достоверности в оцениваемом алгоритме обнаружения. Алгоритмы обнаружения объектов предоставляют оценку достоверности для каждого предсказанного обнаружения, причем данная оценка может формироваться различным образом для разных алгоритмов. Варьирование порога оценки достоверности позволяет менять соотношение полноты и точности для алгоритма обнаружения объектов.

Для сравнения производительности по всем категориям объектов в качестве конечной метрики оценки качества обычно используется среднее значение  $AP$  (mean Average Precision,  $mAP$ ), усредненное по всем категориям объектов

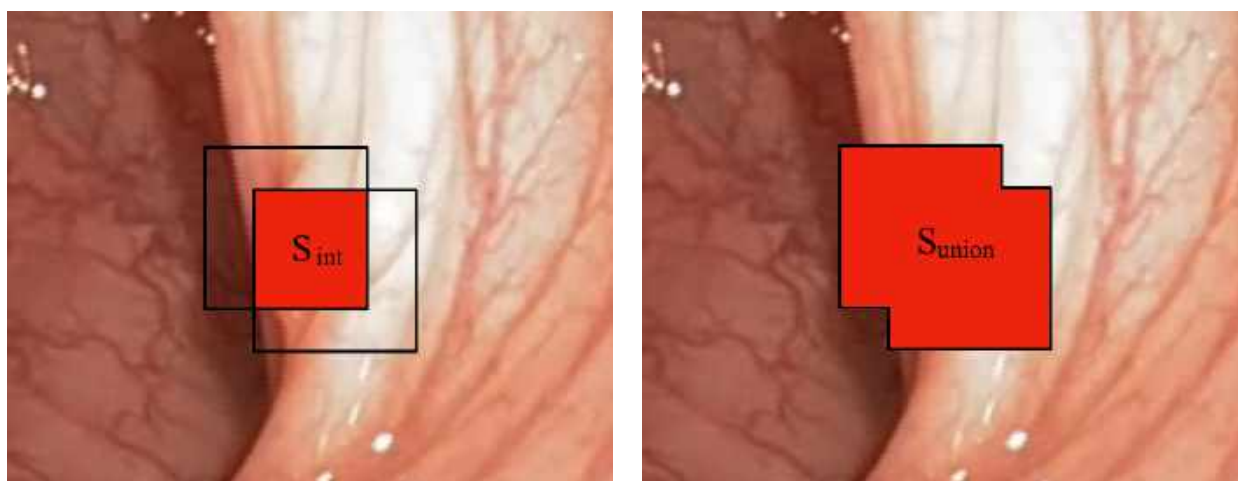
$$mAP = \frac{\sum_{i=1}^K AP_i}{K}, \quad (1.7)$$

где  $K$  – число классов (категорий объектов). Если алгоритм применяется для обнаружения объектов только одного класса, то используется обычная метрика  $AP$ .

Для того чтобы отнести обнаружения к одному из возможных вариантов:  $TP$ ,  $FP$  или  $FN$ , используется отношение пересечения к объединению (Intersection over Union,  $IoU$ ):

$$IoU = \frac{S_{int}}{S_{union}}, \quad (1.8)$$

где  $S_{int}$  – площадь пересечения между экспертной разметкой и обнаружением, предсказанным алгоритмом (Рисунок 1.8 а), а  $S_{union}$  – площадь объединения между экспертной разметкой и обнаружением, предсказанным алгоритмом (Рисунок 1.8 б).



а)

б)

Рисунок 1.8 – Визуализация метрики  $IoU$ :

а) площадь пересечения, б) площадь объединения

Если метрика  $IoU$  между предсказанной ограничивающей рамкой и зоной экспертной разметки превышает заданный порог (например, 0,5), тогда объект будет идентифицирован как истинно положительный, в противном случае будет идентифицирован как ложноотрицательный. Таким образом,

именно метрика  $mAP$ , рассчитанная при пороге  $IoU=0,5$  ( $mAP@0,50$ ), стала общепринятой для оценки алгоритмов обнаружения объектов.

После 2014 г. из-за популярности наборов данных MS-COCO [109] исследователи стали уделять больше внимания точности определения местоположения ограничительной рамки на изображении. В связи с этим наряду с  $mAP@0,50$  общепринятой для оценки алгоритмов обнаружения объектов также становится метрика  $mAP$ , усреднённая по ряду значений  $IoU$  от 0,5 до 0,95 с шагом 0,05. Такая метрика  $mAP@[0,5..0,95]$  позволила лучше оценивать точность локализации объекта на изображении и может иметь большое значение для некоторых популярных практических приложений (например, техническое зрение).

Другой популярной метрикой оценки качества алгоритмов обнаружения объектов на изображении является среднее гармоническое точности и полноты ( $F1$ ):

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}. \quad (1.9)$$

Упомянутые метрики (1.4-1.9) используются для сравнения различных алгоритмов обнаружения объектов друг с другом на стандартизованных наборах изображений (датасетах) с заранее размеченными объектами.

#### **1.4.2 Классические алгоритмы обнаружения объектов на изображениях**

Большинство ранних алгоритмов обнаружения объектов строились на основе вручную спроектированных признаков, описывающих изображение. Из-за отсутствия эффективного представления изображений в то время, научное сообщество было сосредоточено на разработке сложных признаковых представлений и различных подходов к ускорению соответствующих алгоритмов, чтобы применять их в условиях ограниченных вычислительных ресурсов.

Одним из самых значимых в истории обнаружения объектов алгоритмом является алгоритм обнаружения лиц на изображениях, представленный в 2001 г. Р. Viola и М. Jones [146, 147]. Авторы достигли скорости обнаружения человеческих лиц, позволяющей алгоритму работать в режиме реального времени без каких-либо ограничений. Реализованный на процессоре Pentium III с частотой 700 МГц алгоритм работал в десятки или даже сотни раз быстрее, чем любые другие алгоритмы того времени при сопоставимой точности обнаружения. Данный алгоритм обнаружения, позже названный «алгоритмом Виолы-Джонса», на линии времени (Рисунок 1.7) обозначен VJ – инициалами авторов в память об их значительном вкладе в рассматриваемую область.

Алгоритм VJ подходит к задаче обнаружения наиболее прямолинейно, в его основе лежит метод скользящего окна. Он анализирует все возможные местоположения и масштабы на изображении, чтобы обнаружить, содержит ли какое-либо окно человеческое лицо. Для своего времени алгоритм VJ значительно улучшил скорость обнаружения лиц на изображениях за счет использования трех важных подходов: **«интегральное изображение»**, **«выбор признаков»** и **«каскадная структура»**.

- **Интегральное изображение** – вычислительный метод, позволяющий ускорить фильтрацию фрагментов изображений или процесс свертки. Как и другие алгоритмы обнаружения объектов того времени [119, 121 122], алгоритм VJ использует вейвлет Хаара в качестве признаков для представления изображения. Интегральное изображение делает вычислительную сложность обработки каждого окна в алгоритме VJ независимой от размера этого окна.
- **Выбор признаков**: вместо использования набора вручную выбранных базовых фильтров Хаара авторы использовали алгоритм Adaboost [66] для выбора небольшого набора признаков, наиболее полезных для обнаружения лиц, из огромного набора (около 180 тыс.) случайных пулов признаков.



- **Каскадная структура:** процесс многоступенчатого (каскадного) обнаружения введен в алгоритме VJ, чтобы уменьшить его вычислительные затраты за счет уменьшения количества вычислений, приходящихся на окна, не содержащие лиц, и увеличение – на окна с лицами. Структура каскадного алгоритма приведена на Рисунке 1.9.

Таким образом, алгоритм VJ для обнаружения лиц на изображении примерно на десятилетие стал самым практически используемым и стимулировал активное развитие других методов и подходов в рассматриваемой области.

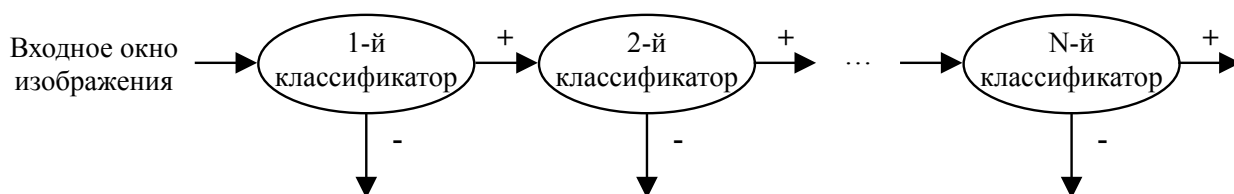


Рисунок 1.9 – Структура каскадного алгоритма обнаружения объектов

Вторым знаковым событием в области стала разработка алгоритма обнаружения объектов, основанного на гистограммах ориентированных градиентов. Соответствующий дескриптор признаков (Histogram of Oriented Gradients, HOG) первоначально предложен в 2005 г. N. Dalal и В. Triggs [56]. Его можно рассматривать как важное улучшение масштабно-инвариантного преобразования признаков (scale-invariant feature transform, SIFT) [112, 113] и контекстов формы [44]. Чтобы сбалансировать инвариантность признаков (включая сдвиги, масштаб, освещение и др.) и нелинейность (при разделении различных категорий объектов), HOG-дескриптор вычислялся на плотной сетке из равномерно расположенных ячеек, а для повышения точности на блоках применялась локальная нормализация контраста. Алгоритм на основе HOG-дескриптора можно использовать для обнаружения различных классов объектов, но изначально он разрабатывался для решения задачи обнаружения пешеходов на изображении. Для обнаружения объектов разного размера в

данном алгоритме многократно изменяется масштаб входного изображения, сохраняя при этом размер окна обнаружения. Данный подход долгое время был важной основой для многих других алгоритмов обнаружения объектов [62, 64, 114] и множества приложений технического зрения.

Пиком развития классических алгоритмов обнаружения объектов стал алгоритм на основе «Модели деформируемых частей» (Deformable Part-based Model, DPM). Данный подход стал победителем соревнований по обнаружению объектов VOC-07, VOC-08 и VOC-09.

Алгоритм на основе DPM следует идее «разделяй и властвуй», где обучение можно рассматривать просто как изучение правильного способа разложения объекта, а процесс предсказания – как совокупность обнаружений различных частей объекта. Например, задачу обнаружения объекта класса «автомобиль» можно рассматривать как последовательное обнаружение его окна, кузова и колес. Эта часть работы, также известная как «звездная модель», предложена Р. Felzenszwalb и соавторами в [64]. Позже R. Girshick расширил звездную модель до «смеси моделей» [62, 63, 70, 73], чтобы улучшить работу алгоритма при работе с объектами реального мира, имеющими большую внутриклассовую вариативность.

Алгоритм DPM состоит из корневого фильтра и нескольких фильтров частей. Вместо того, чтобы вручную задавать конфигурации фильтров частей (например, размер и местоположение), в DPM разработан слабо контролируемый метод обучения, где все конфигурации фильтров частей могут быть изучены автоматически как скрытые переменные. Позже R. Girshick сформулировал этот процесс как частный случай многоэкземплярного обучения [40]. Кроме того, для повышения точности обнаружения здесь применяются и некоторые другие важные методы, получившие широкое распространение в других подходах, такие как отбор сложных отрицательных примеров (hard negative mining), регрессия ограничивающей рамки (bounding box regression) и контекстное заполнение (context priming). Чтобы ускорить работу алгоритма, R. Girshick разработал

метод «компиляции моделей обнаружения» в гораздо более быструю модель, которая реализует каскадную архитектуру, которая обеспечивает ускорение более чем в 10 раз без ущерба для точности [62, 70].

Хотя сегодняшние методы обнаружения объектов намного превзошли алгоритм DPM с точки зрения точности обнаружения, многие из разработанных позже методов находились под сильным влиянием заложенных в данном подходе идей.

### **1.4.3 Двухэтапные алгоритмы обнаружения объектов на изображениях на основе сверточных нейронных сетей**

Нейросетевые алгоритмы обнаружения объектов, захватившие внимание исследователей в последние 10 лет, разделяются на две основные категории: «**двухэтапные алгоритмы обнаружения**» и «**одноэтапные алгоритмы обнаружения**» объектов. Начнем рассмотрение с двухэтапных алгоритмов, появившихся несколько раньше.

По мере того как производительность классических методов обнаружения объектов достигла предела, после 2010 г. прогресс в этой области заметно замедлился [70]. В 2012 г. в научном сообществе сильно возрос интерес к сверточным нейронным сетям [98]. В 2014 г. R. Girshick предложил новый подход к обнаружению объектов на основе модели регионов с признаками генерируемых сверточной нейронной сетью (Regions with CNN features, R-CNN) [71, 72]. С тех пор область обнаружения объектов на изображениях начала развиваться с беспрецедентной скоростью.

Идея, лежащая в основе алгоритма R-CNN, заключается в извлечении набора объектов-кандидатов на первом этапе путем выборочного поиска [136]. Затем на втором этапе каждый фрагмент изображения с объектом-кандидатом масштабируется до изображения фиксированного размера и передается в СНС без слоев классификации в конце, обученную на наборе изображений ImageNet [133] (например, на сеть архитектуры AlexNet [98]) для извлечения признаков. Полученные признаки используются далее для

классификации с помощью метода опорных векторов [53]. Схема алгоритма R-CNN приведена на Рисунке 1.10.

Алгоритм R-CNN показал значительный прирост в качестве распознавания на соревновании VOC-07. Метрика  $mAP@0,50$  прошлого лидера – алгоритма DPM-v5 – равнялась 33,7%, в то время как алгоритм R-CNN достиг значения 58,5%. Несмотря на это, данный подход имел существенный недостаток: избыточное вычисление признаков для большого количества перекрывающихся фрагментов изображения (более 2000 шт. на одном изображении) приводили к чрезвычайно низкой скорости работы алгоритма (примерно 14 с на анализ изображения при использовании имеющихся на тот момент графических процессоров – GPU).

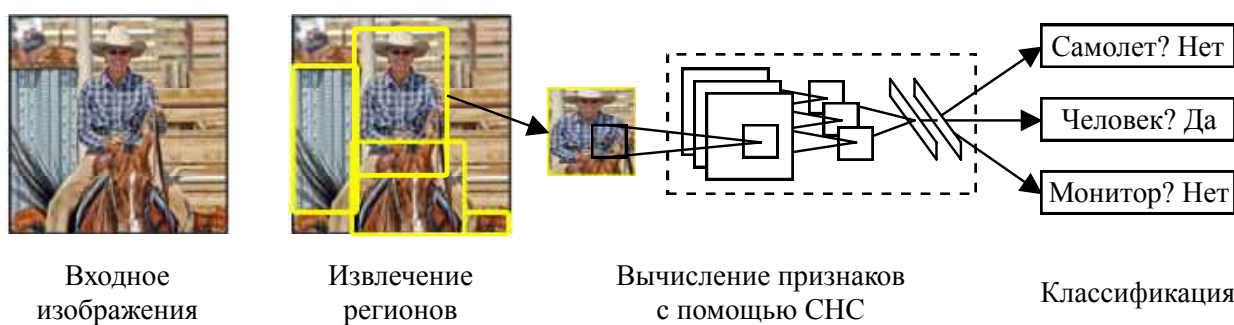


Рисунок 1.10 – Схема алгоритма R-CNN

В том же 2014 г. предложена нейронная сеть SPPNet [80], в которой эта проблема уже решалась. Не К. и др. предложили сети с моделью объединения пространственных пирамид (Spatial Pyramid Pooling Networks, SPPNet) [80]. Используемые до этого модели СНС требовали подачи входных данных фиксированного размера, например, изображения разрешением  $224 \times 224$  пикселя для архитектуры AlexNet [98]. Основным вкладом разработчиков SPPNet является введение слоя нейронной сети Spatial Pyramid Pooling (SPP), который позволяет ей генерировать представление фиксированной длины независимо от размера входного изображения или его фрагмента без масштабирования. При использовании SPPNet в задаче

обнаружения объектов, карты признаков могут быть вычислены из всего изображения только один раз, а затем могут быть сгенерированы представления фиксированной длины для произвольных областей изображения для обучения алгоритма обнаружения, что позволяет избежать повторного вычисления признаков с помощью СНС. Установлено, что сеть SPPNet работает более чем в 20 раз быстрее, чем сеть R-CNN, без потери точности в качестве обнаружения.

Хотя нейронная сеть SPPNet существенно увеличила скорость обнаружения, у данного подхода также есть некоторые недостатки: во-первых, обучение по-прежнему являлось многоэтапным, во-вторых, в SPPNet полносвязные слои обучаются отдельно от остальных слоев.

В 2015 г. R. Girshick предложил архитектуру Fast R-CNN [69], которая является дальнейшим усовершенствованием R-CNN и SPPNet [72, 80], позволяющую одновременно обучать и блок обнаружения, и блок регрессии ограничивающей рамки. В наборе данных VOC-07 сеть архитектуры Fast R-CNN увеличила показатель  $mAP@0,50$  с 58,5% (R-CNN) до 70,0%, имея при этом скорость обнаружения более чем в 200 раз выше, чем у базовой сети R-CNN. Хотя архитектура Fast-R-CNN успешно объединяла преимущества R-CNN и SPPNet, скорость обнаружения все еще ограничивалась поиском объектов-кандидатов на первом этапе работы.

В 2015 г. S. Ren и др. предложили архитектуру Faster R-CNN [130, 131] – первый алгоритм на основе глубокого обучения, состоящий только из модели СНС, и первый, работающий практически в режиме реального времени (показатели метрик на ключевых датасетах: COCO  $mAP@0,50=42,7\%$ , COCO  $mAP@[0,5..0,95]=21,9\%$ , VOC-07  $mAP@0,50=73,2\%$ , VOC-12  $mAP@0,50=70,4\%$ , скорость обработки 17 кадров в секунду с ZF-Net [157]). Основным вкладом архитектуры Faster R-CNN является введение блока Region Proposal Network (RPN), позволяющего генерировать объекты-кандидаты с небольшими вычислительными затратами. При эволюции архитектуры от R-CNN до Faster R-CNN

большинство отдельных блоков алгоритма обнаружения объектов, например, обнаружение объектов-кандидатов, извлечение признаков, регрессия ограничивающей рамки и др., постепенно интегрировались в унифицированную сквозную нейросетевую структуру.

Несмотря на успехи архитектуры Faster R-CNN в первом из двух этапов работы, на втором этапе по-прежнему сохраняется избыточность вычислений. Позднее было предложено множество усовершенствований, в том числе архитектуры R-FCN [55] и R-CNN Lighthouse [105].

В 2017 г. Т.-У. Lin и др. предложили нейросетевую архитектуру Feature Pyramid Networks (FPN) [107] на основе Faster R-CNN. До этого момента большинство алгоритмов на основе глубокого обучения использовали для обнаружения признаки, полученные только с верхнего уровня сети, отвечающего за формирования признаков. Эти признаки плохо подходят для локализации объектов и, фактически, пропускают маленькие объекты на изображении. С этой целью в архитектуре FPN разработана нисходящая модель с побочными связями для агрегирования признаков, хорошо описывающих объекты разного размера. Поскольку СНС естественным образом формирует пирамиду признаков посредством своего прямого распространения, архитектура FPN демонстрирует хорошие результаты в обнаружении объектов с широким разнообразием масштабов. Модель сети FPN существенно улучшает результат обнаружения на стандартном наборе изображений MS-COCO ( $\text{COCO } mAP@0,50=59,1\%$ ,  $\text{COCO } mAP@[0,5..0,95]=36,2\%$ ). Она в свою очередь стала центральным компонентом многих современных нейросетевых подходов к решению задачи обнаружения объектов на изображениях.

#### **1.4.4 Одноэтапные алгоритмы обнаружения объектов на изображениях на основе сверточных нейронных сетей**

Первый одноэтапный алгоритм обнаружения объектов на основе нейронной сети YOLO (You Only Look Once) предложил R. Joseph и др. в

2015 г. [127]. Алгоритм YOLO обладает высокой скоростью анализа изображения. Было предложено две модификации алгоритма: одна более быстрая, другая более точная. Быстрая модификация YOLO достигает скорости обработки 155 кадров в секунду (при использовании современных GPU) с  $mAP@0,50=52,7\%$  на наборе данных VOC-07, а его более точная модификация работает со скоростью 45 кадров в секунду с  $mAP@0,50=63,4\%$  на наборе VOC-07 и  $mAP@0,50=57,9\%$  на наборе VOC-12.

Уже из названия нейросетевой архитектуры YOLO видно, что авторы сделали упор на подход, противоположный парадигме двухэтапных алгоритмов. Вместо этого предложенный алгоритм следует совершенно другой идее: применить одну нейронную сеть ко всему изображению за один раз. Данный подход заключается в том, чтобы разделить изображение на фиксированные неперекрывающиеся блоки и предсказывать ограничивающие рамки и классы в каждом блоке параллельно. Идея данного подхода проиллюстрирована на Рисунке 1.11. Позже R. Joseph сделал ряд улучшений для базового алгоритма YOLO и предложил его версии v2 и v3 [128, 129], которые еще больше повысили точность обнаружения, сохраняя при этом высокую скорость обработки изображения, позволяющую его практическое использование в режиме реального времени.

Несмотря на увеличение скорости обнаружения, ранние версии данного подхода имеют точность локализации ниже, чем двухэтапные алгоритмы, особенно для небольших объектов. В последующих версиях YOLO [128, 129] этой проблеме уделялось больше внимания. Значительно позже, почти одновременно от различных групп авторов появились модификации этого подхода YOLOv4 [49] и YOLOv5 [90], которые превзошли двухэтапные алгоритмы как по точности работы, так и по скорости анализа изображения. Данные алгоритмы более подробно рассматриваются во второй главе настоящей работы.

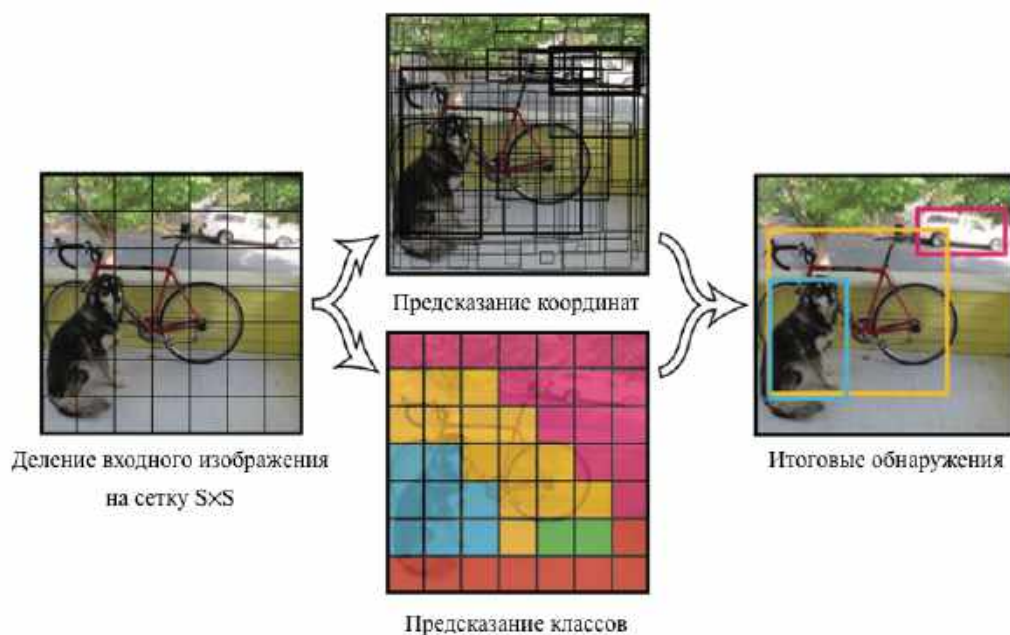


Рисунок 1.11 – Визуализация принципа работы алгоритма YOLO

В 2015 г. W. Liu и др. предложили нейросетевую архитектуру Single Shot MultiBox Detector (SSD) [111] – второй одноэтапный алгоритм обнаружения. Основным вкладом SSD является введение методов обнаружения с несколькими эталонами и несколькими разрешениями, что значительно повышает точность обнаружения одноэтапного алгоритма, особенно для некоторых небольших объектов. Архитектура SSD имеет преимущества как в скорости работы, так и в точности (метрики на типовых датасетах: VOC-07  $mAP@0,50=76,8\%$ , VOC-12  $mAP@0,50=74,9\%$ , COCO  $mAP@0,50=46,5\%$ ,  $mAP@[0,5..0,95]=26,8\%$ , при скорости обработки 59 кадров в секунду) по сравнению с ранними версиями YOLO. Подробнее данный подход также рассматривается во второй главе.

Несмотря на свою высокую скорость и простоту, одноэтапные алгоритмы длительное время уступали по точности двухэтапным алгоритмам. В 2017 г. Т.-У. Lin и др. изучили причины такого положения одноэтапных подходов и предложили с учетом этого новую архитектуру RetinaNet [108]. Они утверждали, что главной причиной проблемы точности является крайний дисбаланс классов объектов интереса и фона, возникающий при обучении алгоритмов с предсказанием по плотной сетке



возможных местоположений объектов. Для исправления этого недостатка в RetinaNet ввели новую функцию потерь, называемую «фокусной потерей» (англ. focal loss). Она заменяла стандартную функцию кросс-энтропии, чтобы алгоритм во время обучения уделял больше внимания сложным, неправильно классифицированным примерам. Модифицированная функция потерь позволяет одноэтапным алгоритмам достигать точности, сравнимой с двухэтапными алгоритмами, при сохранении высокой скорости работы (метрики на типовых датасетах: COCO  $mAP@0,50=59,1\%$ ,  $mAP@[0,5..0,95]=39,1\%$ ).

Таким образом, из приведенного выше анализа научно-технической литературы можно сделать вывод о том, что именно одноэтапные алгоритмы обнаружения объектов на основе СНС имеют большой потенциал для решения поставленной задачи анализа эндоскопических видеоизображений в режиме реального времени.

### **1.5 Краткие выводы**

Результаты проведенного анализа существующих задач, методов и алгоритмов в области построения систем прикладного телевидения в эндоскопии позволяют сделать следующие основные выводы:

- медицинские СПТ являются перспективными с точки зрения эффективного внедрения методов машинного обучения и цифровой обработки изображений. В частности, одной из актуальных задач в данной области является разработка алгоритмов обнаружения патологий на эндоскопических видеоизображениях;
- в результате приведенного анализа алгоритмов обнаружения объектов на изображениях можно сделать вывод о необходимости разработки программно-аппаратных систем на основе методов глубокого машинного обучения и СНС. Так, например, практически важной является задача разработки алгоритмов обнаружения полипов на видеоизображениях колоноскопических исследований кишечника;

- одним из технологических барьеров является высокая вычислительная сложность многих современных нейросетевых алгоритмов, что позволяет применять алгоритмы анализа видеоданных в режиме реального времени только на высокопроизводительных графических процессорах или даже специализированных суперкомпьютерах. Поэтому возникает задача нахождения баланса соотношения «точность/производительность», необходимая для применения на практике алгоритмов обработки видеоданных в эндоскопических СПТ;
- другим технологическим барьером является тот факт, что для обучения алгоритмов на основе СНС необходимо использование больших (десятки тысяч) экспертно размеченных баз изображений, что в области задачи обнаружения полипов на эндоскопических изображениях является актуальной задачей. Еще большей проблемой является отсутствие доступных баз видеоданных с экспертной разметкой местоположений полипов, при том, что анализ видеопотока – существенно более сложная задача по сравнению с анализом статических изображений;

Таким образом, проведенный анализ позволил сформулировать следующие основные задачи диссертации:

- разработка и исследование нейросетевых алгоритмов обнаружения объектов на видеоизображениях эндоскопических исследований;
- разработка методики полуавтоматического создания размеченной базы эндоскопических видеоданных для обучения нейросетевых моделей;
- усовершенствование алгоритмов предобработки и постобработки видеоизображений, учитывающих специфику видеоэндоскопических данных;
- тестирование разработанного алгоритма для анализа эндоскопического видеопотока в реальной медицинской системе прикладного телевидения.

## ГЛАВА 2

# ИССЛЕДОВАНИЕ АЛГОРИТМОВ ОБНАРУЖЕНИЯ ПОЛИПОВ НА ЭНДСКОПИЧЕСКИХ ИЗОБРАЖЕНИЯХ КИШЕЧНИКА

### 2.1 Вводные замечания

Проблема автоматического обнаружения полипов на эндоскопических изображениях является актуальной для специалистов, разрабатывающих СПТ в медицине. В ранних научно-технических работах особое внимание уделялось алгоритмам машинного обучения, полагающимся на ручной подбор дескрипторов, основанных на таких признаках как цвет и текстура полипов [39, 93]. Позднее методы, основанные на СНС, привлекли значительное внимание [138, 143] и стали часто применяться при анализе общедоступных наборов данных [38, 45].

Wang и др. [151] разработали алгоритмы и программные модули для быстрого обнаружения краев полипов и самих полипов, в том числе программную систему оповещения о полипах. Shin и др. [139] использовали СНС на основе классификации регионов для автоматического обнаружения полипов на видеоданных и изображениях колоноскопических исследований. Они применяли предварительно обученную сеть Inception ResNet для трансферного обучения и методы постобработки для надежного обнаружения полипов в колоноскопических исследованиях. Позже Shin и др. [140] использовали генеративную состязательную сеть GAN [74], где они показали, что сгенерированные изображения полипов не являются качественно реалистичными, но в то же время они могут помочь повысить эффективность обнаружения. Lee и др. [103] использовали архитектуру YOLOv2 [127, 128] для разработки алгоритма обнаружения и локализации полипов. Алгоритм обеспечивает высокую чувствительность и производительность, близкую к режиму реального времени. Yamada и др.

[156] разработали систему искусственного интеллекта, которая может автоматически определять признаки рака толстой кишки во время колоноскопического исследования с высокой чувствительностью и специфичностью.

В дополнение к работам, связанным с автоматическим обнаружением, попиксельная классификация (сегментация) патологий позволяет найти точную границу полипа и, следовательно, также имеет большое значение для клинического наблюдения и соответствующих процедур. Bernel и др. [45] представили результаты решения задачи автоматического обнаружения полипов, которая была частью соревнования по техническому зрению на конференции Medical Image Computing and Computer Assisted Intervention (MICCAI) 2015. В этой работе сравнивалась эффективность восьми алгоритмов, применяемых к предоставленным тестовым наборам изображений полипов. Wang и др. [149] предложил алгоритм на основе архитектуры SegNet [41], которая имела производительность, достаточную для работы в режиме реального времени со скоростью обработки более чем 25 кадров в секунду. Guo и др. [75] использовали полностью сверточные сети дилатации на наборе данных сегментации полипов Gastrointestinal Image ANalysis (GIANA). Jha и др. [87] предложили использовать архитектуру ResUNet ++, демонстрирующую улучшение на 10% по сравнению с широко известной базовой архитектурой сети UNet на наборе данных Kvasir-SEG. Они также применили обученную модель к набору данных CVC-ClinicDB [46], показав улучшение более чем на 15% по сравнению с UNet. Али и др. [38] провели всестороннюю оценку как алгоритмов обнаружения, так и сегментации артефактов, присутствующих при клинической эндоскопии, включая данные колоноскопии [37]. Wang и др. [150] предложили нейронную сеть с учетом границ (BA-Net) для сегментации медицинских изображений. Позже Jha и др. [89] предложили архитектуру DoubleUNet для сегментации, которая применялась к четырем наборам биомедицинских изображений. Такая нейронная сеть представляла собой комбинацию двух расположенных

друг над другом UNet с некоторыми дополнительными блоками. Предложенный алгоритм достиг наилучших экспериментальных результатов на наборах изображений полипов CVC-Clinic и ETIS-Larib. Подобные исследования были также продолжены в работах [42, 116, 118, 135].

Таким образом, согласно данным, приведенным в научно-технической литературе, на сегодняшний день наиболее эффективным подходом к обнаружению полипов на эндоскопических изображениях является использование СНС. Обнаружение и локализация полипов обычно имеют решающее значение во время проведения диагностической колоноскопической процедуры, в то время как пиксельная сегментация становится жизненно важной для автоматизации определения границ полипа во время хирургических процедур [88].

Задачей данного исследования является обнаружение полипов при проведении диагностических колоноскопических исследований, поэтому в дальнейшем рассматриваются только алгоритмы обнаружения, прогнозирующие координаты рамки, внутри которой находится полип. Согласно анализу, проведенному в п. 1.4, для построения нейросетевого алгоритма будем рассматривать одноэтапные методы обнаружения полипов.

## **2.2 Алгоритмы обнаружения полипов на основе одноэтапных нейросетевых методов**

В общем случае алгоритмы обнаружения на базе СНС состоят из входа, основы (экстрактора признаков), агрегатора признаков и выхода. Входными данными могут быть изображения, их фрагменты или пирамиды изображений. Экстрактором могут быть разные архитектуры СНС, такие как VGG16, ResNet50, ResNext-101, Darknet и др. Агрегатор представляет собой такую сеть, которая собирает карты признаков с разных слоев базовой сети. Выходные слои используются для обработки предсказываемых ограничивающих рамок, которые могут быть одноэтапным алгоритмом для предсказания по плотной сетке координат (например, YOLO, RPN и

RetinaNet [108]) и двухэтапным алгоритмом с разреженным предсказанием (например, Faster R-CNN [130] и R-FCN [55]). В последнее время одноэтапные методы привлекают повышенное внимание исследователей из-за своей скорости и способности обеспечивать высокую точность. Это достигается за счет использования сети пирамиды признаков или уровней объединения пространственных пирамид для предсказания возможных ограничивающих рамок, как показано на Рисунке 2.1.

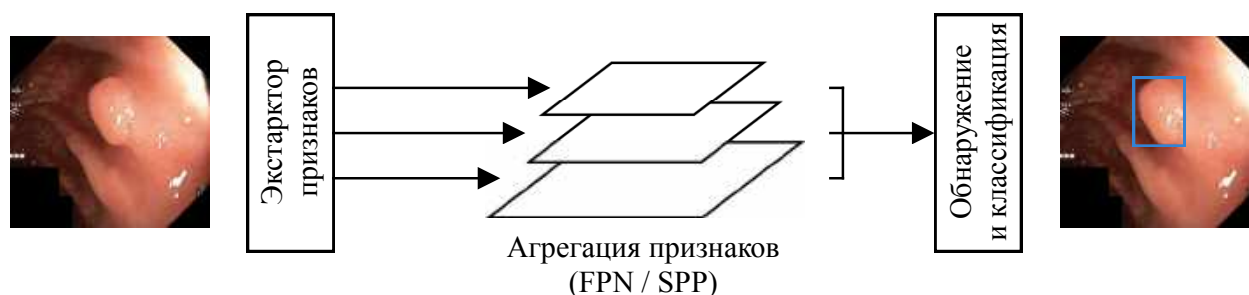


Рисунок 2.1 – Упрощенная архитектура типичного одноэтапного метода обнаружения объектов на основе СНС

В недавно вышедшей работе [88] исследователи используют следующие архитектуры СНС для обнаружения полипов: EfficientDet-D0 [144, 145], Faster R-CNN [130], RetinaNet50 [108], RetinaNet101 [108], YOLOv3+spp [129], YOLOv4 [49] и ColonSegNet [88] и делают вывод, что наиболее подходящим по соотношению точность/производительность является алгоритм YOLOv4. Однако авторы не рассматривают популярный алгоритм, часто дающий хорошие результаты на различных задачах обнаружения объектов – SSD [30, 95, 96, 101, 111], а также недавно появившуюся модификацию одноэтапного алгоритма – YOLOv5 [21, 90]. Рассмотрим эти два нейросетевых алгоритма обнаружения более подробно.

### 2.2.1 Архитектура сверточной нейронной сети SSD

Сверточная нейронная сеть SSD представляет собой сеть прямого распространения, состоящую из нескольких блоков: базовая сеть для извлечения признаков из изображения (может выбираться любая сеть, показывающая высокое качество классификации), собственные сверточные слои более низкой размерности, слои обнаружения и классификации, на которые подается информация с карт признаков различного разрешения, и алгоритм подавления немаксимумов для определения итоговых координат объектов. Архитектура данной сети представлена на Рисунке 2.2.

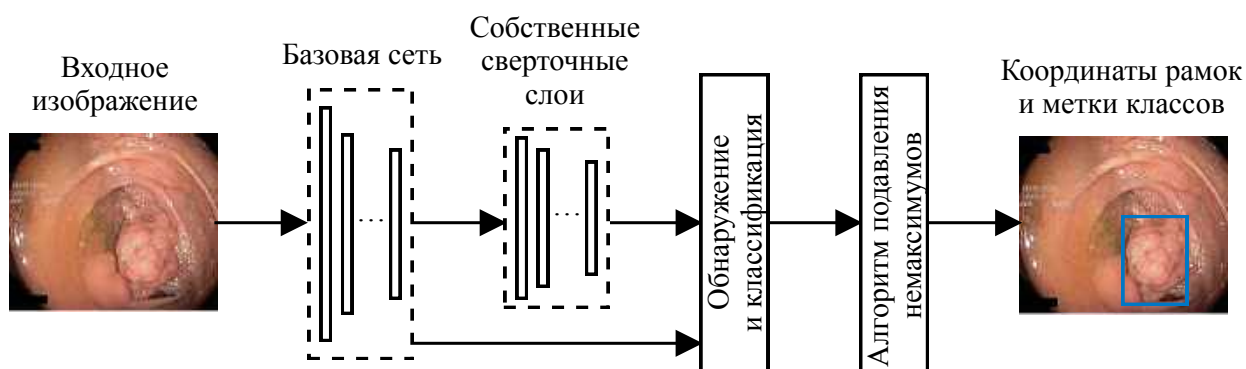


Рисунок 2.2 – Архитектура сверточной нейронной сети SSD

Сеть SSD генерирует набор ограничивающих рамок фиксированных размеров, а также оценку достоверности того, что объект в рамке принадлежит определенной категории объектов. Эти данные подаются на вход алгоритма подавления немаксимумов для окончательного обнаружения объекта в кадре. Ряд начальных слоев сети SSD (базовая сеть), представляют собой стандартную архитектуру сети VGG [67] с отсечением слоев классификации. Далее, к базовой сети добавляется вспомогательная структура, производящая обнаружение со следующими ключевыми особенностями.

- 1) Использование мультимасштабных карт признаков для обнаружения: сверточные слои, расположенные после усеченной базовой сети,

постепенно уменьшаются в размере и позволяют прогнозировать ограничивающие рамки в нескольких масштабах.

- 2) Использование сверточных предикторов обнаружений: каждый добавленный слой признаков (также могут использоваться слои базовой сети) может производить фиксированное количество предсказаний для обнаружения объектов на основе использования набора сверточных фильтров. Для слоя сети размера  $m \times n$  с  $p$  каналами основным элементом для прогнозирования параметров потенциального объекта является небольшое ядро  $3 \times 3 \times p$ . В каждом из  $m \times n$  местоположений, где применяется ядро, оно выдает некоторое выходное значение, характеризующее сдвиг ограничивающей рамки относительно ее исходного положения.
- 3) Использование набора «боксов» и соотношения сторон: каждой ячейке карт признаков ставится в соответствие набор ограничивающих прямоугольников – «боксов» (default boxes), для которых прогнозируется их смещение относительно начального положения «бокса», а также оценка достоверности наличия экземпляра класса в данном «боксе».

В стандартном варианте размер входного слоя сети SSD составляет  $300 \times 300$  (сеть SSD-300), однако, согласно исследованию [111], в качестве базовой сети в данной архитектуре могут успешно использоваться и другие варианты СНС. В связи с тем, что разрешение кадра видеоданных в используемом при создании базы изображений эндоскопическом оборудовании Olympus составляет  $720 \times 576$  пикселей, принято решение использовать в исследовании помимо стандартной сети SSD-300 также архитектуру SSD с базовой сетью VGG, размер входного слоя которой составляет  $512 \times 512$  и используется еще один дополнительный масштаб карт признаков для обнаружения полипов.

Кроме того, в связи с высокими требованиями разрабатываемой системы к скорости работы алгоритма обнаружения, связанной с



необходимостью обработки данных в режиме реального времени, реализована архитектура сети SSD, в которой в качестве базовой сети использовалась часть архитектуры MobileNet [84]. Особенностью такой модели является ее «легковесность» относительно других сетей, связанная с отсутствием слоев пулинга, вместо которых используются свертки с шагом 2.

Таким образом, для дальнейших исследований в рамках данного этапа работ выбраны следующие модели СНС на базе архитектуры SSD:

- 1) Архитектура SSD с базовой сетью VGG, размер входного слоя  $300 \times 300$ .
- 2) Архитектура SSD с базовой сетью VGG, размер входного слоя  $512 \times 512$ .
- 3) Архитектура SSD с базовой сетью MobileNet, размер входного слоя  $300 \times 300$ .

### **2.2.2 Архитектура сверточной нейронной сети YOLOv5**

Базовой сетью недавно предложенной архитектуры YOLOv5 является сеть Darknet53 с добавлением межступенчатых частичных соединений (Cross-Stage-Partial-connections, CSP) [148], служащая для выделения признаков из входного изображения. К базовой части также добавляется блок объединения пространственных пирамид (Spatial pyramid pooling, SPP), поскольку он значительно увеличивает поле восприятия, выделяет наиболее важные характерные особенности и практически не снижает скорость работы сети [80]. Для агрегации признаков с разных слоев экстрактора в данной архитектуре используется сеть агрегации путей (Path Aggregation Network, PANet) [110]. Далее следуют слои обнаружения и классификации из архитектуры YOLOv3 [129], выходы которых обрабатываются алгоритмом подавления немаксимумов. Архитектура СНС YOLOv5 приведена на Рисунке 2.3.

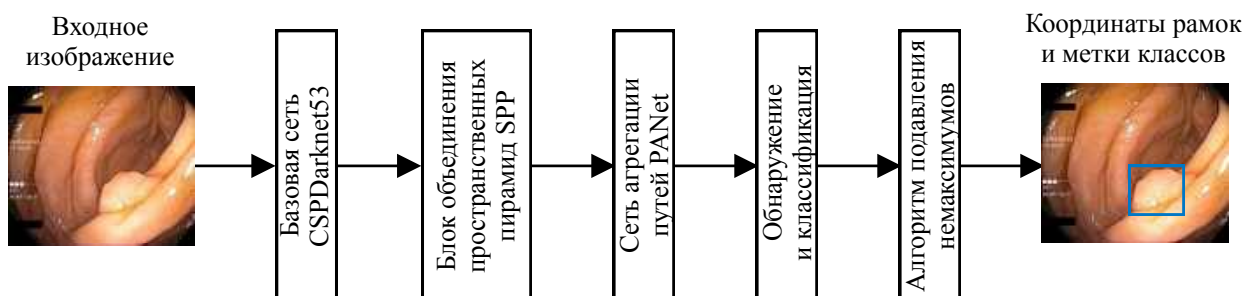


Рисунок 2.3 – Архитектура сверточной нейронной сети YOLOv5

Основным отличием архитектуры YOLOv5 относительно YOLOv4 является автоматический подбор размеров и пропорций якорных ограничивающих рамок на основе распределения ограничивающих рамок в обучающем наборе, что позволяет точнее подстроить нейросетевой алгоритм под целевой обучающий набор, кроме того, предыдущие версии реализованы с использованием собственного фреймворка Darknet, в то время как алгоритм YOLOv5 реализован с использованием популярного фреймворка PyTorch, в связи с чем в реализации присутствуют различные минорные изменения.

Существует 5 модификаций архитектуры YOLOv5 – N, S, M, L, X, отличающиеся количеством слоев и параметров сети. Модели образованы путем масштабирования базовой архитектуры по ширине и глубине [145]. На основе предварительных экспериментов было решено в данном исследовании самую простую модификацию сети YOLOv5N не рассматривать. Таким образом, для дальнейших исследований в рамках данного этапа работ выбраны следующие модели СНС на базе архитектуры YOLO: YOLOv5S, YOLOv5M, YOLOv5L, YOLOv5X.

Для обучения описанных в разделе 2.2. архитектур нейронных сетей требуются специализированные базы изображений с предварительной экспертной разметкой областей интереса. К сожалению, в отличие от собранных за последние два десятилетия гигантских по размеру баз изображений общего назначения (ImageNet, Pascal VOC, MS-COCO) в области эндоскопии подобные исследовательские проекты еще пока находятся в начальной стадии развития.

### 2.3 Базы эндоскопических изображений

Наборы экспертно размеченных изображений играют важную роль для построения алгоритмов анализа изображений методами глубокого машинного обучения. Рассмотрим список открытых баз эндоскопических изображений и видеоданных, полученных с помощью эндоскопа в режиме визуализации в белом свете:

- Kvasir-SEG [86] – 1000 изображений толстой кишки с экспертной разметкой полипов для автоматического решения задач обнаружения и сегментации.
- Kvasir [123] – 8000 изображений всего желудочно-кишечного тракта (ЖКТ) с разметкой следующих классов: полип, эзофагит, язвенный колит, z-линия, привратник желудка, устье слепой кишки, окрашенный полип, окрашенные края резекции и стул.
- Nerthus [124] – 21 видеопоследовательность исследования толстой кишки с разметкой следующих классов: очищенный кишечник/не очищенный кишечник.
- HyperKvasir [50] – 110079 изображений и 373 видеопоследовательности всего ЖКТ, с разметкой 16 различных классов для верхних отделов ЖКТ и 24 различных классов для нижних отделов ЖКТ.
- ETIS-Larib [141] – 196 изображений толстой кишки с разметкой полипов для задачи сегментации изображений.
- SVC-Clinic [46] – 612 изображений толстой кишки с разметкой полипов для задачи сегментации изображений.
- Kvasir-Instrument [85] – 590 изображений толстой кишки с разметкой медицинских инструментов для задач обнаружения и сегментации изображений.

Также следует обратить внимание на базу EDD 2020 [36], которая содержит 386 изображений всего ЖКТ, полученных с помощью эндоскопа,

как в режиме визуализации в белом свете, так и в режиме узкоспектральной визуализации (NBI). Она содержит разметку следующих классов для задач обнаружения и сегментации изображений: полип, пищевод Барретта, дисплазия высокой степени, предраковые участки, рак. Еще одна база KvasirCapsule [142] представляет собой 4820739 изображений и 118 видеопоследовательностей всего ЖКТ, полученных с помощью капсульной эндоскопии и содержит разметку 13 различных классов аномалий ЖКТ для задачи классификации изображений.

Таким образом, базы Kvasir-SEG, ETIS-Larib и CVC-ClinicDB содержат изображения колоноскопических исследований, тогда как базы Kvasir, Nerthus и HyperKvasir содержат изображения всего ЖКТ. Все эти наборы данных содержат изображения полипов. Из девяти доступных наборов только Kvasir-SEG, ETISLarib и CVC-ClinicDB имеют вручную созданную экспертную разметку, необходимую для решения задачи обнаружения полипов. Среди них база Kvasir-SEG предлагает наибольшее количество размеченных примеров, обеспечивающих как маски для сегментации, так и ограничивающие рамки, что дает возможность использовать этот набор как для решения задачи автоматической сегментации, так и для задачи автоматического обнаружения полипов. Таким образом, для исследований в данной главе использовался именно этот набор изображений.

На Рисунке 2.4 показаны изображения, бинарные маски и ограничивающие рамки для полипов из набора Kvasir-SEG. Данный набор содержит изображения полипов, полученных с помощью эндоскопической системы Olympus, соответствующие им маски и информацию об ограничивающих рамках. Изображения и их маски могут использоваться для решения задачи сегментации полипов, а ограничивающие рамки для решения задачи обнаружения полипов. Разрешение изображений в этом наборе данных варьируется от  $332 \times 487$  до  $1920 \times 1072$  пикселей. Набор данных включает изображения 700 «больших полипов» ( $> 160 \times 160$  пикселей), 323

полипов «среднего размера» ( $> 64 \times 64$  пикселей и  $\leq 160 \times 160$  пикселей) и 48 «маленьких полипов» ( $\leq 64 \times 64$  пикселей).

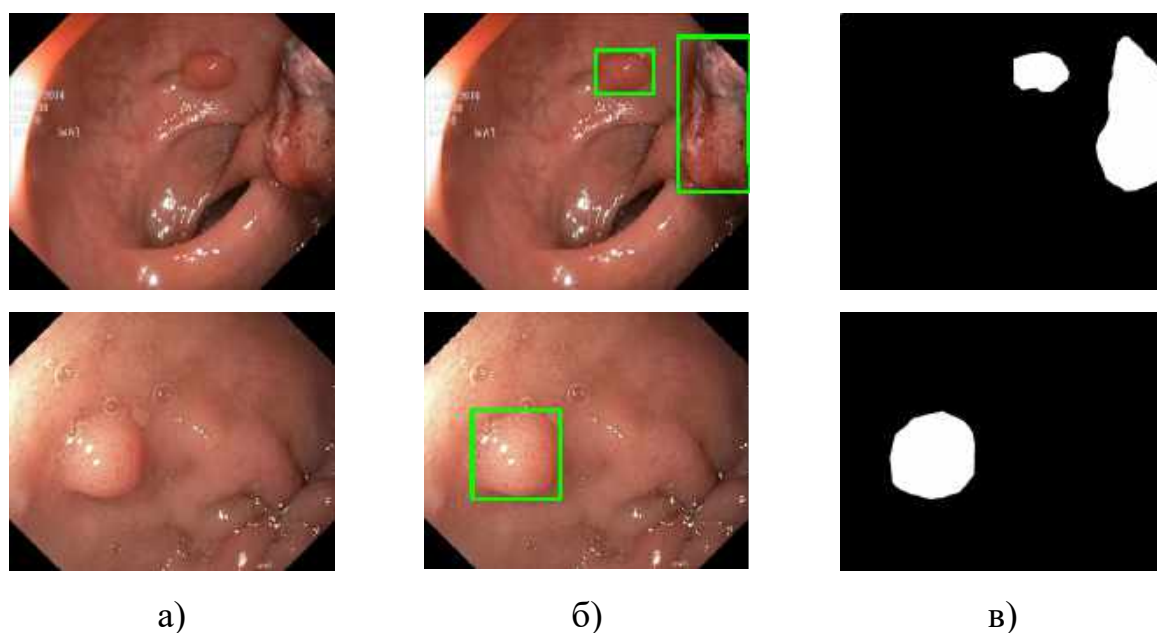


Рисунок 2.4 – Примеры изображений из набора данных Kvasir-SEG:  
 а) исходные изображения; б) изображения с ограничивающими рамками;  
 в) аннотированные бинарные маски

Общее количество полипов в наборе данных Kvasir-SEG составляет 1072 штуки. Следует отметить, что такое число обучающих примеров на современном этапе развития нейросетевых технологий обнаружения объектов является недостаточным, так как типичное число таких изображений обычно составляет десятки и даже сотни тысяч штук. С учетом этого необходимым этапом в построении алгоритма обнаружения полипов является этап размножения (аугментации) данных, который должен производиться с учетом специфики эндоскопических изображений.

#### **2.4 Модификация алгоритма аугментации эндоскопических изображений**

Нейросетевые алгоритмы глубокого обучения требуют большого количества обучающих примеров и показывают улучшение результатов

работы по мере увеличения количества входных данных. Получение таких обучающих примеров посредством сбора и разметки данных – рутинный ручной процесс, требующий значительных ресурсов и трудозатрат как со стороны клинических экспертов, так и специалистов в области обработки изображений. Аугментация данных – распространенный метод увеличения числа обучающих примеров в наборе данных с помощью математических преобразований.

Общепринятые стандартные методы аугментации данных разработаны для применения на объектах реального мира из базы изображений ImageNet [133], в то время как эндоскопические изображения имеют свою специфику. В частности, для объектов из базы ImageNet не применяются повороты изображения относительно горизонтальной оси и повороты на произвольные углы. Это оправдано, так как в данных изображениях присутствует линия горизонта, и расположение объектов относительно этой линии не случайно: пешеходы, автомобили, велосипедисты, животные, как правило, соответствующим образом ориентированы в пространстве. Если повернуть данные изображения на 180 градусов – объекты, изображенные на них, перестанут выглядеть естественно. Однако в случае с эндоскопическими изображениями, это не так. Полипы могут располагаться в разных частях изображения, а сам эндоскоп во время обследования может вращаться на произвольные углы, пока врач производит обследование пациента. Принимая во внимание данный факт, в работе предложены модификации алгоритмов аугментации данных для эндоскопических изображений, позволяющие получить лучшие результаты.

Рассмотрим стандартную аугментацию данных для алгоритма SSD [111]. Для того чтобы сделать модель более устойчивой к различным размерам и формам входных объектов, к каждому изображению обучающей выборки случайным образом применяется один из следующих вариантов преобразований:

- используется все исходное входное изображение;

- из изображения вырезается фрагмент (патч), для которого минимальное значение  $IoU$  с объектом составляет 0,1; 0,3; 0,5; 0,7 или 0,9;
- из изображения вырезается патч случайным образом.

Размер каждого патча составляет от 10% до 100% от размера исходного изображения, и соотношение сторон находится в диапазоне от 1 к 2 до 2 к 1.

После выполнения вышеописанного этапа производится изменение размеров каждого фрагмента до некоторого фиксированного значения и осуществляется отзеркаливание части фрагментов по горизонтали, а также выбирается с вероятностью 0,5 одна из двух последовательностей фотометрических преобразований, аналогичных описанным в [83].

При использовании предложенного **модифицированного варианта аугментации для сети SSD** дополнительно используются преобразования, при которых учитывается специфика эндоскопических изображений. В частности, в отличие от стандартной аугментации, в данном варианте рассматривается также отзеркаливание изображений относительно горизонтальной оси.

В целом, при использовании модифицированного варианта аугментации выполняется преобразование изображения из цветового пространства RGB в HSV, рандомизация яркости, контрастности, насыщенности и оттенка в заданных пределах, осуществляется случайный поворот в пределах 180 градусов, отзеркаливание относительно горизонтальной и вертикальной осей, вырезание из изображения патчей случайным образом и изменение их размера с выбором различного соотношения сторон.

Аналогичным образом произведена модификация аугментации для алгоритма YOLOv5. При стандартной аугментации данных для него к каждому изображению обучающей выборки, также как и в алгоритме SSD, применяются фотометрические и геометрические преобразования для увеличения вариативности входных изображений и повышению устойчивости модели к различным условиям. Однако, в отличии от сети SSD,

для формирования изображения применяется новый метод аугментации – мозаика, который смешивает 4 обучающих изображения. Это позволяет разнообразить обычное окружение объектов. Кроме того, пакетная нормализация (batch normalizaton) оценивает статистику активации по 4 различным изображениям на каждом слое. Это значительным образом снижает потребность в большом размере пакета (batch size) [49]. После формирования мозаики применяются геометрические преобразования, затем фотометрические преобразования и отзеркаливание относительно вертикальной оси.

При использовании **модифицированного варианта аугментации для сети YOLOv5** дополнительно используются вращение изображения на случайный угол в пределах 180 градусов и отзеркаливание относительно горизонтальной оси. Таким образом, предлагаемый модифицированный алгоритм аугментации для сети YOLOv5 состоит из:

- процедуры составления мозаики;
- случайного вращения изображения в пределах 180 градусов;
- случайного изменения масштаба в пределах -50..50%;
- случайных сдвигов по осям x и y;
- перевода изображения в цветовое пространство HSV и процедуры случайных искажений во всех трех компонентах;
- процедуры случайного (с вероятностью 0,5) отзеркаливания изображения относительно горизонтальной оси;
- процедуры случайного (с вероятностью 0,5) отзеркаливания изображения относительно вертикальной оси.

Проведены вычислительные эксперименты для сравнения процессов стандартной и модифицированной аугментации эндоскопических изображений. Для проведения сравнительного анализа в алгоритме SSD используется модель CHC SSD-300, описанная в п. 2.2.1, а для алгоритма YOLO – модификация нейронной сети YOLOv5X, описанная в п. 2.2.2. Набор изображений Kvasir-SEG разделен на два поднабора – обучающий и



тестовый. При этом 880 изображений использовались для обучения нейросетевой модели и 120 изображений – для тестирования соответствующего алгоритма обнаружения полипов.

На Рисунке 2.5 приведены кривые зависимости значений метрики  $AP@0,50$  от числа эпох обучения ( $E$ ) для стандартного и модифицированного алгоритмов аугментации для сети SSD-300. Как видно из приведенных зависимостей, у модифицированного алгоритма значение метрики  $AP@0,50$  достигает большего значения, чем у стандартного алгоритма. Преимущество составляет в среднем 2,2%.

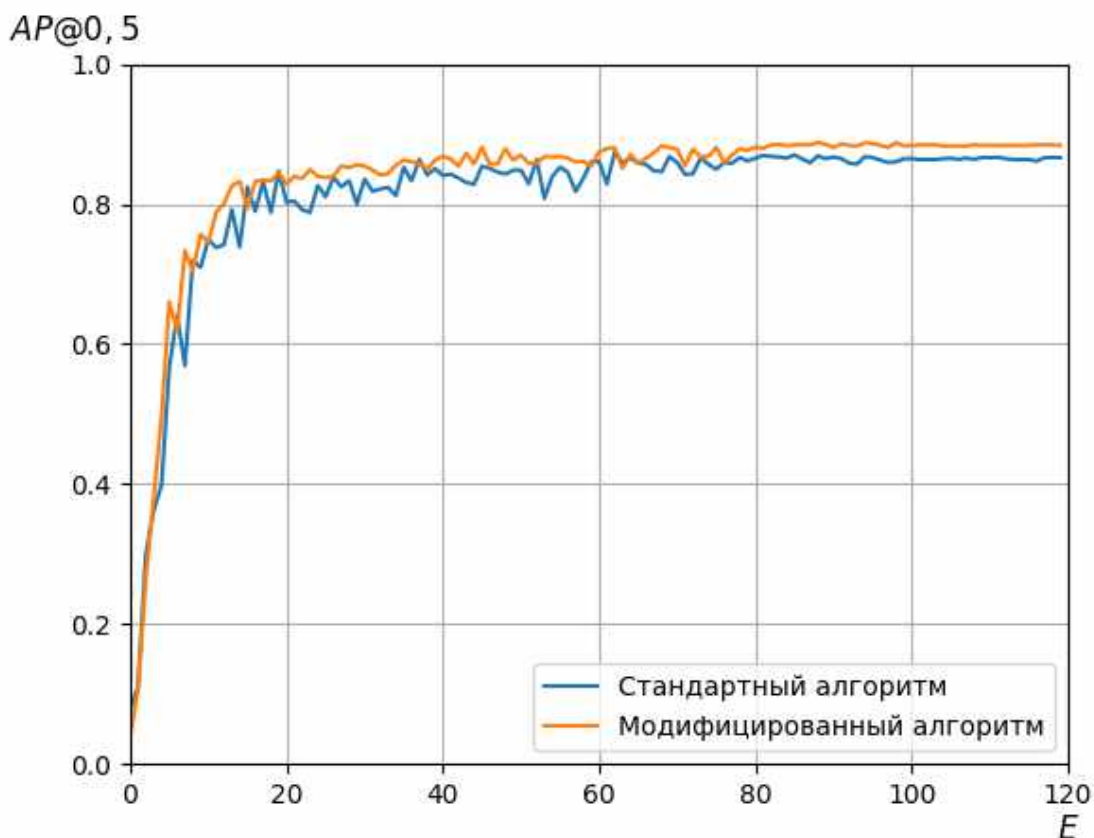


Рисунок 2.5 – Зависимость метрики  $mAP@0,50$  от числа эпох обучения для стандартного и модифицированного алгоритмов аугментации для сети SSD-300

На Рисунке 2.6 представлены графики кривых зависимости точности от полноты для стандартного и модифицированного алгоритмов аугментации

для сети SSD-300. Согласно приведенным результатам, метрика  $AP@0,50$  для модифицированного алгоритма аугментации составляет 93,0%, что на 1,4% больше, чем для стандартного алгоритма.

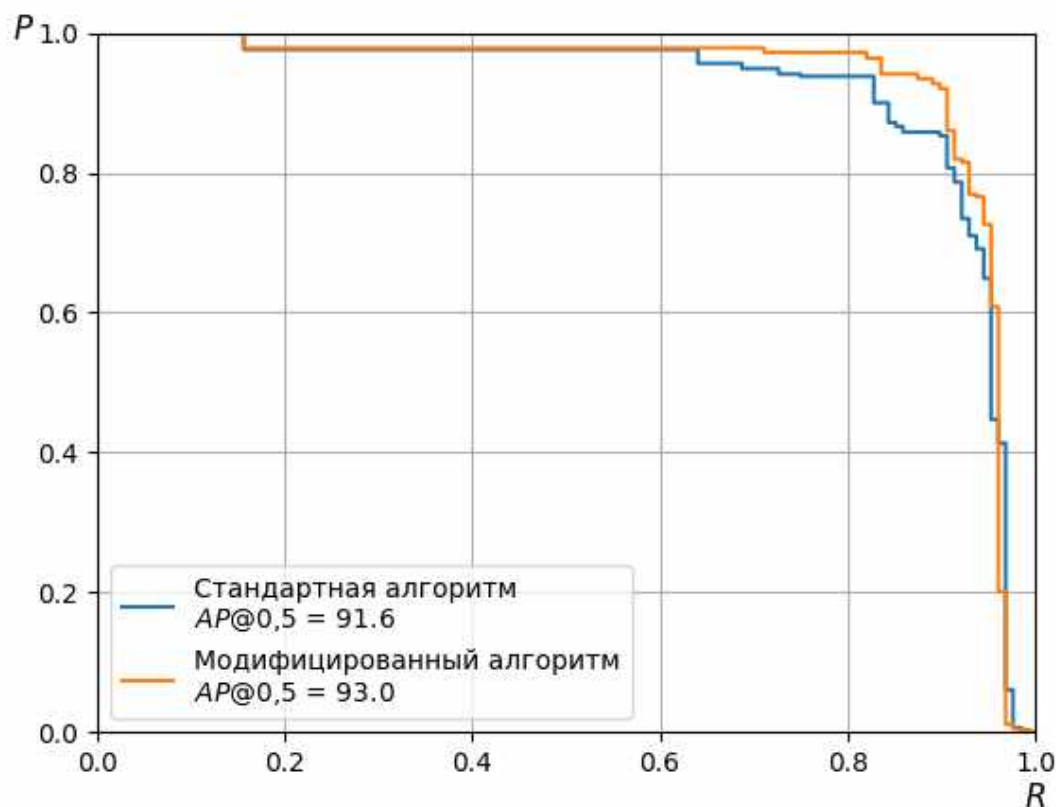


Рисунок 2.6 – Зависимость значений точности от полноты для стандартного и модифицированного алгоритмов аугментации для сети SSD-300

На Рисунке 2.7 приведены кривые зависимости значений метрики  $AP@0,50$  от числа эпох обучения для стандартного и модифицированного алгоритмов аугментации для сети YOLOv5. Приведенные зависимости показывают, что у модифицированного алгоритма медленнее нарастает метрика  $AP@0,50$ , однако после 40-й эпохи обучения модифицированный алгоритм достигает большего значения  $AP@0,50$ , чем стандартный алгоритм (преимущество составляет в среднем 2,1%).

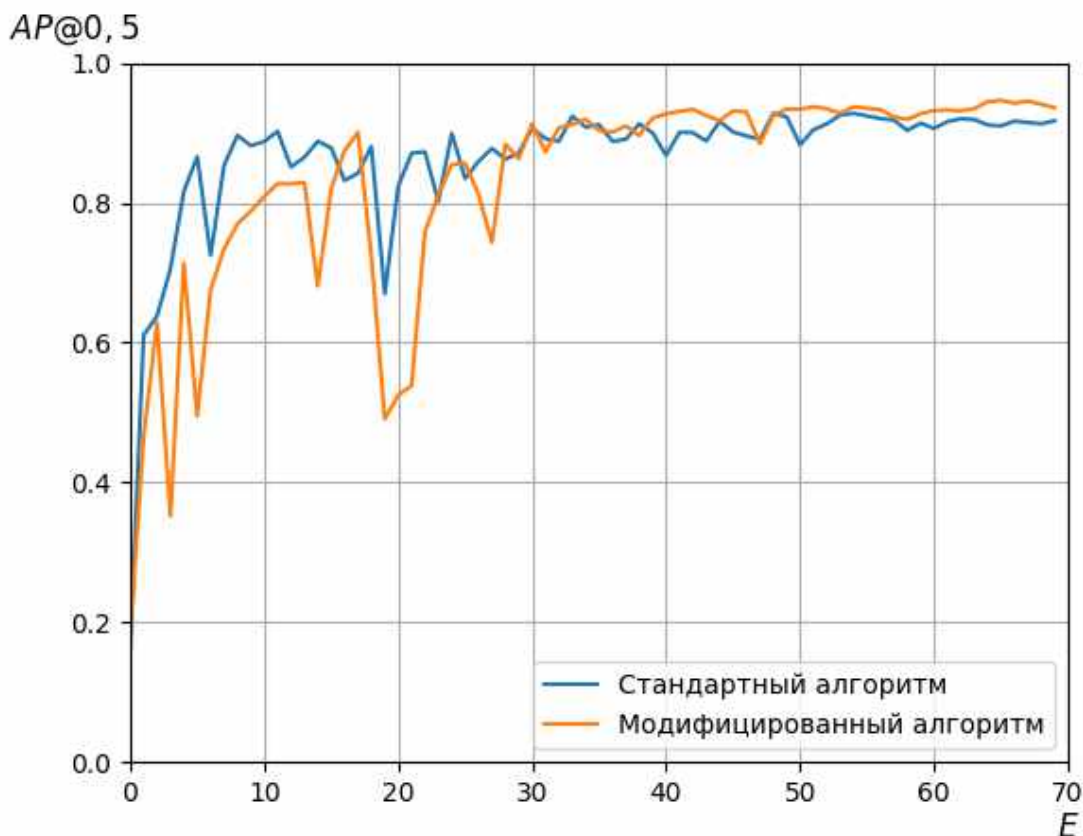


Рисунок 2.7 – Зависимость метрики  $mAP@0,50$  от числа эпох обучения для стандартного и модифицированного алгоритмов аугментации для сети YOLOv5

На Рисунке 2.8 представлены зависимости точности от полноты для стандартного и модифицированного алгоритмов аугментации для сети YOLOv5. Согласно представленным результатам, метрика  $AP@0,50$  для модифицированного алгоритма обнаружения полипов на основе YOLOv5 составляет 94,6%, что на 3,0% больше, чем для стандартного алгоритма.

Таким образом, проведенные вычислительные эксперименты показали, что использование модифицированного с учетом особенностей эндоскопических изображений алгоритма аугментации данных позволяет достичь преимущества как для сверточной нейронной сети SSD-300, так и для СНС YOLOv5.

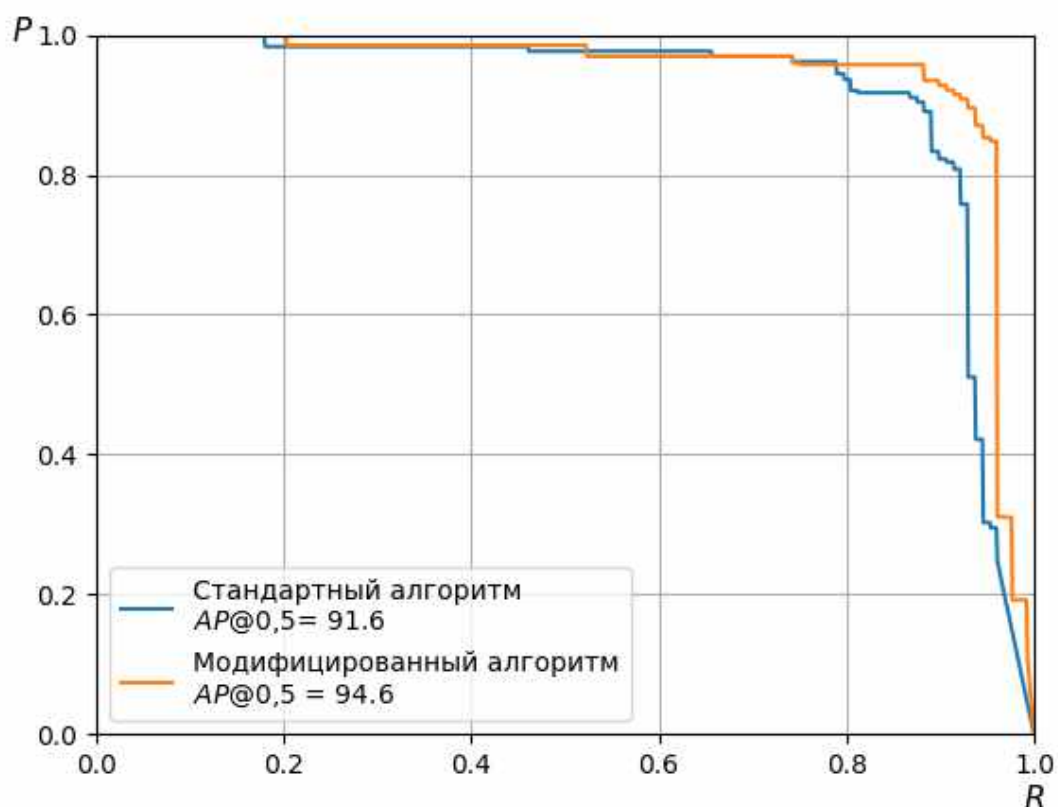


Рисунок 2.8 – Зависимость значений точности от полноты для стандартного и модифицированного алгоритмов аугментации для сети YOLOv5

## 2.5 Тестирование нейросетевых алгоритмов обнаружения полипов

Для решения поставленной задачи предложены и исследованы следующие нейросетевые алгоритмы обнаружения объектов (НАОО) на основе рассмотренных выше архитектур SSD и YOLOv5:

- алгоритм НАОО<sub>SV300</sub> – на основе архитектуры SSD с базовой сетью VGG300;
- алгоритм НАОО<sub>SV512</sub> – на основе архитектуры SSD с базовой сетью VGG512;
- алгоритм НАОО<sub>SM300</sub> – на основе архитектуры SSD с базовой сетью MobileNet;
- алгоритм НАОО<sub>YS</sub> – на основе архитектуры YOLOv5S;

- алгоритм  $\text{HAOO}_{\text{YM}}$  – на основе архитектуры YOLOv5M;
- алгоритм  $\text{HAOO}_{\text{YL}}$  – на основе архитектуры YOLOv5L;
- алгоритм  $\text{HAOO}_{\text{YX}}$  – на основе архитектуры YOLOv5X.

Для алгоритма  $\text{HAOO}_{\text{SV300}}$  использовались следующие параметры обучения: размер батча составлял 8 изображений, в качестве алгоритма оптимизации использовался стохастический градиентный спуск. Начальная скорость обучения, используемая до 80-й эпохи, выбиралась равной  $10^{-4}$ ; с 80-й по 100-ю эпоху составляла  $10^{-5}$ ; начиная с 100-й эпохи обучения, выбиралась равной  $10^{-6}$ , общее число эпох обучения – 120.

Для алгоритма  $\text{HAOO}_{\text{SV512}}$  использовались следующие параметры обучения: размер батча составлял 8 изображений, в качестве алгоритма оптимизации использовался алгоритм Adam (Adaptive Moment Estimation). Он сочетает в себе и идею накопления движения, и идею более слабого обновления весов для типичных признаков. Начальная скорость обучения, используемая до 40-й эпохи, выбиралась равной  $10^{-4}$ ; с 40-й по 80-ю эпоху она составляла  $10^{-5}$ ; начиная с 100-й эпохи обучения, выбиралась равной  $10^{-6}$ , общее число эпох обучения – 120. Другие параметры алгоритма выбирались следующим образом:  $\beta_1 = 0,9$ ;  $\beta_2 = 0,999$ ;  $\varepsilon = 10^{-8}$ .

Для алгоритма  $\text{HAOO}_{\text{SM300}}$  использовались следующие параметры обучения: размер батча составлял 64 изображения, в качестве алгоритма оптимизации использовался алгоритм Adam. Начальная скорость обучения, используемая до 100-й эпохи выбрана равной  $10^{-3}$ ; с 100-й по 200-ю эпоху составляла  $10^{-4}$ ; начиная с 200-й эпохи – равна  $10^{-5}$ , общее число эпох обучения – 300. Другие параметры алгоритма выбирались следующим образом:  $\beta_1 = 0,9$ ;  $\beta_2 = 0,999$ ;  $\varepsilon = 10^{-8}$ .

Так как алгоритмы  $\text{HAOO}_{\text{YS}}$ ,  $\text{HAOO}_{\text{YM}}$ ,  $\text{HAOO}_{\text{YL}}$  и  $\text{HAOO}_{\text{YX}}$  имеют одну и ту же архитектуру и различаются только числом слоев и количеством обучаемых параметров, то для всех перечисленных алгоритмов использовались одинаковые параметры обучения: размер батча составлял 16

изображений, в качестве алгоритма оптимизации использовался стохастический градиентный спуск, начальная скорость обучения равнялась  $10^{-2}$ ; далее в процессе обучения скорость обучения менялась по косинусной функции, общее число эпох обучения – 70. Другие параметры алгоритма выбирались следующим образом: параметр  $\text{momentum} = 0,937$ , в течение первых трех эпох этот параметр равен 0,8, параметр  $\text{decay} = 5 * 10^{-4}$ , параметры функции потерь для локализации  $\text{box} = 0,04$ , порог отношения пересечения к объединению во время обучения – 0,2.

Для сравнения с аналогичными алгоритмами, приведенными в [88], база эндоскопических изображений Kvasir-SEG разделялась на тестовый и обучающий набор, 120 изображений отбиралось для тестового набора и 880 – для обучающего, однако, так как в работе не указано, какие именно изображения выбраны для тестов, то для сравнения результатов в данном исследовании случайным образом сформировано 10 обучающих и тестовых наборов, после чего каждый алгоритм обучался и тестировался на каждой из 10 пар, а в финальные результаты выбиралось усреднённое по 10 наборам значение метрики оценки качества работы алгоритмов. При проведении вычислительных экспериментов использовался модифицированный вариант алгоритма аугментации, описанный выше в разделе 2.4.

Для оценивания качества работы алгоритмов в [88] предложено использовать метрики средней точности, рассчитанные при значениях порога  $IoU$  0,25, 0,50 и 0,75, данные метрики обозначены как  $AP@0,25$ ,  $AP@0,50$  и  $AP@0,75$  соответственно. Кроме того, для комплексной оценки по множеству порогов  $IoU$  предложена метрика  $AP@[0,25..0,75]$ , которая рассчитывается как среднее 11-ти значений  $AP$ , рассчитанных при 11 различных порогах  $IoU$  в пределах от 0,25 до 0,75 с шагом 0,05.

В Таблицах 2.1-2.3 приведены результаты тестирования алгоритмов на основе архитектуры CHC SSD по 10 случайно сформированным наборам из базы Kvasir-SEG. Наилучший результат среди алгоритмов, основанных на архитектуре SSD, получен для алгоритма HAOO<sub>SV12</sub>.

Таблица 2.1

Тестирование алгоритма  $HAOO_{SV300}$ 

Номер набора	1	2	3	4	5	6	7	8	9	10	Ср. знач.
$AP@[0,25..0,75]$	90,3	91,9	85,5	88,8	91,3	87,4	91,4	89,2	89,6	88,1	89,3
$AP@0,25$	96,3	98,1	94,5	96,0	97,1	95,8	96,8	97,1	94,4	96,6	96,3
$AP@0,50$	93,0	96,7	88,8	93,2	95,9	92,4	94,9	93,6	92,6	92,2	93,3
$AP@0,75$	74,1	74,3	65,5	64,8	75,3	63,4	72,1	68,9	73,4	68,7	70,0

Таблица 2.2

Тестирование алгоритма  $HAOO_{SV512}$ 

Номер набора	1	2	3	4	5	6	7	8	9	10	Ср. знач.
$AP@[0,25..0,75]$	90,5	91,4	90,1	90,4	93,8	89,1	90,4	91,4	88,1	87,3	90,2
$AP@0,25$	97,3	98,0	96,1	96,4	97,8	95,9	96,4	98,3	93,8	96,5	96,6
$AP@0,50$	93,1	96,8	92,5	94,7	96,2	93,5	93,9	94,4	90,7	92,7	93,8
$AP@0,75$	76,7	72,4	74,8	73,7	82,1	71,0	76,1	80,0	75,3	66,6	74,9

Таблица 2.3

Тестирование алгоритма  $HAOO_{SM300}$ 

Номер набора	1	2	3	4	5	6	7	8	9	10	Ср. знач.
$AP@[0,25..0,75]$	76,8	80,9	71,2	73,9	79,3	74,9	76,4	79,6	75,0	76,3	76,4
$AP@0,25$	90,6	93,9	87,7	88,2	94,6	88,3	89,7	94,3	88,3	90,7	90,6
$AP@0,50$	81,0	86,3	79,4	81,6	85,3	84,9	83,5	84,5	80,1	82,3	82,9
$AP@0,75$	45,8	48,4	37,5	44,7	51,3	38,5	39,6	50,4	41,2	49,1	44,6

В Таблицах 2.4-2.7 приведены результаты тестирования алгоритмов на основе архитектуры СНС YOLOv5 по 10 случайно сформированным наборам

из базы Kvasir-SEG. Наилучший результат среди алгоритмов, основанных на архитектуре YOLOv5, получен для алгоритма НАОО<sub>YL</sub>.

Таблица 2.4

Тестирование алгоритма НАОО<sub>YS</sub>

Номер набора	1	2	3	4	5	6	7	8	9	10	Ср. знач.
<i>AP@[0,25..0,75]</i>	90,8	91,5	86,5	88,1	91,9	87,7	88,4	89,5	85,5	86,6	88,6
<i>AP@0,25</i>	96,5	95,6	92,7	92,5	95,5	92,9	94,4	95,6	90,9	93,9	94,0
<i>AP@0,50</i>	93,5	93,5	89,8	89,8	93,2	90,7	90,8	90,6	87,7	88,1	90,8
<i>AP@0,75</i>	71,1	78,5	66,5	75,9	80,9	70,7	72,4	78,0	68,5	75,5	73,8

Таблица 2.5

Тестирование алгоритма НАОО<sub>UM</sub>

Номер набора	1	2	3	4	5	6	7	8	9	10	Ср. знач.
<i>AP@[0,25..0,75]</i>	93,4	94,2	89,5	93,0	92,4	90,6	93,2	91,8	90,0	89,8	91,8
<i>AP@0,25</i>	97,2	96,9	94,1	95,8	95,4	94,9	97,1	97,5	93,4	96,3	95,9
<i>AP@0,50</i>	94,4	95,3	91,0	94,2	93,6	93,5	94,9	94,5	90,2	91,4	93,3
<i>AP@0,75</i>	86,0	87,1	79,8	84,5	85,7	74,8	84,0	81,0	82,3	82,8	82,8

Таблица 2.6

Тестирование алгоритма НАОО<sub>YL</sub>

Номер набора	1	2	3	4	5	6	7	8	9	10	Ср. знач.
<i>AP@[0,25..0,75]</i>	93,7	95,6	91,1	92,0	94,1	92,5	94,0	92,8	89,6	90,1	92,5
<i>AP@0,25</i>	97,4	98,6	95,8	95,7	97,0	96,5	97,2	97,9	93,4	97,3	96,7
<i>AP@0,50</i>	95,0	97,0	93,0	93,8	94,8	94,3	95,8	93,9	90,3	91,4	93,9
<i>AP@0,75</i>	87,8	89,7	81,8	82,6	88,2	83,3	84,9	84,4	85,4	83,2	85,1



Таблица 2.7

Тестирование алгоритма НАОО<sub>УХ</sub>

Номер набора	1	2	3	4	5	6	7	8	9	10	Ср. знач.
$AP@[0,25..0,75]$	92,9	95,2	89,9	91,3	92,0	90,8	92,2	91,6	89,9	89,5	91,5
$AP@0,25$	96,1	97,3	95,1	94,9	95,9	94,8	95,5	97,0	93,3	95,7	95,6
$AP@0,50$	94,6	96,4	89,2	92,6	92,4	91,4	93,4	92,5	89,9	90,6	92,3
$AP@0,75$	83,3	89,9	84,5	81,5	86,3	83,0	83,2	84,9	85,6	83,7	84,6

На Рисунке 2.9 представлены кривые зависимости метрики точности от полноты с указанием метрики  $AP@0,50$ , построенные для алгоритмов на основе архитектуры SSD, а на Рисунке 2.10 – для алгоритмов на основе архитектуры YOLOv5, обученных на первом из десяти случайных наборов. Полученные результаты подтверждают полученные выше закономерности относительно превосходства алгоритмов НАОО<sub>SV512</sub> и НАОО<sub>УЛ</sub>.

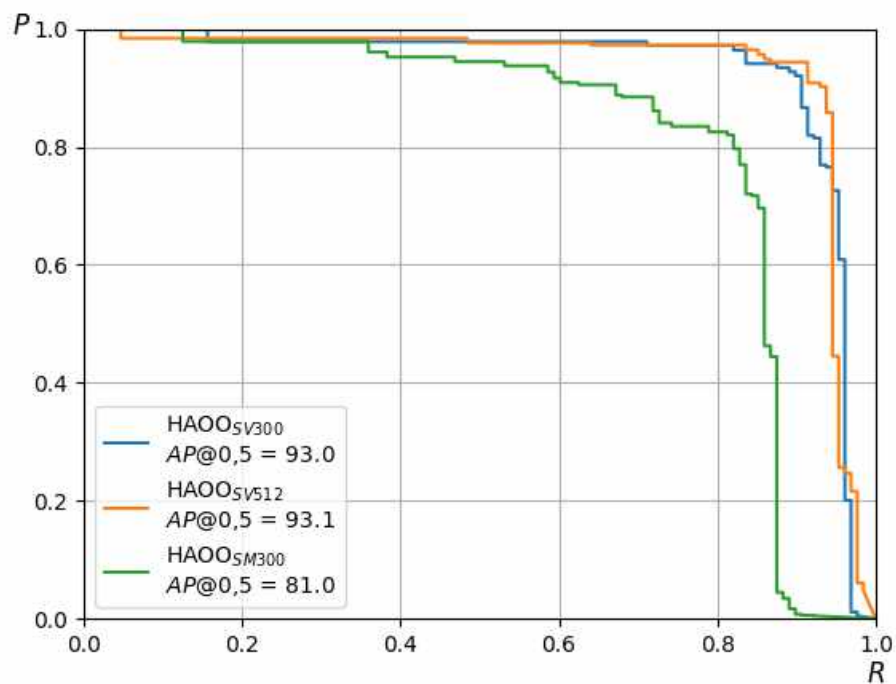


Рисунок 2.9 – Зависимости точности от полноты для нейронных сетей на основе SSD для изображений из базы Kvasir-SEG

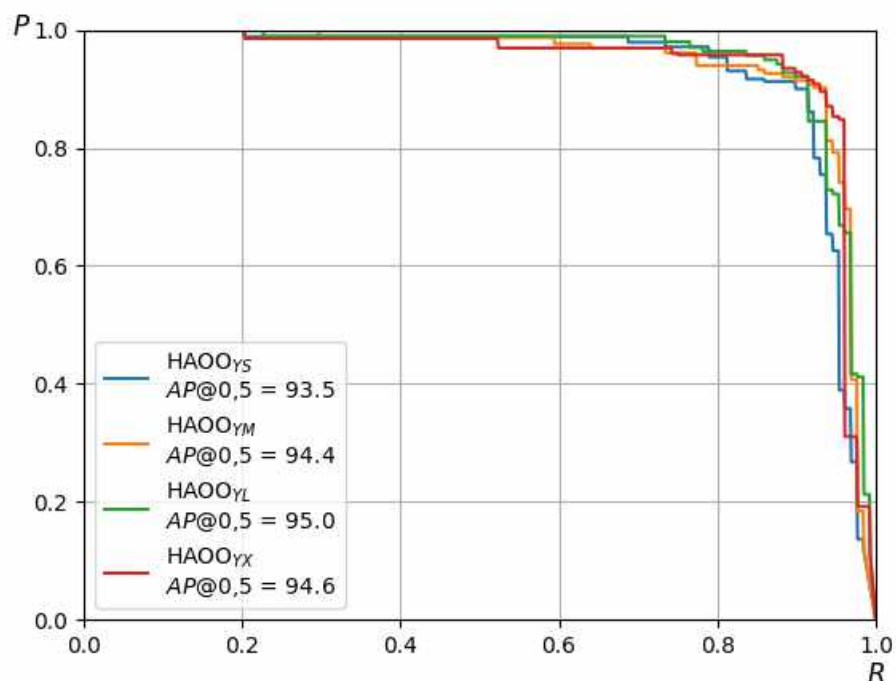


Рисунок 2.10 – Зависимости точности от полноты для нейронных сетей на основе YOLOv5 для изображений из базы Kvasir-SEG

## 2.6 Сравнительный анализ нейросетевых алгоритмов обнаружения полипов на базе Kvasir-SEG

Выбор стандартной базы изображений полипов Kvasir-SEG позволил провести сравнительное исследование предложенных нейросетевых алгоритмов обнаружения с известными из литературы аналогами. В Таблице 2.8 приведены усредненные по 10-ти наборам метрики оценки качества для предложенных алгоритмов, а также результаты для других известных аналогов, взятых из работы [88].

Для оценки скорости работы алгоритма используется стандартная метрика – число кадров, которое алгоритм может обработать за одну секунду (FPS). Для оценки метрики FPS исследуемых алгоритмов использовался персональный компьютер с графическим процессором Nvidia RTX 2080Ti, как и в работе [88].

Сравнение алгоритмов обнаружения полипов на базе Kvasir-SEG

Метрика	$AP@[0,25..0,75]$	$AP@0,25$	$AP@0,5$	$AP@0,75$	$FPS$
<b>Известные аналоги из [88]</b>					
EfficientDet-D0	47,6	68,5	50,5	22,8	35,00
Faster R-CNN	78,7	89,5	84,2	56,6	8,00
RetinaNet50	87,0	93,9	90,9	69,7	16,20
RetinaNet101	<b>87,4</b>	<b>94,8</b>	<b>90,9</b>	71,3	16,80
YOLOv3+spp	81,0	88,6	85,3	<b>75,9</b>	45,01
YOLOv4	85,1	91,2	82,3	<b>75,9</b>	48,00
ColonSegNet	80,0	90,0	81,7	67,1	<b>180,00</b>
<b>Предложенные алгоритмы</b>					
HAOOSV300	89,3	96,3	93,3	70,0	71,97
HAOOSV512	90,2	96,6	93,8	74,9	46,33
HAOOSM300	76,4	90,6	82,9	44,6	80,16
HAOOSYS	88,6	94,0	90,8	73,8	79,36
HAOOSYM	91,8	95,9	93,3	82,8	61,35
HAOOSYL	<b>92,5</b>	<b>96,7</b>	<b>93,9</b>	<b>85,1</b>	54,05
HAOOSYX	91,5	95,6	92,3	84,6	43,85

Среди известных из [88] аналогов лучшие значения метрик средней точности при малых и средних порогах  $IoU$  достигает алгоритм RetinaNet101, а лучшее значение метрики оценки скорости работы алгоритма соответствует алгоритму ColonSegNet. Среди предложенных нейросетевых алгоритмов обнаружения лучшие результаты демонстрирует HAOOSYL. Стоит отметить, что алгоритм HAOOSV512 показал сопоставимый результат при низких и средних порогах  $IoU$ , однако при более высоких значениях  $IoU$  предпочтительней работают алгоритмы на базе архитектуры YOLOv5.

Еще одним важным результатом исследования является тот факт, что самый вычислительно сложный алгоритм  $\text{HАОО}_{\text{УХ}}$  показывает результат ниже, чем  $\text{HАОО}_{\text{УЛ}}$ , несмотря на большее число слоев нейронной сети. Это связано с тем, что обучающая выборка с учетом проведенной процедуры аугментации, тем не менее, имеет недостаточное число примеров для обучения данной сложной СНС.

На Рисунке 2.11 приведены изображения из тестового набора Kvasir-SEG с нанесенной эталонной разметкой и областями обнаружения полипа, предсказанными алгоритмом  $\text{HАОО}_{\text{УЛ}}$ .

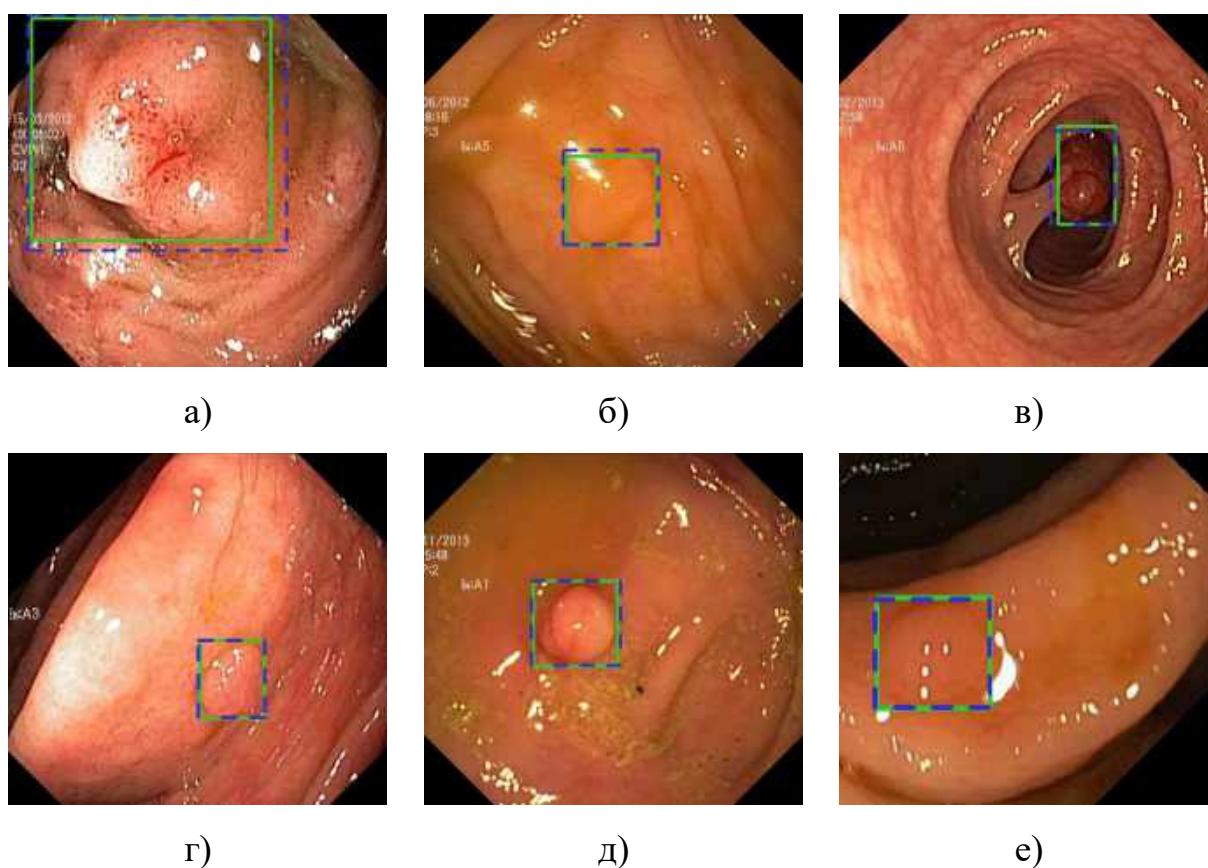


Рисунок 2.11 – Изображения из тестового набора Kvasir-SEG с нанесенной эталонной разметкой (зеленая сплошная линия) и рамками, предсказанными алгоритмом (синяя пунктирная линия)

Таким образом, предложенный алгоритм  $\text{HАОО}_{\text{УЛ}}$  достигает наилучшего результата по метрикам средней точности среди всех

рассматриваемых алгоритмов обнаружения полипов. Данный подход имеет высокий потенциал для построения на его основе модуля анализа видеопотока в эндоскопической системе. Структурная схема алгоритма приведена на Рисунке 2.12.



Рисунок 2.12 – Структурная схема разработки алгоритма обнаружения полипов на эндоскопических изображениях кишечника

## 2.7 Краткие выводы

Результаты проведенных исследований в части разработки алгоритмов обнаружения полипов на статических изображениях колоноскопических исследований позволяют сделать следующие основные выводы:

- Для решения задачи обнаружения полипов на изображениях колоноскопических исследований предложены 3 алгоритма на основе архитектуры нейронной сети SSD и 4 алгоритма на основе архитектуры YOLOv5. Обучение и тестирование нейросетевых алгоритмов проводилось с использованием суперкомпьютера NVIDIA DGX-1, предназначенного для решения задач глубокого обучения.
- Предложенная модификация процесса аугментации данных позволила улучшить значение метрики  $AP@0,50$  в среднем на 1,4% для алгоритма на основе CHC SSD и на 3,0% для алгоритма на основе CHC YOLOv5;
- В результате тестирования предложенных алгоритмов обнаружения патологических объектов на изображениях с использованием открытого набора изображений Kvasir-SEG, наилучший результат получен для предложенного нейросетевого алгоритма  $HAOO_{YL}$ , для которого значение метрики  $AP@[0,25..0,75]$  составило 92,5%, что превосходит известные аналоги на 5% и более. Некоторым недостатком такого алгоритма является повышенная в 3 раза вычислительная сложность по сравнению с известным из литературы алгоритмом ColonSegNet.
- Исследования, проведенные в этой главе послужили основой для построения на основе разработанного нейросетевого алгоритма обнаружения полипов модуля анализа видеопотока в реальной эндоскопической системе.

## ГЛАВА 3

# РАЗРАБОТКА И АНАЛИЗ АЛГОРИТМА ОБНАРУЖЕНИЯ ПОЛИПОВ НА ЭНДОСКОПИЧЕСКИХ ВИДЕОИЗОБРАЖЕНИЯХ

### 3.1 Вводные замечания

Во многих медицинских СПТ ставятся задачи обнаружения объектов на видеоданных [16, 19, 88]. Это более сложная научно-техническая задача в сравнении с автоматическим анализом статических изображений. Важно отметить, что в реальной ситуации эндоскопическая система выдает на выход видеопоток, обладающий определенными характеристиками, зависящими от класса используемого оборудования, поэтому именно обнаружение полипов на видеоизображениях представляет собой наиболее актуальную практическую задачу.

На видеопоследовательностях зачастую наблюдаются эффекты размытия, отсутствие фокусировки на объекте и прочие артефакты, связанные с движением объекта и камеры относительно друг друга. В связи с этим, несмотря на последние успехи алгоритмов на основе глубокого обучения для анализа изображений, до сих пор существует значительный разрыв в качестве процедуры обнаружения объектов на статических изображениях и на видеопоследовательностях [158].

Для повышения эффективности процесса обнаружения объектов на видеоданных существуют две основные стратегии. С одной стороны, существуют модели обнаружения объектов, специально разработанные для работы с видеопотоками [153, 159]. Они реализуют агрегацию признаков из соседних кадров и достигают более высокой точности, чем модели, использующие признаки одиночных изображений. Однако данные подходы, как правило, являются вычислительно сложными, что сильно ограничивает практическую область их применения.

С другой стороны, используются методы постобработки видеопотока, такие, например, как алгоритмы Seq-NMS [76] и Seq-Bbox-Matching [43]. Данные подходы позволяют улучшить производительность системы обнаружения за счет применения соответствующего алгоритма не в каждом кадре, а, например, в каждом пятом кадре, интерполируя результаты в остальных кадрах. Данные методы основаны на связывании предсказанных объектов между кадрами и использовании этих связей для уточнения результатов обнаружения.

Рассмотрим указанные стратегии работы с видеопотоком более подробно.

### **3.2 Особенности алгоритмов обнаружения объектов на видеоизображениях**

Рассмотрим подробнее два типа подходов к обнаружению объектов на видеоданных с учетом специфических проблем, таких как размытие, потеря фокуса, редкие ракурсы объектов, а также с учетом использования дополнительной временной информации, содержащейся в согласованных видеоданных.

Алгоритмы обнаружения объектов на видеопоследовательностях разработаны для использования окружающего контекста кадра и обычно используют признаки объекта с серии кадров. Например, алгоритм FGFA (Flow-Guided Feature Aggregation, агрегация признаков, управляемая потоком) [159] собирает признаки объекта вдоль его траектории движения, заданной оценкой оптического потока. Метод D&T (Detection and Tracking, обнаружение и сопровождение) [61] применяет модель нейронной сети ConvNet для обнаружения и сопровождения объектов, используя корреляцию между картами признаков и уточняя оценки связанных таким образом предсказаний. Алгоритм SELSA (Sequence Level Semantics Aggregation, агрегация семантики на уровне последовательности) [153] извлекает предполагаемые объекты из разных кадров видеоданных и объединяет их



признаки в зависимости от их семантического сходства. При подходе на основе TSM (Temporal Shift Module, модуль временного сдвига) [106] сдвигают объекты во временном измерении для их обнаружения с помощью двумерной СНС. Алгоритм TCD (Tracklet-Conditioned Detection, треклет-обусловленное обнаружение) [158] уточняет предсказания в каждом кадре с помощью треклетов, рассчитанных на предыдущих этапах.

Описанные выше алгоритмы обычно вычислительно более затратны, чем алгоритмы, работающие со статическими изображениями на входе. Это связано с повышенной сложностью модели, необходимостью обработки большего количества данных и расчета дополнительных параметров, таких, например, как оптический поток. Вследствие этого применимость таких алгоритмов к обнаружению объектов на эндоскопическом видеопотоке в режиме реального времени ограничена.

Другим возможным подходом к обнаружению объектов на видеоданных является применение специальной постобработки к выходу алгоритма обнаружения объектов на изображениях. Методы постобработки добавляют информацию о временной корреляции кадров, содержащейся в видеоданных, в выходные предсказания стандартных алгоритмов, работающих на изображениях. Применительно к покадровой обработке видеоданных такие методы позволяют повысить качество обнаружения объектов.

Основой методов постобработки является способ связывания обнаруженных объектов между последовательными кадрами. Это связывание обычно основано на эвристическом алгоритме сравнения обнаружений. При этом обычно используется значение метрики  $IoU$  между обнаруженными объектами из соседних кадров. Такой подход имеет несколько ограничений. Во-первых, базовый покадровый алгоритм обнаружения не всегда предсказывает надежные координаты ограничивающей рамки. Во-вторых, значения  $IoU$  сильно зависят от смещения из-за движения камеры или объекта. Например, быстро движущиеся объекты могут не иметь

достаточного перекрытия, чтобы получить надежную связь между кадрами. Наконец, одновременное присутствие нескольких объектов в сцене также усложняет разработку алгоритма сравнения обнаружений и делает его недостаточно робастным.

Некоторые из таких методов постобработки основаны на вариациях фильтра Калмана и используют идеи трекирования (сопровождения), чтобы сделать обнаружение более робастным или консистентным (согласованным). Данные подходы, например, часто применяются для повторной идентификации объекта при наблюдении с нескольких камер с фиксированным положением [52, 117]. Так, алгоритм Deep SORT [152] улучшает алгоритм сортировки [47], основанный на фильтрах Калмана и Венгерском алгоритме, путем добавления информации о внешнем виде к каждой предсказанной ограничивающей рамке.

Другие подходы в методах с постобработкой основаны на интерполяции ограничивающей рамки или сопоставлении обнаружений в соседних кадрах. Например, алгоритм T-CNN (Tubelets with Convolutional Neural Networks, треки со сверточными нейронными сетями) [92] использует контекстную информацию для подавления ложноположительных срабатываний, а оптический поток – для интерполяции обнаружений по кадрам для уменьшения ложноотрицательных результатов. Алгоритм Seq-NMS (Sequence Non-Maximum Suppression, подавление немаксимумов последовательности) [76] сопоставляет обнаружения с большим перекрытием между кадрами в одном видеофрагменте, чтобы сделать результаты более робастными. Подход на основе Seq-Bbox-Matching (Sequence Bbox Matching, сопоставление ограничительных рамок последовательности) [43] связывает перекрывающиеся обнаружения из последовательных пар кадров для создания и повторной оценки экземпляров объектов и использует их для предсказания пропущенных обнаружений.

Рассматриваемый в данной главе подход к обработке видеопотока относится к методам постобработки, однако связывание обнаружений между

кадрами происходит за счет применения классификатора на основе логистической регрессии вместо эвристического алгоритма сравнения обнаружений [134].

### **3.3 Применение алгоритмов обнаружения полипов для анализа видеоданных колоноскопических исследований**

Самым простым способом обработки видеопотока для обнаружения объектов является независимое применение разработанного алгоритма к каждому кадру видеопоследовательности. Данный подход применён для тестирования алгоритмов на базе архитектуры YOLOv5, исследованных ранее в главе 2. Для обучения и тестирования алгоритмов необходима база видеоизображений с экспертно размеченными полипами в каждом кадре колоноскопического исследования.

В связи с отсутствием подходящих открытых наборов видеоданных, было принято решение о создании совместно с врачами эндоскопического отделения Ярославской областной клинической онкологической больницы собственной базы видеоданных (база ЯОКОБ), состоящей из видеопоследовательностей 50 колоноскопических исследований 50 различных пациентов. Видеопоследовательности разбиты на кадры, в каждом из которых рамками выделены полипы. Суммарно обучающая база содержит 165 665 изображений, из которых 114 754 кадра содержат полипы. Еще 9 видеофрагментов колоноскопических исследований 9 различных пациентов использовались для создания тестового набора видеоданных, необходимого для анализа качества работы исследуемых алгоритмов. Эти видеопоследовательности также разбивались на кадры, в каждом из которых рамками выделялись полипы. Тестовый набор содержит 11 192 изображения, из которых 7057 содержат полипы.

И в обучающем, и в тестовом наборах данных длительность видеопоследовательностей варьируется от 6 до 30 минут, а длительность пребывания полипа в кадре – от 6 секунд до 8 минут. Видеопоток

записывался с современных эндоскопических систем OLYMPUS EVIS EXERA II и OLYMPUS EVIS EXERA III. Однако данные видеофрагменты являются архивными данными колоноскопических исследований, в силу чего они отличаются от оригинальных видеопотоков с эндоскопических систем. Видеоданные подвергались перекодированию и сжатию с помощью видеокодека H.264, а их разрешение варьируется от  $626 \times 532$  до  $760 \times 436$  пикселей. Такие ограничения позволяют считать базу ЯОКОБ пригодной для лабораторного тестирования алгоритмов анализа видеопотока, но несколько устаревшей от современных реалий эндоскопического оборудования.

На Рисунке 3.1 приведены зависимости точности от полноты с указанием метрики  $AP@0,50$  для алгоритмов  $HAOO_{YS}$ ,  $HAOO_{YM}$ ,  $HAOO_{YL}$  и  $HAOO_{YX}$ , полученные на базе видеоданных ЯОКОБ без процедуры переобучения. То есть тестированию подвергались алгоритмы, обученные на базе статических изображений Kvasir-SEG.

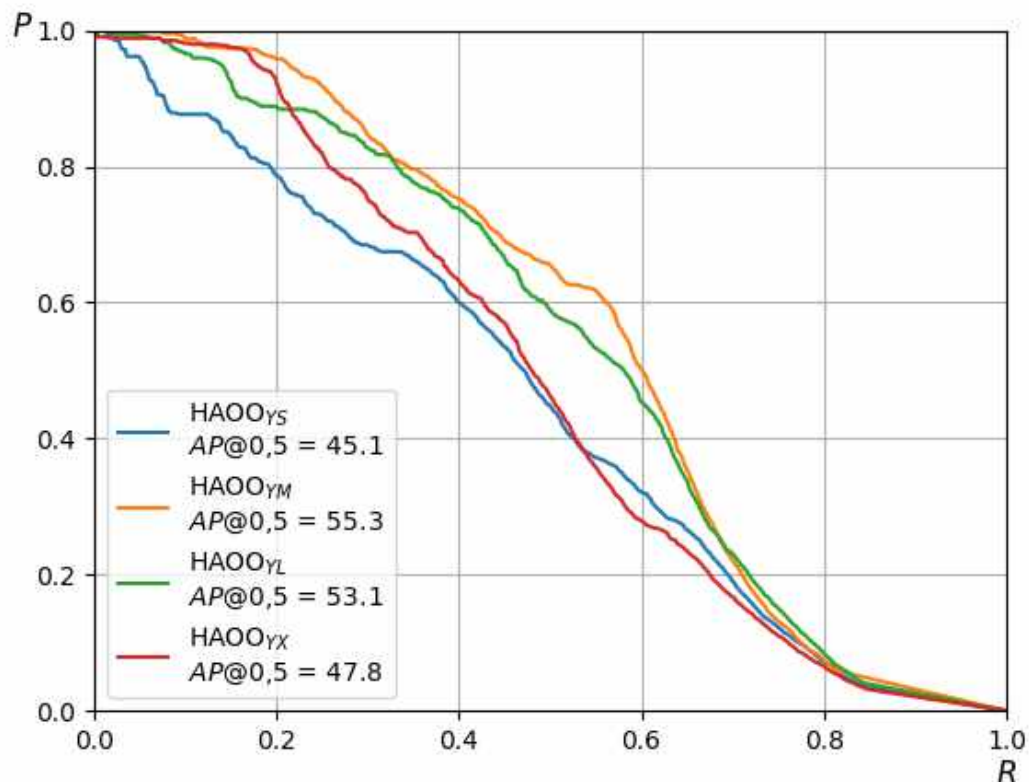


Рисунок 3.1 – Зависимости точности от полноты для нейросетевых алгоритмов, полученные на базе видеоданных ЯОКОБ без переобучения

Из приведенных зависимостей видно, что результаты на базе видеоданных ЯОКОБ существенно ниже, чем результаты, полученные для этих же алгоритмов ранее. Это экспериментально подтверждает тезис, приведенный выше в этой главе о том, что задача обнаружения объектов на видеоданных существенно более сложная по сравнению с задачей обнаружения объектов на статических изображениях.

После переобучения алгоритмов на обучающем наборе базы ЯОКОБ, качество работы алгоритмов обнаружения объектов существенно возросло, что можно видеть на Рисунке 3.2, где приведены кривые зависимости точности от полноты с указанием метрики  $AP@0,50$  для рассматриваемых алгоритмов.

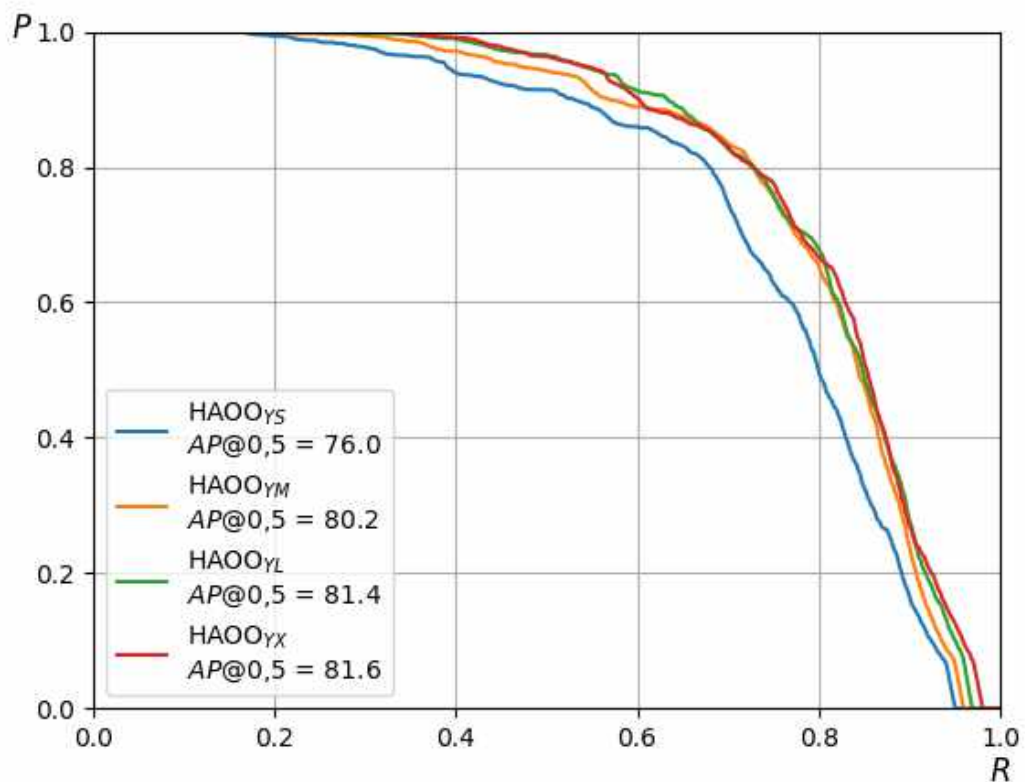


Рисунок 3.2 – Зависимости точности от полноты для нейросетевых алгоритмов, полученные после переобучения на базе ЯОКОБ

Таким образом, показано, что прямое использование разработанных алгоритмов обнаружения полипов после переобучения на видеоданных базы ЯОКОБ приводит к практически востребованным результатам с метрикой

$AP@0,50=81,6$ . Важным вопросом при этом остается способность нейросетевых алгоритмов обрабатывать видеопоток в режиме реального времени.

### **3.4 Модификация алгоритмов обнаружения полипов для анализа видеопотока в режиме реального времени**

В соответствии с исследованиями из предыдущего пункта можно сделать вывод, что после переобучения алгоритмов  $HAOO_{YS}$ ,  $HAOO_{YM}$ ,  $HAOO_{YL}$  и  $HAOO_{YX}$  на базе видеоданных ЯОКОБ, наилучший результат демонстрирует алгоритм  $HAOO_{YX}$ , однако он в то же время является вычислительно самым сложным.

В эндоскопических системах для передачи видеопотока чаще всего используются стандарт HD-SDI SMPTE 292M с разрешением 1080i и частотой обновления 50 Гц, что делает построение системы анализа видеопотока с эндоскопической системы в режиме реального времени сложной задачей и формирует высокие требования к задержкам, вносимым алгоритмами анализа видеопотока.

Выше в Таблице 2.8 приведена оценка по скорости работы рассматриваемых алгоритмов на современном компьютере с производительным графическим процессором Nvidia RTX 2080Ti. Среднее время обработки одного изображения для алгоритма  $HAOO_{YX}$  составило 22,8 мс. При такой скорости обработки за секунду возможно обработать не более 43 кадров, что недостаточно для обработки видеопотока HD-SDI SMPTE 292M с разрешением 1080i и частотой обновления 50 Гц в режиме реального времени. Однако используя метод квантования весов и другие подходы к улучшению производительности алгоритмов на основе СНС, можно существенно улучшить их производительность без значительного ухудшения качества работы. Это позволяет применять более точные нейросетевые алгоритмы для обработки видеопотока с эндоскопической системы в режиме реального времени.

В данном исследовании для квантования весов алгоритма HAOOYX выбрана библиотека NVIDIA TensorRT, представляющая собой фреймворк, в рамках которого можно оптимизировать и развертывать обученные нейронные сети для анализа данных [120].

Среди особенностей данной библиотеки, приводящих к повышению производительности, можно выделить следующие:

- применение оптимального алгоритма свертки для используемого графического процессора;
- вертикальное объединение слоев нейронной сети (вертикальный fusion): вместо последовательного выполнения свертки, добавления смещения и применения функции активации ReLU формируется единый слой, в котором все эти три операции выполняются за раз, что позволяет избежать дополнительных обращений к памяти для чтения и записи матриц;
- горизонтальное объединение слоев нейронной сети (горизонтальный fusion): данная операция подразумевает объединение тех сверток, которые применяются к одним и тем же входным данным, в одну более широкую свертку, что является вычислительно гораздо более эффективным шагом;
- использование вместо шага конкатенации сверток в один тензор запись сверток в нужные участки памяти, что позволяет вовсе отбросить слой конкатенации.

Кроме оптимизации структуры используемой нейронной сети с помощью библиотеки TensorRT можно выполнить квантование параметров сети, представленных в формате представления вещественных чисел FP32 (представление чисел с плавающей точкой в 32-битном формате, числа одинарной точности), в другие форматы. В данном исследовании для квантования параметров сети выбрано представление FP16 (представление чисел с плавающей точкой в 16-битном формате, числа половинной

точности), так как в процессе квантования до целочисленной арифметики происходит, как правило, существенная деградация точности работы сети.

После квантования параметров среднее время обработки одного изображения для алгоритма НАООУХ уменьшилось до 13,1 мс, что эквивалентно скорости обработки 76 кадров в секунду, при этом средняя точность алгоритма НАООУХ уменьшилась не более чем на 0,1%.

Таким образом, применяя процедуру квантования весов СНС до половинной точности представления чисел с плавающей точкой, можно существенно улучшить производительность работы нейросетевых алгоритмов, что позволит использовать вычислительно более сложные архитектуры сетей для анализа видеопотока с эндоскопической системы в режиме реального времени, а также снизить требования к аппаратной части модуля анализа видеопотока.

### **3.5 Разработка алгоритма обнаружения полипов на видеоданных с постобработкой результатов**

Простой процесс применения алгоритма обнаружения объектов к каждому кадру видеопоследовательности, описанный в п. 3.3, хоть и позволяет анализировать видеопоток, но такой подход не использует корреляцию соседних кадров видеопоследовательности. Учесть эту особенность можно с помощью добавления в алгоритм обнаружения объектов этапа постобработки. При таком подходе выполняется связывание обнаружений между кадрами, как описано в работах [43, 76] и предлагается использовать классификатор на основе логистической регрессии как предложено в [134], чтобы различать, принадлежат ли два обнаружения одному и тому же экземпляру объекта или нет. Данный подход может использоваться с любым алгоритмом обнаружения объектов на изображениях. В данной главе далее будет использоваться алгоритм НАООУХ, который продемонстрировал наилучший результат в экспериментах из раздела 3.3.



Структурная схема всего алгоритма постобработки обнаружений на видеопотоке на основе обучаемого классификатора изображена на Рисунке 3.3 а. На первом этапе происходит составление пар обнаружений, полученных с соседних кадров видеопотока. Данный шаг более подробно изображен на Рисунке 3.3 б.

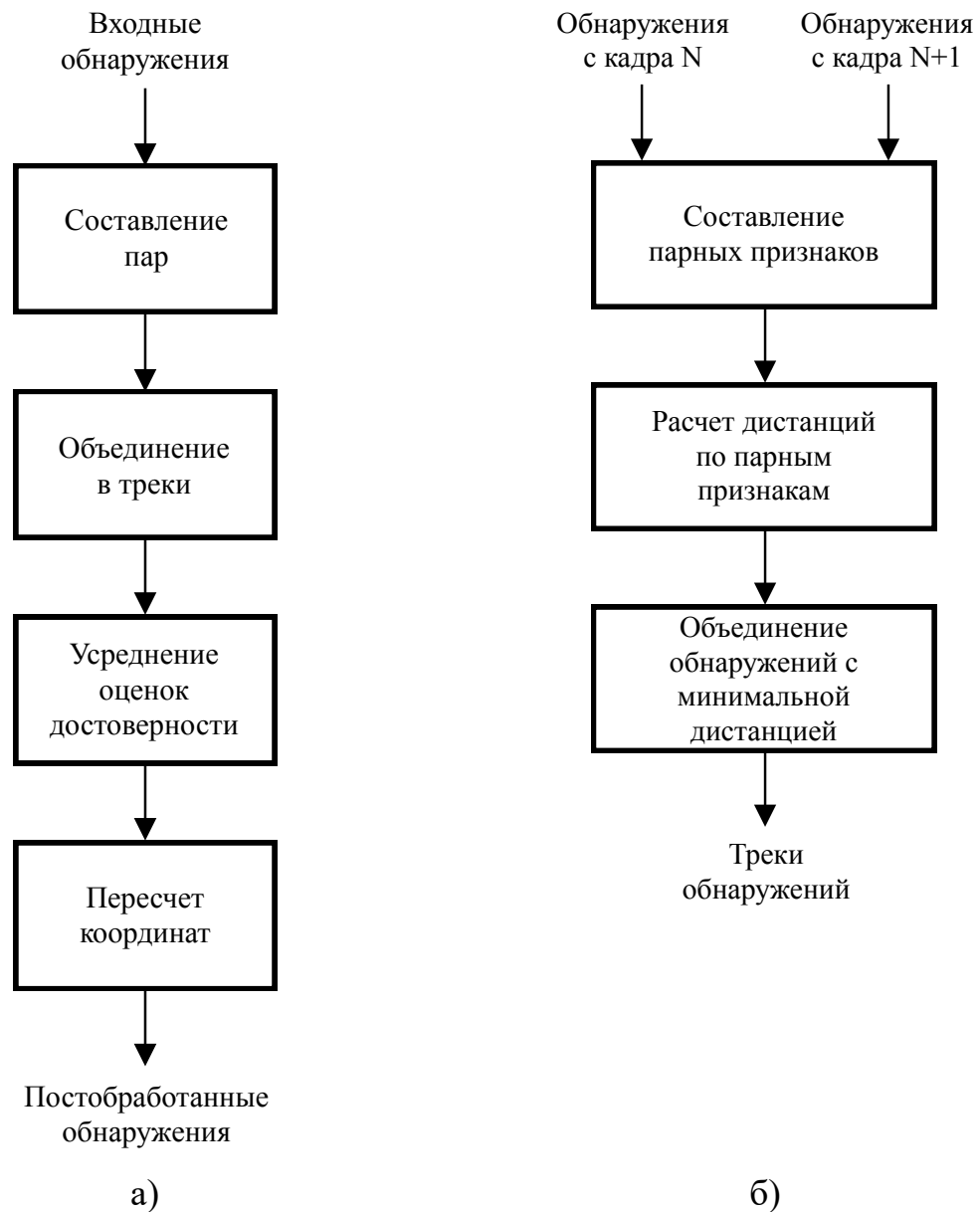


Рисунок 3.3 – Структурные схемы: а) процедуры постобработки обнаружений на видеоданных; б) процедуры объединения обнаружений в треки

Применяя алгоритм НАООУХ в каждом кадре  $t$  видеопотока, получаем набор обнаружений, каждое из которых ( $d_t^i$ ) имеет следующие свойства:

- Расположение и геометрия, т.е. информация об ограничивающей его рамке  $bb_t^i = \{x, y, w, h\}$ .
- Семантическая информация, представляет собой вектор оценок достоверности принадлежности обнаружения к определенному классу объектов  $cc_t^i$ . В частности, в алгоритме НАООУХ рассматривается один класс – полип, поэтому в применении к этому алгоритму длина вектора будет равна единице.

Полученные обнаружения связываются в треки, то есть наборы обнаружений одного и того же экземпляра объекта по всем кадрам видеопотока, как предложено в [43]. Процесс связывания происходит последовательно от кадра к кадру видеоданных. Построение треков начинается с обнаружений между первой парой кадров и продолжается до тех пор, пока соответствующие объекты все еще находятся в следующих кадрах. При этом новые треки могут начинаться в любом кадре с теми обнаружениями, которые не были включены в существующие треки.

Чтобы связать обнаружения между двумя последовательными кадрами, предлагается использовать функцию подобия, основанную на методах машинного обучения, которая объединяет различные признаки и предназначена для повышения робастности к различным скоростям движения объектов. Данная функция подобия определяется через следующие парные признаки, вычисляемые для каждой возможной пары обнаружений  $d_t^i$  и  $d_{t+1}^j$ :

$$f_{loc} = \{IoU, d_{centers}\}, \quad (3.1)$$

$$f_{geo} = \{ratio_w, ratio_h\}, \quad (3.2)$$

$$f_{sem} = cc_t^i \cdot cc_{t+1}^j, \quad (3.3)$$

где  $IoU$  – отношение площади пересечения между обнаружениями к площади объединения обнаружений,  $d_{centers}$  – евклидово расстояние между центрами обнаружений,  $ratio_w$  и  $ratio_h$  – соотношения ширин и высот обнаружений, соответственно,  $cc_t^i$  и  $cc_{t+1}^j$  – вектора оценок достоверности принадлежности к определенному классу объектов для рассматриваемой пары обнаружений в текущем и следующем кадрах, а  $f_{sem}$  – их скалярное произведение.

Таким образом, используя выражения (3.1–3.3), функцию подобия можно определить как:

$$L(d_t^i, d_{t+1}^j) = f_{sem} \cdot \varphi(f_{loc}, f_{geo}), \quad (3.4)$$

где  $\varphi$  – классификатор на основе логистической регрессии, предсказывающий принадлежат ли два обнаружения одному и тому же экземпляру объекта, учитывая их парные признаки.

Определим также величину, обратную предложенной функции подобия, – дистанцию между двумя обнаружениями:

$$D(d_t^i, d_{t+1}^j) = 1 - L(d_t^i, d_{t+1}^j). \quad (3.5)$$

Далее рассчитывается матрица дистанций, каждый элемент которой – дистанция между парой обнаружений с индексами  $ij$ , после чего с помощью жадного алгоритма оптимизации происходит связывание обнаружений в трек. Выбирается наименьшая дистанция, обнаружения соответствующие данной дистанции связываются в трек, а затем соответствующие данной дистанции индексы исключаются. Процесс повторяется до тех пор, пока все индексы не будут исключены, или не будут исключены индексы, соответствующие обнаружениям одного из рассматриваемых кадров, так что образовать новую связь невозможно. Для каждой пары кадров создается новый трек, если рассматриваемая пара обнаружений не является продолжением уже существующего трека.

Поскольку алгоритм НАООУХ, как и другие алгоритмы обнаружения объектов, не всегда делает точные прогнозы, некоторые обнаружения склонны генерировать ложноположительные треки. Как правило, в этом

случае алгоритм выдает низкое значение функции подобия. Пары, которые имеют значение функции подобия ниже установленного порога, отбрасываются.

После получения треков происходит усреднение векторов оценок достоверности принадлежности к определенному классу объектов между обнаружениями в треке, что дает возможность исправлять ошибочно классифицированные обнаружения или устранять неоднозначность с низким значением оценки достоверности принадлежности обнаружения к определенному классу, например в кадрах, где присутствует потеря фокусировки за счет обнаружений в хорошо сфокусированных кадрах, в которых объект обнаруживается уверенно. Опционально производится пересчет координат обнаружений в треке. Из-за быстрого движения объекта на видео, в некоторых кадрах точность локализация объекта может снижаться. Тогда можно рассматривать каждую координату тренируемого объекта как зашумленный временной ряд. Чтобы удалить или ослабить такой шум, используется сглаживание, в частности, применяется одномерный фильтр Гаусса вдоль каждого временного ряда. Затем сглаженный ряд используется как набор координат объекта в треке.

На Рисунке 3.4 изображены зависимости точности от полноты для алгоритма  $HAOO_{YX}$  без применения этапа постобработки и с постобработкой (алгоритм  $HAOO_{YXVID}$ ), полученные на базе ЯОКОБ.

Из приведенных зависимостей следует, что алгоритм постобработки позволил улучшить качество обнаружения полипов на видеопоследовательностях, значение метрики  $AP@0,50$  выросло с 81,6% до 83,3%. Кроме того, информация о треках полипов позволяет в дальнейшем реализовать автоматизированный подсчет числа полипов, осуществлять выбор лучшего кадра из трека для составления медицинского отчета, создавать аннотации для навигации по видеоданным исследования и т.д.

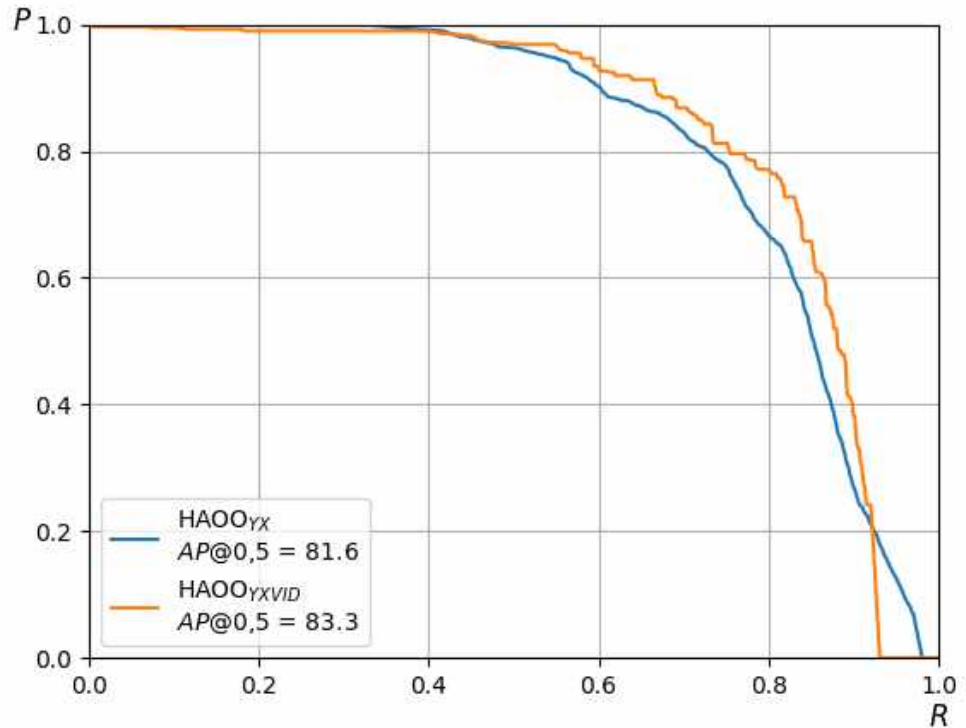


Рисунок 3.4 – Зависимости точности от полноты для алгоритмов HAOO<sub>YX</sub> и HAOO<sub>YXVID</sub>, полученные на базе ЯОКОБ

Пример работы алгоритма постобработки на серии последовательных кадров приведен на Рисунке 3.5 б. Для сравнения на Рисунке 3.5 а приведены исходные результаты базового алгоритма обнаружения полипов.

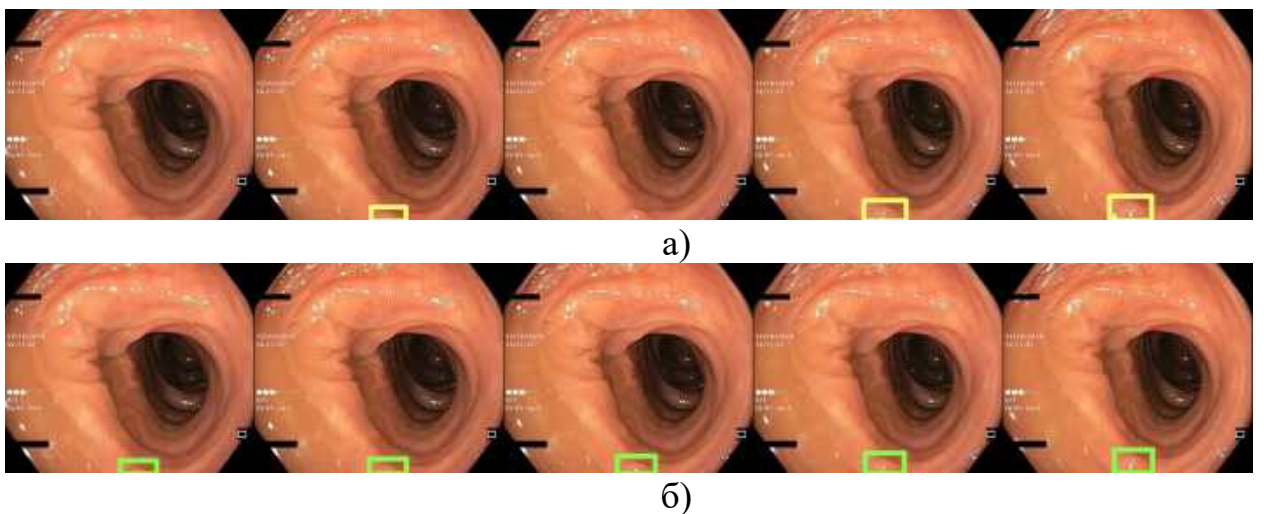


Рисунок 3.5 – Визуализация работы алгоритмов на последовательности кадров: а) HAOO<sub>YX</sub>; б) HAOO<sub>YXVID</sub>

На данной визуализации видно, что недостающие обнаружения базового алгоритма восстановлены за счет обнаружений из соседних кадров с помощью алгоритма постобработки. Данное усовершенствование алгоритма является чрезвычайно важным для практического использования его в работе врача-эндоскописта, для которого «мерцающие» обнаружения являются крайне нежелательными.

### **3.6 Создание и разметка базы видеоданных колоноскопических исследований в реальной эндоскопической системе прикладного телевидения**

Стоит напомнить, что база видеопоследовательностей, собранная из архивных видеоматериалов, имеет существенные различия с видеоданными на выходе реальной современной эндоскопической системы. Поэтому возникла задача сбора базы видеоданных в реальных условиях работы системы эндоскопического прикладного телевидения.

Одним из ведущих производителей эндоскопического оборудования является фирма Olympus. В частности, отделения Ярославской областной клинической онкологической больницы оборудованы эндоскопическими системами EVIS EXERA II и EVIS EXERA III. Эндоскопическая система EVIS EXERA II оборудована HDTV видеопроцессором CV-180 и ксеноновым источником света CLV-180. Эндоскопическая система EVIS EXERA III оборудована HDTV видеопроцессором CV-190 и ксеноновым источником света CLV-190. Видеопроцессоры CV-180 и CV-190 имеют аналоговые видеовыходы HDTV RGB и YPbPr, аналоговые видеовыходы SDTV VBS, Y/C и RGB, а также цифровые видеовыходы HD-SDI (SMPTE 292M), SD-SDI (SMPTE 259M), DV (IEEE 1394) и DVI (WUXGA, 1080p или SXGA). Более подробные технические характеристики видеопроцессоров CV-180 и CV-190 приведены в Приложении А к настоящей работе.

### **3.6.1. Разработка программно-аппаратного тестового стенда с модулем анализа видеопотока**

Для проведения исследований в условиях, максимально приближенных к практическим, необходимо производить захват цифрового видеопотока с эндоскопической системы. Для этого нужно оборудовать компьютер, на котором будет происходить обработка видеопотока, устройством видеозахвата. В качестве такого устройства использована PCI express плата видеозахвата BLACKMAGIC серии DeckLink, с помощью которой можно подключить эндоскопическую систему к компьютеру для обработки видеопотока через цифровой последовательный интерфейс SDI.

В соответствии со стандартом SD-SDI SMPTE 259M, может быть обеспечена передача видео с номинальной скоростью передачи данных 270 Мбит/с, 360 Мбит/с, 143 Мбит/с и 177 Мбит/с, что позволит передавать видео с разрешением 480i, 576i. В соответствии со стандартом HD-SDI SMPTE 292M, может быть обеспечена передача видео с номинальной скоростью передачи данных 1,485 Гбит/с и 1,485/1,001 Гбит/с, что позволит передавать видео с разрешением 720p, 1080i. Модель платы видеозахвата DeckLink Mini Recorder имеет видеовход SDI с поддержкой стандартов SMPTE 292M и SMPTE 259M, что позволяет использовать ее для видеозахвата сигнала с эндоскопической системы.

Таким образом, для анализа алгоритма обнаружения полипов на видеопотоке с эндоскопической системы собран специальный стенд, оборудованный платой видеозахвата. Он дополняет эндоскопическую систему модулем анализа видеопотока и дополнительным монитором, на который выводится обработанный видеопоток. Для удобства использования врачом-эндоскопистом стенд оснащен двумя педалями, позволяющими управлять процессом записи и отмечать временные промежутки присутствия полипа в кадрах видеопотока. Структурная схема данного стенда изображена на Рисунке 3.6.

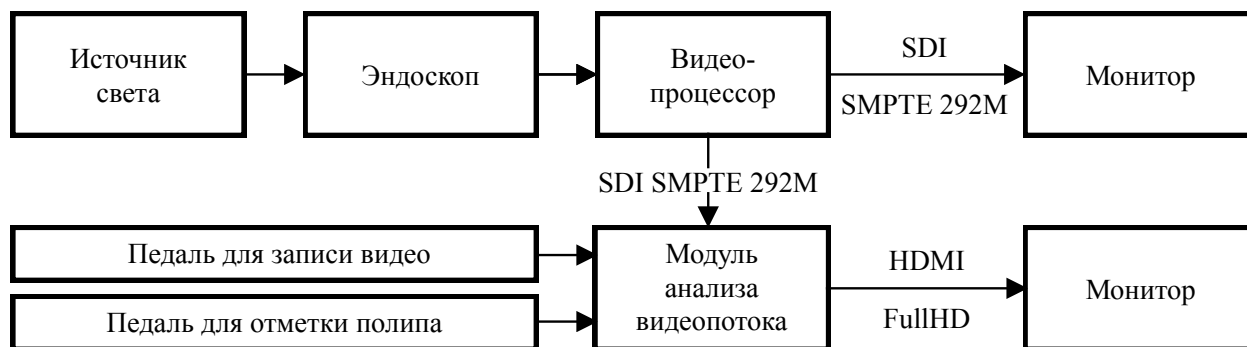


Рисунок 3.6 – Схема разработанного стенда с модулем анализа видеопотока

С помощью данного стенда в ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России (Пироговский центр, г. Москва) во временном интервале с сентября 2021 года по февраль 2022 года записаны 450 видеопоследовательностей колоноскопических исследований в условиях реальной клинической практики. Из них 243 исследования не содержали полипы и не обрабатывались, а 207 оставшихся – содержали полипы, временные интервалы которых отмечались врачами-эндоскопистами с помощью педалей на тестовом стенде. Длительность видеозаписей варьировалась от 6 секунд до 29 минут, а время присутствия в кадре полипа – от 8 секунд до 10 минут. Разрешение видеозаписей составляло  $1400 \times 1080$  пикселей (в оригинале FULL HD  $1920 \times 1080$  пикселей, но обрезалась неинформативная область кадров с персональными данными пациентов). Из данных видеозаписей отобраны 43 наиболее информативных, которые разбивались на кадры, в каждом из которых отмечались местоположения полипов. Таким образом, собрана тестовая база видеофрагментов (ПРГВ), состоящая из 31 020 кадров и 27 133 экспертно размеченных рамок, локализирующих полипы в кадрах.

Рассмотрим процесс экспертной разметки областей полипов более подробно.



### 3.6.2. Разработка полуавтоматической методики разметки видеоизображений

Для того чтобы снизить огромные трудозатраты эндоскопистов экспертного класса в процессе разметки видеофрагментов предложена методика полуавтоматической разметки областей интереса для последующего обучения и/или тестирования нейросетевых моделей. Данная методика предполагает формирование основы для разметки алгоритмом, обученным на предыдущих этапах исследования, с доработкой полученной разметки специалистом по работе с изображениями («разметчик»), который, в свою очередь, консультируется с врачами-эндоскопистами.

Структурная схема предложенной методики приведена на Рисунке 3.7.

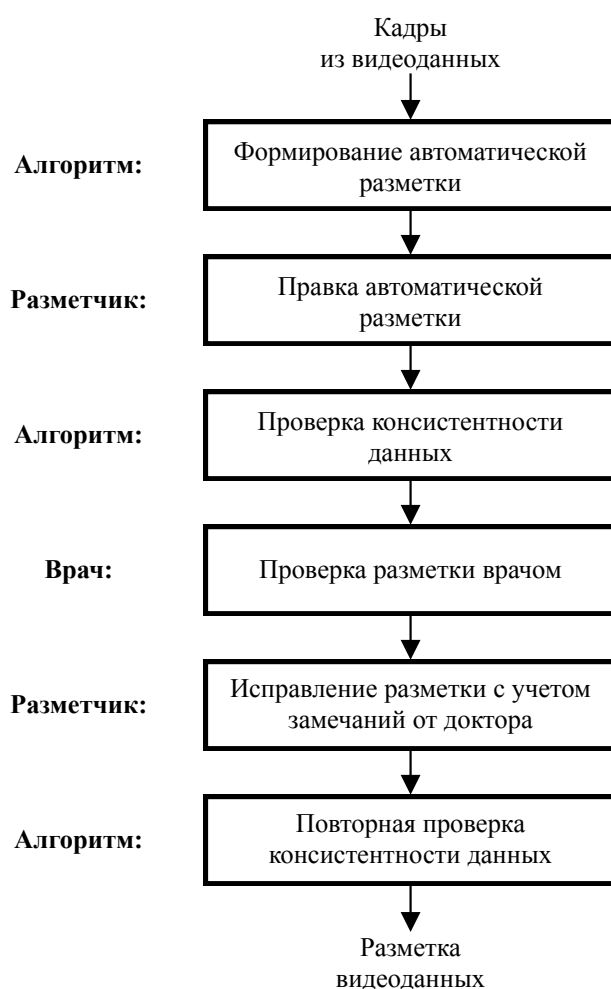


Рисунок 3.7 – Структурная схема методики полуавтоматической разметки областей интереса на видеоданных для последующего обучения и/или тестирования нейросетевых моделей

Она состоит из 6 этапов:

1. Разметка видеофрагментов обученным на предыдущих этапах алгоритмом  $HAOO_{YXVID}$ . Сохраняются кадры с треками полипов, с добавлением по 5 кадров до появления полипа в кадре и после, чтобы сохранялась контекстная информация для дальнейших этапов разметки. Отбор предсказанных алгоритмом ограничивающих рамок, для которых оценка достоверности принадлежности объекта к классу полип превышает 0,3; присвоение класса `detection` и сохранение вместе с отметками врача в специальном текстовом формате сериализации – `json`, с которым можно далее работать в программе разметки `labelme` [28].
2. Разметчик просматривает полученные последовательности кадров с разметкой в хронологическом порядке и удаляет ложноположительные предсказания алгоритма, ошибочность которых для него очевидна и объединяет в один трек треки одного и того же объекта. Если объект на рассматриваемой последовательности кадров соответствует отметке врача, полученной с тестового стенда, любой рамке из данного трека присваивается класс `polyp`, а соответствующей отметке врача – идентификационный номер трека рассматриваемого объекта.
3. Проверяется согласованность данных – все объекты одного трека приводятся к одному классу. Выбираются все треки, соответствующие классу `detection` (объекты, которым не нашлось соответствующей отметки от врача), и последовательности кадров с данными треками сохраняются в отдельные директории. При этом на кадры наносятся ограничивающие рамки, визуально локализирующие трекируемый (сопровождаемый) объект.
4. Врач или группа врачей рассматривают полученные на предыдущем этапе директории и выносят экспертную оценку, соответствует ли

объект на серии кадров полипу или нет, а также отмечают новый полип, в случае его пропуска на предыдущих этапах.

5. Разметчик, руководствуясь результатами экспертизы, полученной на предыдущем этапе, присваивает трекам с классом detection класс polyp, если это соответствует экспертной разметке, а остальные треки удаляет. Если присутствуют пропущенные полипы, то он размечает их в ручном режиме.

6. Снова проверяется согласованность данных.

Руководствуясь данной методикой, выполнена разметка видеофрагментов с эндоскопической системы с целью формирования тестового набора изображений из реальной клинической практики.

Стоит отметить, что данную методику можно применять итеративно для разметки большого количества видеоданных. На каждой итерации стоит размечать часть данных и дообучать алгоритм на новых данных, тогда следующая итерация будет требовать меньше трудозатрат, т.к. качество автоматической разметки будет возрастать. Как показывает практика трудозатраты при создании размеченных баз изображений уменьшаются на 30-40%.

Предложенную методику можно применять для разметки других видов эндоскопических видеоданных практически без изменения, а для других видов медицинских видеоданных с небольшой адаптацией в пунктах, связанных с консультациями с профильными специалистами.

### **3.7 Тестирование итогового нейросетевого алгоритма обнаружения**

Для оценки качества работы финального нейросетевого алгоритма обнаружения на базе ПРГВ использовались два подхода. В первом из них использовались метрики средней точности для каждого обнаружения алгоритма, независимо от остальных. На Рисунке 3.8 изображена кривая зависимости точности от полноты для алгоритма  $HAOO_{YXVID}$ . Значение

метрики  $AP@0,50$  составило 78,3%. Дополнительно, на Рисунке 3.9 приведена кривая зависимости  $F1$  от порога оценки достоверности. Наибольшее значение  $F1$  составило 74,1%.

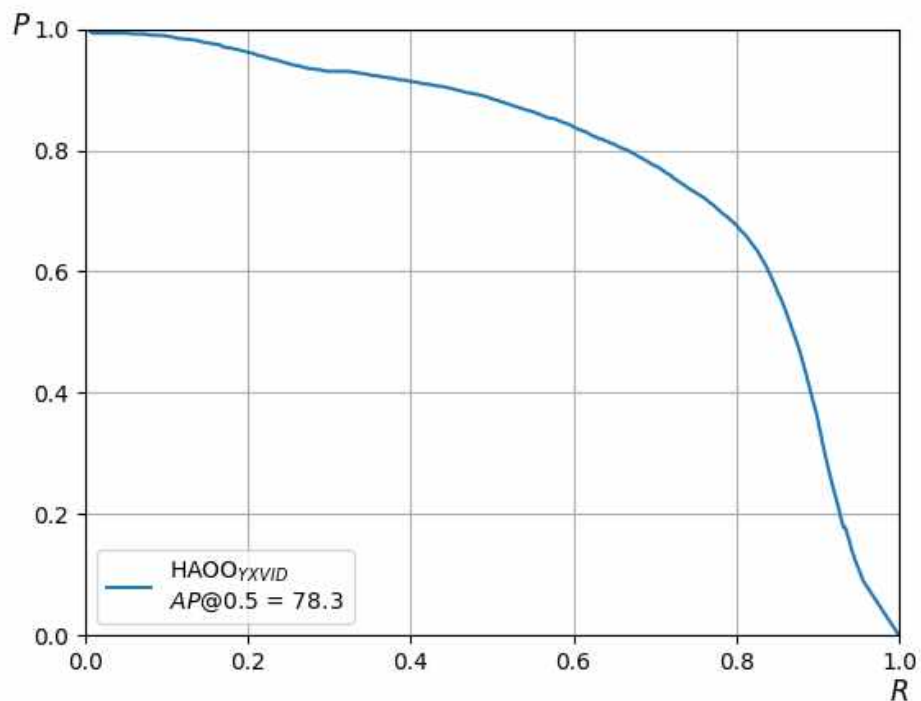


Рисунок 3.8 – Зависимость точности от полноты для алгоритма HAOO\_UXVID, полученная на базе ПРГВ

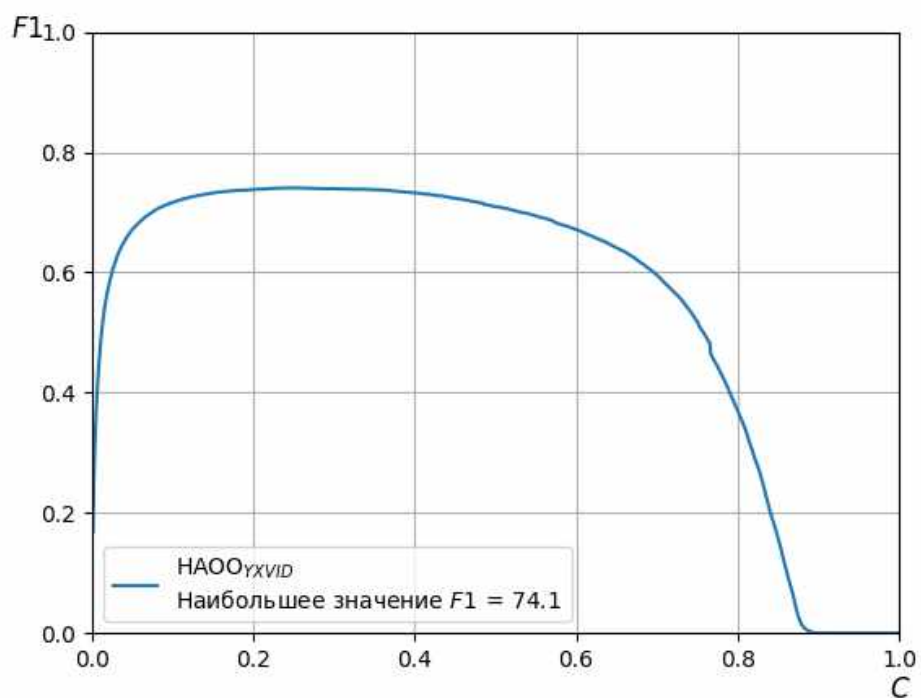


Рисунок 3.9 – Зависимость метрики F1 от порога оценки достоверности

Из приведенных выше зависимостей следует, что метрика средней точности на базе ПРГВ достигает меньшего значения по сравнению с аналогичной метрикой, полученной на базе ЯОКОБ. Это можно объяснить тем, что база ПРГВ является более сложной для анализа, так как в ней присутствует существенно большее разнообразие изображений полипов, связанное с наличием большего числа исследований различных пациентов.

Второй подход учитывает тот факт, что на практике важными являются не только метрики, оценивающие качество обнаружения полипов в каждом кадре, но также и метрики, которые оценивают, был ли обнаружен тот или иной полип, даже если он обнаружен не во всех кадрах видеопоследовательности [17]. В этом случае для подобной оценки, как и раньше, использовались метрики точности, полноты, средней точности и  $F1$ , но с другим определением значений параметров, на основе которых они рассчитываются. Для наглядности добавим к данным метрикам и параметрам индекс *track*.

Определим  $TP_{track}$  как число треков из экспертной разметки, для которых есть пересечение с треками, предсказанными алгоритмом,  $FN_{track}$  – как число треков из экспертной разметки, для которых нет пересечений с треками, предсказанными алгоритмом, а  $FP_{track}$  – как число треков, предсказанных алгоритмом, для которых нет пересечений с треками из экспертной разметки. Тогда, используя вышеопределенные параметры в формулах 1.4, 1.5, 1.6 и 1.9, получим потрековые метрики точности, полноты, средней точности и  $F1$ :  $P_{track}$ ,  $R_{track}$ ,  $AP@0.5_{track}$  и  $F1_{track}$ .

На Рисунке 3.10 изображена кривая зависимости точности от полноты для алгоритма НАООУХVID при потрековом анализе. Значение метрики  $AP@0,50_{track}$  составило 81,0%. Дополнительно на Рисунке 3.11 приведена кривая зависимости метрики  $F1_{track}$  от порога оценки достоверности. В данном случае наибольшее значение  $F1_{track}$  составило 82,9%.

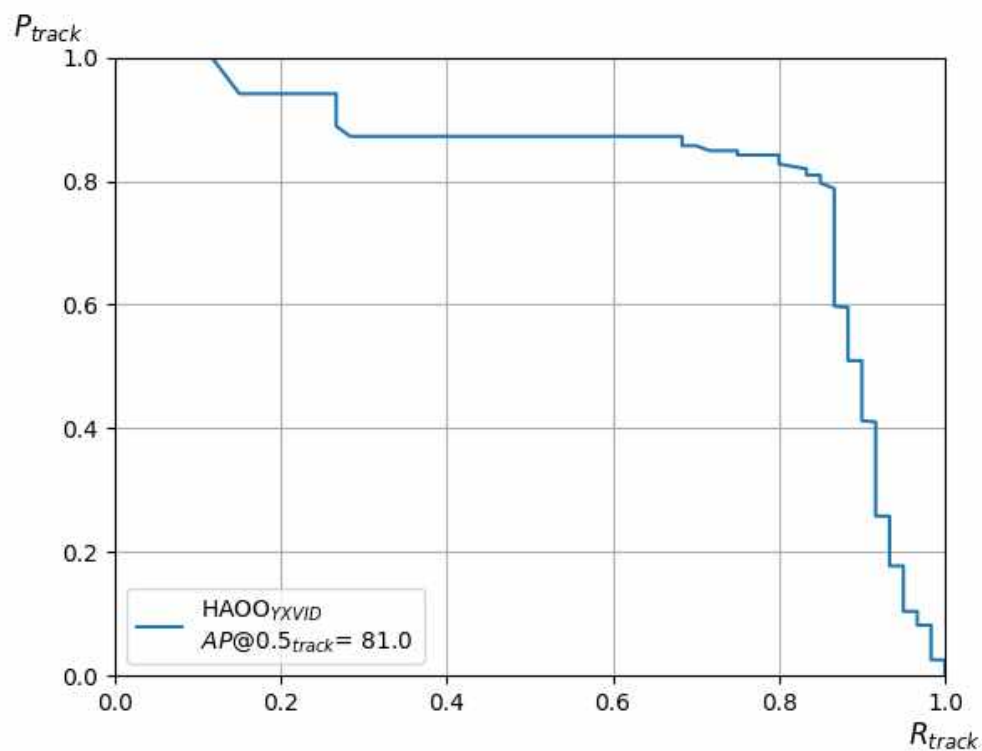


Рисунок 3.10 – Зависимость метрики точности  $P_{track}$  от метрики полноты  $R_{track}$  для алгоритма HAOO<sub>UXVID</sub>, полученная на базе ПРГВ

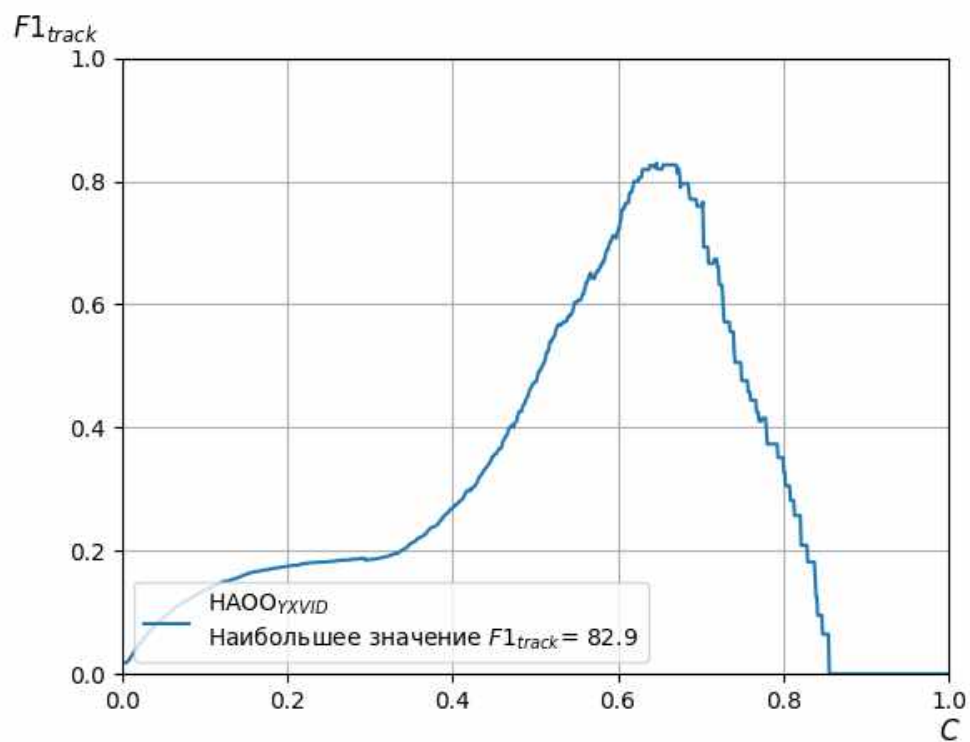


Рисунок 3.11 – Зависимость метрики  $F1_{track}$  от порога оценки достоверности

Для итоговой всесторонней оценки алгоритма  $\text{HAOO}_{\text{YXVID}}$  используются метрики средней точности, точности, полноты и среднего гармонического точности и полноты:  $AP@0,50$ ,  $P$ ,  $R$ , и  $F1$ . При этом для расчета  $P$  и  $R$ , из множества значений  $F1$  выбрано максимальное значение, а  $P$  и  $R$  соответствуют выбранному значению  $F1$ .

Таким способом, указанные выше метрики рассчитаны для баз ЯОКОБ и ПРГВ. Также для базы ПРГВ рассчитаны потрековые аналоги выбранных метрик. Их значения приведены в Таблице 3.1.

Таблица 3.1

Итоговые результаты тестирования алгоритма  $\text{HAOO}_{\text{YXVID}}$ 

Метрика	База ЯОКОБ	База ПРГВ	База ПРГВ (track)
$AP@0,50$	83,3	78,3	81,0
$P$	76,4	72,5	82,0
$R$	81,0	75,7	83,3
$F1$	78,6	74,1	82,6

Из анализа научно-технической литературы следует, что в среднем врач-эндоскопист во время диагностики обнаруживает порядка 80% полипов [91], что соответствует  $R_{\text{track}}=80,0\%$ . Таким образом, на основании приведенных результатов можно сделать вывод о том, что итоговый нейросетевой алгоритм демонстрирует результат обнаружения полипов, сравнимый с работой врача-эндоскописта.

Структурная схема итогового нейросетевого алгоритма анализа видеопотока в эндоскопической системе приведена на Рисунке 3.12. Ее можно применить почти без изменений для обнаружения других типов объектов на эндоскопических видеоданных. Потребуется небольшая модификация первого блока для разработки алгоритмов анализа других медицинских видеоданных.

Перенос данной схемы на другие прикладные области анализа видеопотока может потребовать более существенной адаптации, однако структурно процесс разработки нейросетевого алгоритма обнаружения объектов будет отличаться незначительно.



Рисунок 3.12 – Структурная схема разработки алгоритма обнаружения объектов в видеопотоке с эндоскопической системы

### 3.8 Краткие выводы

Результаты проведенных исследований в части разработки алгоритмов обнаружения объектов на видеоизображениях колоноскопических исследований позволяют сделать следующие основные выводы:

- Для обучения и тестирования алгоритма на видеопоследовательностях совместно с врачами-эндоскопистами создана база видеоданных из видеоархивов колоноскопических исследований (ЯОКОБ). Она состоит из обучающего и тестового наборов, видеофрагменты которых разбиты на кадры с экспертной разметкой полипов. Таким образом, обучающий



набор содержит 165 665 изображений, из которых 114 754 содержат полипы, а тестовый – 11 192 изображения, из которых 7057 содержат полипы.

- Предложенные нейросетевые алгоритмы переобучены и протестированы на базе изображений ЯОКОБ. Наилучший результат на тестовом наборе получен для алгоритма  $\text{HAOO}_{\text{YX}}$ , где значение метрики средней точности работы алгоритма составило 81,6%.
- Применяя алгоритмы квантования весов СНС к алгоритму  $\text{HAOO}_{\text{YX}}$ , удалось уменьшить среднее время обработки одного кадра видеопоследовательности с 22,8 мс до 13,1 мс, что эквивалентно достижению скорости обработки – 76 кадров в секунду. При этом значение метрики  $AP@0,50$  деградировало не более чем на 0,1%. Таким образом, алгоритм  $\text{HAOO}_{\text{YX}}$  после применения методов квантования весов СНС можно использовать для обработки видеопотока с эндоскопической системы в режиме реального времени.
- Разработан алгоритм обнаружения полипов для работы на видеопоследовательностях, который позволил улучшить значение метрики  $AP@0,50$  на базе видеоданных ЯОКОБ с 81,6% до 83,3%. Кроме того, в результате работы алгоритма  $\text{HAOO}_{\text{YXVID}}$  появляется информация о треках объектов, что позволяет использовать данные результаты для дополнительной аналитики, например для автоматизированного подсчета числа полипов, выбора лучшего кадра из трека, создания аннотаций для навигации по видеоданным эндоскопического исследования и т.д.
- Предложена и апробирована на наборе из более чем 31 000 видеок кадров методика полуавтоматической подготовки базы видеоданных колоноскопических исследований с разметкой областей интереса для обучения нейросетевых моделей, использованная для создания базы видеоданных ПРГВ, максимально приближенной к

реальным клиническим условиям. Соответствующие трудозатраты при использовании предложенной методики сокращаются на 30-40%.

- Тестирование алгоритма  $\text{HAOO}_{\text{YXVID}}$  на базе видеоданных ПРГВ дает следующие значения метрик оценки качества:  $AP@0,50=78,3\%$ ,  $P=72,5\%$ ,  $R=75,7\%$ ,  $F1=74,1\%$  – при покадровом анализе работы алгоритма и  $AP@0,50_{\text{track}}=81,0\%$ ,  $P_{\text{track}}=82,0\%$ ,  $R_{\text{track}}=83,3\%$ ,  $F1_{\text{track}}=82,6\%$  – при потрековом анализе работы. Данные результаты являются практически востребованными для задачи обнаружения полипов на подобном классе видеоданных.

## ЗАКЛЮЧЕНИЕ

Основные выводы и результаты диссертационной работы можно сформулировать в следующем виде.

1. Разработаны методика, нейросетевые алгоритмы и программное обеспечение для обнаружения патологических объектов на статических изображениях и видеоданных колоноскопических исследований в медицинской системе прикладного телевидения.
2. Для решения задачи обнаружения полипов на изображениях колоноскопических исследований предложены 3 алгоритма на основе архитектуры нейронной сети SSD и 4 алгоритма на основе архитектуры YOLOv5. Обучение и тестирование нейросетевых алгоритмов проводилось с использованием суперкомпьютера NVIDIA DGX-1, предназначенного для задач глубокого обучения.
3. Предложенная модификация процесса аугментации данных позволила улучшить значение метрики  $AP@0,50$  в среднем на 1,4% для алгоритма на основе сверточной нейронной сети SSD и на 3,0% для алгоритма на основе СНС YOLOv5.
4. В результате тестирования предложенных алгоритмов обнаружения патологических объектов на изображениях с использованием открытого набора изображений Kvasir-SEG, наилучший результат получен для предложенного нейросетевого алгоритма  $HAOO_{YL}$ , для которого значение метрики  $AP@[0,25..0,75]$  составило 92,5%, что превосходит известные аналоги на 5% и более. Некоторым недостатком такого алгоритма является повышенная в 3 раза вычислительная сложность по сравнению с известным из литературы алгоритмом ColonSegNet.
5. Для обучения и тестирования алгоритма на видеопоследовательностях совместно с врачами-эндоскопистами создана база видеоданных из видеоархивов колоноскопических исследований (ЯОКОБ). Она состоит из обучающего и тестового наборов, видеофрагменты которых разбиты

на кадры с экспертной разметкой полипов. Таким образом, обучающий набор содержит 165 665 изображений, из которых 114 754 содержат полипы, а тестовый – 11 192 изображения, из которых 7057 содержат полипы.

6. Предложенные нейросетевые алгоритмы переобучены и протестированы на базе изображений ЯОКОБ. Наилучший результат на тестовом наборе получен для алгоритма  $HAOO_{YX}$ , где значение метрики средней точности работы алгоритма составило 81,6%.
7. Применяя алгоритмы квантования весов СНС к алгоритму  $HAOO_{YX}$ , удалось уменьшить среднее время обработки одного кадра видеопоследовательности с 22,8 мс до 13,1 мс, что эквивалентно достижению скорости обработки – 76 кадров в секунду. При этом значение метрики  $AP@0,50$  деградировало не более чем на 0,1%. Таким образом, алгоритм  $HAOO_{YX}$  после применения методов квантования весов СНС можно использовать для обработки видеопотока с эндоскопической системы в режиме реального времени.
8. Разработан алгоритм обнаружения полипов для работы на видеопоследовательностях, который позволил улучшить значение метрики  $AP@0,50$  на базе видеоданных ЯОКОБ с 81,6% до 83,3%. Кроме того, в результате работы алгоритма  $HAOO_{YXVID}$  появляется информация о треках объектов, что позволяет использовать данные результаты для дополнительной аналитики, например для автоматизированного подсчета числа полипов, выбора лучшего кадра из трека, создания аннотаций для навигации по видеоданным эндоскопического исследования и т.д.
9. Предложена и апробирована на наборе из более чем 31 000 видеокадров методика полуавтоматической подготовки базы видеоданных колоноскопических исследований с разметкой областей интереса для обучения и/или тестирования нейросетевых моделей, использованная для создания базы видеоданных ПРГВ, максимально

приближенной к реальным клиническим условиям. Соответствующие трудозатраты при использовании предложенной методики сокращаются на 30-40%.

10. Тестирование алгоритма  $\text{HAOO}_{\text{YXVID}}$  на базе видеоданных ПРГВ дает следующие значения метрик оценки качества:  $AP@0,50=78,3\%$ ,  $P=72,5\%$ ,  $R=75,7\%$ ,  $F1=74,1\%$  – при покадровом анализе работы алгоритма и  $AP@0,50_{\text{track}}=81,0\%$ ,  $P_{\text{track}}=82,0\%$ ,  $R_{\text{track}}=83,3\%$ ,  $F1_{\text{track}}=82,6\%$  – при потрековом анализе работы. Данные результаты являются практически востребованными для задачи обнаружения полипов на подобном классе видеоданных.

11. Цель и задачи диссертационной работы успешно выполнены. Разработанные нейросетевые алгоритмы обнаружения объектов интегрированы в следующие программы для ЭВМ:

- AI.ColonoscopySolution 1.0 – программа для анализа колоноскопических изображений на основе технологий машинного обучения;
- Endoscopy\_Polyps 1.0 – программа для детектирования и распознавания полипов в кишечнике с использованием эндоскопических изображений.

Для данных программных продуктов получены свидетельства о государственной регистрации программы для ЭВМ (Приложение Б). Также для созданной в процессе работы базы видеоданных колоноскопических исследований получено свидетельство государственной регистрации баз данных:

- Colonoscopy.DB – база данных эндоскопических изображений толстой кишки.

## ЛИТЕРАТУРА

1. Авшаров Е.М., Абгарян М. Г., Сафарянц С.А. Обработка медицинских изображений как необходимый инструментарий медицинского диагностического процесса // Вестник рентгенологии и радиологии. – 2010. – № 3. – С. 54–61.
2. Батухтин Д.М., Пеганова Е.В., Митракова Н.Н. Анализ узкоспектральных эндоскопических изображений на внутренней поверхности пищевода // Вестник Поволжского государственного технологического университета. Серия: радиотехнические и инфокоммуникационные системы. – 2014. – № 4 (23). – С. 45–57.
3. Безруков В.Н., Балобанов В.Г. Системы цифрового вещательного и прикладного телевидения. Учебное пособие для вузов. М.: Горячая линия – Телеком, 2015. – 608 с.
4. Блинов Н.Н., Мазуров А.И. Системы прикладного телевидения: (Цветное телевидение в биологии и медицине). – М.: Знание, 1987. - 62 с.
5. Блинов, Н. Н. Микропроцессорная техника и ЭВМ в медицинской интроскопии Н. Н. Блинов. - М.: Знание, 1986. – 64 с.: ил. - (Новое в жизни, науке, технике. Радиоэлектроника и связь; вып. 4).
6. Грязин Г.Н. Системы прикладного телевидения / СПб.: Политехника, 2000. – 277 с.
7. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение // ДМК-Пресс, 2017. – 652 с.
8. Джакония В.Е., Гоголь А.А., Друзин Я.В. Телевидение: Учебник для вузов. М: Горячая линия – Телеком, 4-е издание. 2007. – 616 с.
9. Завьялов Д.В., Кашин Н.В., Нестеров П.В., Нечипай А.М., Рыбачков В.В., Малашенко В.Н., Хрящев В.В. Алгоритм уточняющей диагностики и внутрипросветного эндоскопического удаления эпителиальных новообразований толстой кишки // Колопроктология. – 2021. – №20(1). – С. 17-22.

10. Кирсанова А.В. Современное состояние и перспективы развития экспертных медицинских систем // Новый университет. Серия «Технические науки», 2015, – № 11-12. – С. 45-46.
11. Кондратьев А.Г., Лукин М.И. Техника промышленного телевидения. – Лениздат, 1976. – 503 с.
12. Королев М.П., Кашин С.В., Завьялов Д.В., Рассадина Е.А. и др. Принципы организации и критерии качества колоноскопии, выполняемой с целью скрининга колоректального рака. Электронный ресурс]. Режим доступа: [http://www.endoscopy-uroslavl.ru/content/100513\\_rekom](http://www.endoscopy-uroslavl.ru/content/100513_rekom) (Дата обращения: 25.03.2022).
13. Коротаев В.В., Краснящих А.В. Телевизионные измерительные системы / Учебное пособие. – СПб: СПбГУ ИТМО, 2008. – 108 стр.
14. Куваев Р.О., Кашин С.В. Современное эндоскопическое исследование желудка с использованием методик узкоспектральной и увеличительной эндоскопии: техника проведения и алгоритмы диагностики // Доказательная гастроэнтерология. – 2016. – Т. 2, № 5. – С. 3–13.
15. Лебедев А. А., Хрящев В. В., Кашин С. В., Среднякова А. С., Казина Е. М. Применение методов глубокого обучения для поддержки врачебного решения при эндоскопическом исследовании желудка // Модели, системы, сети в экономике, технике, природе и обществе. – 2021. – №2(38). – С. 95-106.
16. Лебедев А.А. Разработка алгоритма детектирования полипов на видеопотоке с эндоскопической системы с использованием методов глубокого машинного обучения // Перспективные технологии в средствах передачи информации - ПТСПИ-2021: Материалы 14-ой международной научно-технической конференции. – 2021. – С. 400-403.
17. Лебедев А.А. Разработка нейросетевого алгоритма обнаружения объектов на видеоизображениях в системах прикладного телевидения //

- Радиоэлектронные устройства и системы для инфокоммуникационных технологий (REDS–2022): докл. 77-й всероссийской конференции (с международным участием). – Москва. – 2022.
18. Лебедев А.А., Приоров А.Л., Андержанова А.С., Среднякова А.С. Разработка алгоритмов детектирования полипов на эндоскопических изображениях // Проблемы информатики в образовании, управлении, экономике и технике: сборник статей XXI Международной научно-технической конференции. – Пенза: Приволжский Дом знаний. – 2021. – С. 51-56
19. Лебедев А.А., Приоров А.Л., Хрящев В.В., Среднякова А.С. Постобработка результатов работы нейросетевого детектора полипов на видеоизображениях колоноскопических исследований // Цифровая обработка сигналов. – 2021. – №4. – С. 68-71.
20. Лебедев А.А., Степанова О.А., Юрченко Е.А., Хрящев В.В. Разработка алгоритмов анализа изображений для классификации патологий слизистой оболочки желудка // Цифровая обработка сигналов и ее применение (DSPA-2018): докл. 20-й междунар. конф. – Москва. – 2018. – Т. 8. №3. – С. 223-227.
21. Лебедев А.А., Хрящев В.В. Разработка алгоритма детектирования полипов на видеопотоке с эндоскопической системы в режиме реального времени. 77-я Научно-техническая конференция Санкт-Петербургского НТО РЭС им. А.С. Попова, посвященная Дню радио: сб. докладов [Электронный ресурс] / СПбГЭТУ «ЛЭТИ». – Санкт-Петербург. – 2022. – С. 53-56.
22. Лебедев А.А., Хрящев В.В., Казина Е.М., Среднякова А.С., Журавлева А.С. Распознавание устья червеобразного отростка на эндоскопических изображениях прямой кишки на основе сверточной нейронной сети // Цифровая обработка сигналов и ее применение (DSPA-2020): докл. 22-й междунар. конф. – Москва. – 2020. – С. 638-642.



23. Лебедев А.А., Хрящев В.В., Среднякова А.С., Казина Е.М. Разработка алгоритма детектирования полипов на эндоскопических изображениях с использованием сверточных нейронных сетей // Цифровая обработка сигналов. – 2021. – №2. – С. 55-60.
24. Лебедев А.А., Хрящев В.В., Степанова О.А. Распознавание патологий желудка по видеоизображениям эндоскопического исследования // Новые информационные технологии и системы: Сборник научных статей по материалам XVII Международной научно-технической конференции. – 2020. – С. 197-201.
25. Лукьяница А.А., Шишкин А.Г. Цифровая обработка видеоизображений. – М.: Ай-Эс-Эс Пресс, 2009. – 518 с.
26. М. С. Фролова, С. В. Фролов, И. А. Толстухин Системы поддержки принятия решений для задач оснащения лечебных учреждений медицинской техникой // Университет им. В.И. Вернадского. Специальный выпуск (52). – 2014. – С. 106-111.
27. Нестеров П.В., Ухарский А.В., Кислов Н.В. Региональная клинико-экономическая модель скрининга колоректального рака // Исследования и практика в медицине. – 2020. – № 3, – С. 146-159.
28. Программа для разметки изображений [Электронный ресурс]. Режим доступа: <https://github.com/wkentaro/labelme> (Дата обращения: 25.03.2022).
29. Старков Ю.Г., Солодина Е.Н., Шишин К.В. Эволюция диагностических технологий в эндоскопии и современные возможности выявления опухолей желудочно-кишечного тракта // Тихоокеанский медицинский журнал. – 2009. – Т. 2, № 36. – С. 35 – 39.
30. Степанова О.А., Лебедев А.А., Хрящев В.В., Приоров А.Л. Использование сверточной нейронной сети SSD для детектирования патологий при эндоскопии желудка // Цифровая обработка сигналов и ее применение (DSPA-2019): докл. 21-й междунар. конф. – Москва. – 2019. – С. 533-537.

31. Федотов А.А. Основы цифровой обработки биомедицинских изображений: учеб. пособие / А.А. Федотов. – Самара: Изд-во СГАУ, 2013. – 108 с.
32. Хрящев В.В., Ганин А.Н., Лебедев А.А., Степанова О.А., Кашин С.В., Куваев Р.О. Разработка и анализ алгоритма детектирования патологий на эндоскопических изображениях желудка на основе сверточной нейронной сети // Цифровая обработка сигналов. – 2018. – №3. – С.70-75.
33. Черенков В.Г. Клиническая онкология / Медицинская книга, 2010. – 448 с.
34. Шайн А.А. Онкология: учебник для студентов медицинских вузов / Тюмень: Издат. центр «Академия», 2004. – 544 с.
35. Abel E., Xi W., White P. Methods for removing glare in digital endoscope images // Surg Endosc. – 2011. – vol 25(12). – pp. 3898-905.
36. Ali S. et al. Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy // Medical Image Analysis. – 2021. – vol. 70. – p. 102002.
37. Ali S. et al. Endoscopy artifact detection (EAD 2019) challenge dataset // arXiv preprint arXiv:1905.03209, 2019.
38. Ali S., Zhou F., Braden B. et al. An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy // Scientific reports. – 2020. – vol. 10(1). – pp. 1–15.
39. Ameling S., Wirth S., Paulus D., Lacey G., Vilarino F. Texture-Based Polyp Detection in Colonoscopy // Bildverarbeitung für die Medizin. – 2009. – pp. 346–350.
40. Andrews S., Tsochantaridis I., Hofmann T. Support vector machines for multiple-instance learning // Advances in Neural Information Processing Systems. – 2003. – pp. 577–584.
41. Badrinarayanan V., Kendall A., Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation // IEEE Transactions

- on Pattern Analysis and Machine Intelligence. – 2017. – vol. 39(12). – pp. 2481–2495.
42. Baldeon-Calisto M., Lai-Yuen S.K. AdaResU-Net: Multiobjective adaptive convolutional neural network for medical image segmentation // Neurocomputing. – 2020. – vol. 392. – pp. 325–340.
43. Belhassen H., Zhang H., Virginie F., Bourennane E.-B. Improving Video Object Detection by Seq-Bbox Matching // Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. – 2019. – pp. 226-233.
44. Belongie S., Malik J., Puzicha J. Shape matching and object recognition using shape contexts // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2002. – vol. 24(4). – pp. 509-522.
45. Bernal J. et al. Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results from the MICCAI 2015 Endoscopic Vision Challenge // IEEE Transactions on Medical Imaging. – 2017. – vol. 36(6). – pp. 1231–1249.
46. Bernal J., Sánchez F. J., Fernández-Esparrach G., Gil D., Rodríguez C., and Vilariño F. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians // Computerized Medical Imaging and Graphics. – 2015, – vol. 43. – pp. 99–111.
47. Bewley A., Ge Z., Ott L., Ramos F., Upcroft B. Simple online and realtime tracking // 2016 IEEE International Conference on Image Processing. – 2016. – pp. 3464-3468.
48. Bisschops R., Areia M., Coron E. Performance measures for upper gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative // Endoscopy, – 2016. – vol. 48, № 9. – pp. 843–864.
49. Bochkovskiy A., Wang C.-Y., Liao H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection // arXiv preprint arXiv:2004.10934, 2020.

50. Borgli H. et al. Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy // *Scientific Data*. – 2020. – vol. 7(1). – pp. 1–14.
51. Castaneda D., Popov V.B., Verheyen E., Wander P., Gross S.A. New technologies improve adenoma detection rate, adenoma miss rate, and polyp detection rate: a systematic review and meta-analysis // *Gastrointestinal endoscopy*. – 2018. – vol. 88(2). – pp. 209–222.
52. Chavdarova T., Baqu P., Bouquet S., Maksai A., Jose C., Bagautdinov T., Lettry L., Fua P., Van Gool L., Fleuret F. WILDTRACK: A Multi-camera HD Dataset for Dense Unscripted Pedestrian Detection // *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. – 2018. – pp. 5030-5039.
53. Cortes C., Vapnik V. Support-Vector Networks // *Machine Learning*. – 1995. – vol. 20(3). – pp. 273–297.
54. Dai J., He K., Sun J. Instance-aware Semantic Segmentation via Multi-task Network Cascades // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. – 2016. – pp. 3150–3158.
55. Dai J., Li Y., He K., Sun J. R-FCN: Object Detection via Region-Based Fully Convolutional Networks // *Advances in Neural Information Processing Systems*. – 2016. – pp. 379–387.
56. Dalal N., Triggs B. Histograms of oriented gradients for human detection // *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. – 2005. – vol. 1. – pp. 886–893.
57. De Lange T., Halvorsen P., Riegler M. Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy // *World journal of gastroenterology*. – 2018, – vol. 24(45). – p. 5057.
58. Dollar P., Wojek C., Schiele B., Perona P. Pedestrian detection: An evaluation of the state of the art // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2012. – vol. 34(4). – pp. 743–761.

59. Dollár P., Wojek C., Schiele B., Perona P. Pedestrian detection: A benchmark // IEEE Conference on Computer Vision and Pattern Recognition. – 2009. – pp. 304–311.
60. Everingham M., Van Gool L., Williams C., Winn J., Zisserman A. The PASCAL Visual Object Classes (VOC) Challenge // International Journal of Computer Vision. – 2010. – vol. 88(2). – pp. 303–338.
61. Feichtenhofer C., Pinz A., Zisserman A. Detect to Track and Track to Detect // IEEE International Conference on Computer Vision. – 2017. – pp. 3057–3065.
62. Felzenszwalb P., Girshick R., McAllester D. Cascade object detection with deformable part models // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – 2010. – pp. 2241–2248.
63. Felzenszwalb P., Girshick R., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2010. – vol. 32(9). – pp. 1627–1645.
64. Felzenszwalb P., McAllester D., Ramanan D. A discriminatively trained, multiscale, deformable part model // IEEE Conference on Computer Vision and Pattern Recognition. – 2008. – pp. 1–8.
65. Fernandez-Gonzalez R., Deschamps T., Idica A., Malladi R., Solorzano C.O. Automatic segmentation of histological structures in mammary gland tissue sections // Journal of biomedical optics. – 2004. – vol. 9(3). – pp. 444–454.
66. Freund Y., Schapire R., Abe N. A Short Introduction to Boosting // Journal- Japanese Society for Artificial Intelligence. – 1999. – vol. 14(5). – pp. 771–780.
67. Fully convolutional reduced VGGNet [Электронный ресурс]. Режим доступа: <https://gist.github.com/weiliu89/2ed6e13bfd5b57cf81d6> (Дата обращения: 25.03.2022)

68. Galeeva M.A. et al. On specific features of the endoscopic image processing // *Journal of Physics: Conference Series*. – 2019. – vol. 1236(1). – pp. 012036.
69. Girshick R. Fast R-CNN // *Proceedings of the IEEE international conference on computer vision*. – 2015. – pp. 1440–1448.
70. Girshick R. From Rigid Templates to Grammars: Object Detection with Structured Models: Ph.D. dissertation/Girshick R. – 2012.
71. Girshick R., Donahue J., Darrell T., Malik J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2016. – vol. 38(1). – pp. 142-158.
72. 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. – pp. 580– 587.
73. Girshick R., Felzenszwalb P., Mcallester D. Object detection with grammar models // *Advances in Neural Information Processing Systems*. – 2011. – pp. 442–450.
74. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Nets // *Advances in Neural Information Processing Systems*. – 2014. – pp. 2672– 2680.
75. Guo Y.B. and Matuszewski B. Giana polyp segmentation with fully convolutional dilation neural networks // *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. – 2019. – pp. 632–641.
76. Han W., Khorrami P., Paine T. L., Ramachandran P., Babaeizadeh M., Shi H., Li J., Yan S., Huang T.S. Seq-NMS for Video Object Detection // *CoRR*. – vol. abs/1602.08465. – 2016.
77. Hariharan B., Arbeláez P., Girshick R., Malik J. Hypercolumns for Object Segmentation and Fine-grained Localization // *Proceedings of the IEEE*

- Conference on Computer Vision and Pattern Recognition. – 2015, – pp. 447–456.
- 78.Hariharan B., Arbeláez P., Girshick R., Malik J. Simultaneous detection and segmentation // European Conference on Computer Vision. Springer. – 2014. – pp. 297–312.
- 79.He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN // 2017 IEEE International Conference on Computer Vision (ICCV). – 2017. – pp. 2980-2988.
- 80.He K., Zhang X., Ren S., Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition // European conference on computer vision. Springer. – 2014. – pp. 346–361.
- 81.Holme Ø., Bretthauer M., Fretheim A., Odgaard-Jensen J., Hoff G. Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals // Cochrane Database of Systematic Reviews. – 2013. – vol. 9.
- 82.Holzheimer R. G., Mannick J. A. Surgical treatment: Evidence-Based and Problem-Oriented. Munich: Zuckschwerdt. – 2001.
- 83.Howard A.G. Some Improvements on Deep Convolutional Neural Network Based Image Classification // arXiv preprint arXiv:1312.5402, 2013.
- 84.Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications // arXiv preprint arXiv:1704.04861, 2017.
- 85.Jha D. et al. Kvasir-Instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy // International Conference on Multimedia Modeling. – 2021. – pp. 218–229.
- 86.Jha D. et al. Kvasir-seg: A Segmented Polyp Dataset // Proceedings of International Conference on Multimedia Modeling. – 2020. – pp. 451–462.
- 87.Jha D. et al. ResUNet++: An Advanced Architecture for Medical Image Segmentation // Proceedings of IEEE International Symposium on Multimedia (ISM). – 2019, – pp. 225–2255.

88. Jha D., Ali S., Tomar N. K., Johansen H.D., Johansen D., Rittscher J., Riegler M.A., Halvorsen P. Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning // *Computer Vision and Pattern Recognition*. – 2021. – pp. 40496 – 40510.
89. Jha D., Riegler M.A., Johansen D., Halvorsen P., Johansen H.D., DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation // *IEEE 33rd International Symposium on Computer-Based Medical Systems*. – 2020. – pp. 558-564.
90. Jocher G. et al. ultralytics/yolov5: v6.1. [Электронный ресурс]. Режим доступа: <https://github.com/ultralytics/yolov5> (Дата обращения: 25.03.2022)
91. Kaminski M.F. et al. Quality indicators for colonoscopy and the risk of interval cancer // *New England Journal of Medicine*. – 2010. – vol. 362(19). – pp. 1795–1803.
92. Kang K., Li H., Yan J., Zeng X., Yang B., Xiao T., Zhang C., Wang Z., Wang R., Wang X. et al. T-CNN: Tubelets with Convolutional Neural Networks for Object Detection from Videos // *IEEE Transactions on Circuits and Systems for Video Technology*. – 2018. – vol. 28(10). – pp. 2896–2907.
93. Karkanis S.A., Iakovidis D.K., Maroulis D.E., Karras D.A., Tzivras M. Computer-aided tumor detection in endoscopic video using color wavelet features // *IEEE Transactions on Information Technology in Biomedicine*. – 2003. – vol. 7(3). – pp. 141–152.
94. Karpathy A., Fei-Fei L., Deep Visual-Semantic Alignments for Generating Image Descriptions // *Proceedings of the IEEE conference on computer vision and pattern recognition*. – 2015. – pp. 3128–3137.
95. Khryashchev V., Lebedev A., Stepanova O., Srednyakova A. Analysis of Pathologies on Endoscopic Images of the Stomach Using SSD and RetinaNet Neural Network Architecture // *Proceedings of 2021 IEEE East-West Design and Test Symposium (EWDTS 2021)*. – 2021. – pp. 119-123.



96. Khryashchev, V.V., Stepanova, O.A., Lebedev, A.A., Kashin, S.V., Kuvaev, R.O. Deep learning for gastric pathology detection in endoscopic images // ACM International Conference Proceeding Series, 3rd International Conference on Graphics and Signal Processing, ICGSP 2019. – 2019. – pp 90-94.
97. Ko C., Dominitz J. Complications of Colonoscopy: Magnitude and Management // Gastrointestinal Endoscopy Clinics of North America. – 2010. – vol. 20(4). – pp. 659-671.
98. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. – 2012. – pp. 1097–1105.
99. Kronborg O., Regula J. Population screening for colorectal cancer: advantages and drawbacks // Digestive Diseases. – 2007. – vol. 25(3). – pp. 270-273.
100. Lebedev A., Khryashchev V., Kazina E., Zhuravleva A., Kashin S., Zavyalov D. Automatic Identification of Appendiceal Orifice on Colonoscopy Images Using Deep Neural Network. In 2020 IEEE East-West Design & Test Symposium (EWDTS), Varna, Bulgaria. – 2020. – pp. 1-5.
101. Lebedev A., Khryashchev V., Stefanidi A., Stepanova O., Kashin S., Kuvaev R. Convolutional neural network for early detection of gastric cancer by endoscopic video analysis // Proc. SPIE 11433, Twelfth International Conference on Machine Vision (ICMV 2019). – 2020. – p. 1143325.
102. Lee J. Resection of diminutive and small colorectal polyps: what is the optimal technique? // Clinical endoscopy. – 2016. – vol. 49(4). – p. 355.
103. Lee J.Y., Jeong J., Song E.M. et al. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets // Scientific Reports. – 2020. – vol. 10(1). – pp. 1–9.

104. Levin B., Lieberman D., McFarland B., Andrews K., Brooks D., Bond J. et al. Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, A Joint Guideline From the American Cancer Society, the US MultiSociety Task Force on Colorectal Cancer, and the American College of Radiology // *Gastroenterology*. – 2008; – vol.134(5). – pp. 1570-1595.
105. Li Z., Peng C., Yu G., Zhang X., Deng Y., Sun J. Light-Head R-CNN: In Defense of Two-Stage Object Detector // *arXiv preprint arXiv:1711.07264*, 2017.
106. Lin J., Gan C., Han S. TSM: Temporal Shift Module for Efficient Video Understanding // *IEEE/CVF International Conference on Computer Vision*. – 2019. – pp. 7082-7092.
107. Lin T.-Y., Dollár P., Girshick R., He K., Hariharan B., and Belongie S. Feature Pyramid Networks for Object Detection // *IEEE Conference on Computer Vision and Pattern Recognition*. – 2017. – pp. 936-944.
108. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. Focal loss for dense object detection // *IEEE International Conference on Computer Vision*. – 2017. – pp. 2999-3007.
109. Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollár P., Zitnick C. Microsoft COCO: Common Objects in Context // *European Conference on Computer Vision*. Springer. – 2014. – pp. 740–755.
110. Liu S., Qi L., Qin H., Shi J., Jia J. Path Aggregation Network for Instance Segmentation // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. – 2018. – pp. 8759–8768.
111. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A. SSD: Single Shot MultiBox Detector // *European conference on computer vision*. Springer. – 2016. – pp. 21–37.
112. Lowe D. Distinctive Image Features from Scale-Invariant Keypoints // *International Journal of Computer Vision*. – 2004. – vol. 60(2). – pp. 91–110.

113. Lowe D. Object Recognition from Local Scale-Invariant Features // Proceedings of the Seventh IEEE International Conference on Computer Vision. – 1999. – vol. 2. – pp. 1150-1157.
114. Malisiewicz T., Gupta A., Efros A.A. Ensemble of exemplar-SVMs for object detection and beyond // International Conference on Computer Vision. – 2011. – pp. 89-96.
115. Matyja M., Pasternak A., Szura M., Wysocki M., Pędziwiatr M., Rembiasz K. How to improve the adenoma detection rate in colorectal cancer screening? Clinical factors and technological advancements // Archives of medical science: AMS. – 2019. – vol. 15(2). – p. 424.
116. Meng Y., Wei M., Gao D., Zhao Y., Yang X., Huang X., Zheng Y. CNN-GCN Aggregation Enabled Boundary Regression for Biomedical Image Segmentation // International Conference on Medical Image Computing and Computer-Assisted Intervention. – 2020. – pp. 352–362.
117. Milan A., Leal-Taixé L., Reid I.D., Roth S., Schindler K. MOT16: A Benchmark for Multi-Object Tracking // CoRR. – vol. abs/1603.00831. – 2016.
118. Minaee S., Boykov Y., Porikli F., Plaza A., Kehtarnavaz N., Terzopoulos D. Image segmentation using deep learning: A survey // arXiv preprint arXiv:2001.05566, 2020.
119. Mohan A., Papageorgiou C., Poggio T. Example-based object detection in images by components // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2001. vol. 23(4). – pp. 349-361.
120. NVIDIA TensorRT: Programmable Inference Accelerator. [Электронный ресурс]. Режим доступа: <https://developer.nvidia.com/tensorrt> (Дата обращения: 25.03.2022).
121. Papageorgiou C., Oren M., Poggio T. A general framework for object detection // IEEE International Conference on Computer Vision. – 1998. – pp. 555-562.

122. Papageorgiou C., Poggio T. A trainable system for object detection // *International Journal of Computer Vision*. – 2000. – vol. 38(1). – pp. 15–33.
123. Pogorelov K. et al. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection // *Proceedings of the 8th ACM on Multimedia Systems Conference*. – 2017. – pp. 164–169.
124. Pogorelov K., Randel K. R. et al. Nerthus: A bowel preparation quality video dataset // *Proceedings of the ACM on Multimedia Systems Conference (MMSys)*. – 2017. – pp. 170–174.
125. Ponugoti P.L., Cummings O.W., Rex D.K. Risk of cancer in small and diminutive colorectal polyps // *Digestive and Liver Disease*. – 2017. vol. 49(1). – pp. 34–37.
126. Rangayyan R.M. *Biomedical Image Analysis*. – CRC Press, Boca Raton, 2005. – p. 1306.
127. 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. – pp. 779–788.
128. Redmon J., Farhadi A. Yolo9000: better, faster, stronger // *IEEE Conference on Computer Vision and Pattern Recognition*. – 2017. – pp. 6517-6525.
129. Redmon J., Farhadi A. Yolov3: An incremental improvement // *arXiv preprint arXiv:1804.02767*, 2018.
130. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks // *Advances in neural information processing systems*. – 2015. – pp. 91–99.
131. 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*. – 2017. – vol. 39(6). – pp. 1137-1149.

132. Riegler M. Eir - A medical multimedia system for efficient computer aided diagnosis: Ph.D. dissertation/ Riegler M. – PhD thesis. University of Oslo. – 2017.
133. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge // *International Journal of Computer Vision*. – 2015. – vol. 115(3). – pp. 211–252.
134. Sabater A., Montesano L., Murillo A.C. Robust and efficient post-processing for video object detection // *IEEE/RSJ International Conference on Intelligent Robots and Systems*. – 2020, – pp. 10536-10542.
135. Saeedizadeh N., Minaee S., Kafieh R., Yazdani S., Sonka M. COVID TV-Unet: Segmenting COVID-19 chest CT images using connectivity imposed Unet // *arXiv preprint arXiv:2007.12303*, 2020.
136. Sande K. E., Uijlings J. R., Gevers T., Smeulders A. W. Segmentation as selective search for object recognition // *International Conference on Computer Vision*. – 2011. – pp. 1879-1886.
137. Sharma P., Bansal A., Mathur S., Wani S., Cherian R., McGregor D., et al. The utility of a novel narrow band imaging endoscopy system in patients with Barrett’s esophagus // *Gastrointest Endosc* 2006. – vol. 64(2). – pp. 167-175.
138. Shin H.-C. et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning // *IEEE Transactions on Medical Imaging*. – 2016. – vol. 35(5). – pp. 1285–1298.
139. Shin Y., Qadir H. A., Aabakken L., Bergsland J., Balasingham I. Automatic Colon Polyp Detection Using Region Based Deep CNN and Post Learning Approaches // *IEEE Access*. – 2018. – vol. 6. – pp. 40950–40962.
140. Shin Y., Qadir H.A., Balasingham I. Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance // *IEEE Access*. – 2018. – vol. 6. – pp. 56007–56017.

141. Silva J., Histace A., Romain O., Dray X., Granado B. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer // *International Journal of Computer Assisted Radiology and Surgery*. – 2014. – vol. 9(2). – pp. 283–293.
142. Smedsrud, P.H., Thambawita, V., Hicks, S.A. et al. Kvasir-Capsule, a video capsule endoscopy dataset // *Sci Data*. – 2021. – vol. 8. – p. 142.
143. Tajbakhsh N. et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? // *IEEE transactions on medical imaging*. – 2016, – vol. 35(5). – pp. 1299–1312.
144. Tan M., Le Q.V. Efficientnet: Rethinking model scaling for convolutional neural networks // *arXiv preprint arXiv:1905.11946*, 2019.
145. Tan M., Pang R., Le Q.V. Efficientdet: Scalable and efficient object detection // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. – 2020. – pp. 10 781–10 790.
146. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. – 2001. – pp. I–I.
147. Viola P., Jones M. Robust real-time face detection // *International Journal of Computer Vision*. – 2004. – vol. 57(2). – pp. 137–154.
148. Wang C.-Y, Liao H.-Y.M., Wu Y.-H., Chen P.-Y., Hsieh J.-W., Yeh I.-H. CSPNet: A New Backbone that can Enhance Learning Capability of CNN // *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. – 2020. – pp. 1571-1580.
149. Wang P., Xiao X., Glissen Brown J.R. et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy // *Nature biomedical engineering*. – 2018. – vol. 2(10), – pp. 741–748.
150. Wang R., Chen S., Ji C., Fan J., Li Y. Boundary-aware context neural network for medical image segmentation // *arXiv preprint arXiv:2005.00966*, 2020.

151. Wang Y., Tavanapong W., Wong J., Oh J. H., De Groen P. C. Polyp-alert: Near real-time feedback during colonoscopy // *Computer Methods and Programs in Biomedicine*. – 2015. – vol. 120(3). – pp. 164–179.
152. Wojke N., Bewley A., Paulus D. Simple online and realtime tracking with a deep association metric // *IEEE International Conference on Image Processing*. – 2017. – pp. 3645-3649.
153. Wu H., Chen Y., Wang N., Zhang Z. Sequence Level Semantics Aggregation for Video Object Detection // *IEEE/CVF International Conference on Computer Vision*. – 2019. – pp. 9216-9224.
154. Wu Q., Shen C., Wang P., Dick A., Hengel A. Image Captioning and Visual Question Answering Based on Attributes and External Knowledge // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2018. – vol. 40(6), – pp. 1367–1381.
155. Xu K., Ba J., Kiros R., Cho K., Courville A., Salakhudinov R., Zemel R., Bengio Y. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention // *International Conference on Machine Learning*. – 2015, – pp. 2048–2057.
156. Yamada M., Saito Y., Imaoka H. et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy // *Scientific reports*. – 2019, – vol. 9(1), – pp. 1–9.
157. Zeiler M.D., Fergus R. Visualizing and Understanding Convolutional Networks // *European Conference on Computer Vision*. Springer. – 2014. – pp. 818–833.
158. Zhang Z., Cheng D., Zhu X., Lin S., Dai J. Integrated Object Detection and Tracking with Tracklet-Conditioned Detection // *CoRR*. – vol. abs/1811.11167. – 2018.
159. Zhu X., Wang Y., Dai J., Yuan L., Wei Y. Flow-Guided Feature Aggregation for Video Object Detection // *IEEE International Conference on Computer Vision*. – 2017. – pp. 408-417.

## ПРИЛОЖЕНИЕ А. Технические характеристики видеопроцессоров эндоскопических систем

Таблица 1 – Технические характеристики видеопроцессора CV-180

Выход HDTV-сигнала	Может быть выбран RGB-выход или YPbPr-выход.
Выход SDTV-сигнала	Возможен одновременный выход сигналов VBS-составной (PAL), Y/C и RGB.
Настройки баланса белого	Настройка баланса белого при использовании кнопки Wh/B на передней панели прибора.
Контрольная цветная полоса	Возможная следующая регулировка цветового тона: Red, Blue, Chroma, $\pm 8$ ступеней.
Автоматический контроль усиления (AGC)	Возможно электронное усиление изображения при недостаточном освещении, если дистальный конец эндоскопа расположен слишком далеко от объекта.
Контрастность	<p>Можно установить один из следующих трёх режимов контрастности (N, H, L):</p> <ul style="list-style-type: none"> <li>- N (нормальная): нормальное изображение;</li> <li>- H (высокая): Тёмные зоны на изображении – темнее, светлые – светлее, чем на нормальном изображении;</li> <li>- L (низкая): Тёмные зоны на изображении – светлее, светлые зоны – темнее, чем на нормальном изображении.</li> </ul>
Диафрагма	<p>Могут быть выбраны режимы автоматической регулировки диафрагмы:</p> <p>Reak: используется при наблюдении посредством фокусировки на малом освещенном участке;</p> <p>Auto: используется при наблюдении посредством фокусировки на центре изображения.</p>



Средство для записи	xD-карта (1 ГБ /512/256/128/64/32/16 Мб) фирмы Olympus. MAPC-10 может использоваться как адаптер PC-карты.
Формат записи	TIFF: без сжатия, SHQ: приблизительно 1/5, HQ: приблизительно 1/7, SQ: приблизительно 1/10.
Количество записываемых изображений	На 16 Мб, SDTV/HDTV, TIFF: приблизительно 30/6 изображений, SHQ: приблизительно 310/110 изображений; HQ: приблизительно 2000/760 изображений, SQ: приблизительно 2570/430 изображений.

Таблица 2 – Технические характеристики видеопроцессора CV-190

Выход аналогового сигнала HDTV	Можно выбрать выход сигнала в формате RGB или YPbPr (1080/50I: PAL).
Выход аналогового сигнала SDTV	Композитный разъем VBS, Y/C и RGB (576/50I: PAL).
Выход цифрового сигнала	Можно выбрать HD-SDI (SMPTE 292M), SD-SDI (SMPTE 259M), DV (IEEE 1394) и DVI (WUXGA, 1080p или SXGA).
Регулировка баланса белого	Регулировка баланса белого осуществляется с помощью кнопки баланса белого на передней панели.
Вывод стандартной таблицы цветов	Можно отобразить экраны Colour bar (Цветовая полоса) или 50% white (50% белого).

Регулировка оттенка цвета	Возможны следующие варианты регулировки оттенка цвета: с помощью кнопки регулировки уровня оттенка цвета и кнопки выбора оттенка цвета на клавиатуре. Регулировка красного: $\pm 8$ шагов; регулировка синего: $\pm 8$ шагов; регулировка насыщенности цвета: $\pm 8$ шагов.
Автоматическая регулировка усиления (AGC)	Возможно электрическое усиление изображения при недостаточности освещения из-за нахождения дистального конца эндоскопа на большом расстоянии от объекта.
Контрастность	<p>N (обычный): обычное изображение.</p> <p>H (Высокий): темные участки темнее, а светлые светлее по сравнению с обычным изображением.</p> <p>L (Низкий): темные участки светлее, а светлые темнее по сравнению с обычным изображением.</p>
Диафрагма	<p>Можно выбрать различные автоматические режимы диафрагмы, используя переключатель Iris mode (Режим диафрагмы) на передней панели:</p> <p>auto (автоматическое значение): яркость регулируется по самому яркому участку центральной части, а средняя яркость – по периферической части;</p> <p>peak (пиковое значение): яркость регулируется по самому яркому участку эндоскопического изображения;</p> <p>average (среднее значение): яркость регулируется по средней яркости эндоскопического изображения.</p>
Средство для записи	MAJ-1925 (OLYMPUS)

Формат записи	TIFF: без сжатия; JPEG (1/5): сжатие примерно на 1/5; JPEG (1/10): сжатие примерно на 1/10.
Количество записанных изображений	TIFF: примерно 227 изображений; JPEG (1/5): примерно 1024 изображений; JPEG (1/10): примерно 2048 изображений.

**ПРИЛОЖЕНИЕ Б. Свидетельства о государственной  
регистрации интеллектуальной собственности**

РОССИЙСКАЯ ФЕДЕРАЦИЯ



**СВИДЕТЕЛЬСТВО**  
о государственной регистрации программы для ЭВМ  
**№ 2020618382**

**AI.ColonoscopySolution 1.0 - программа для анализа  
колоноскопических изображений на основе технологий  
машинного обучения**

Правообладатель: *Общество с ограниченной ответственностью  
«Смарт Анализ» (RU)*

Авторы: *Лебедев Антон Александрович (RU), Хрящев Владимир  
Вячеславович (RU), Завьялов Дмитрий Вячеславович (RU), Казина  
Евгения Максимовна (RU), Журавлева Анастасия Сергеевна (RU)*

Заявка № **2020617175**  
Дата поступления **07 июля 2020 г.**  
Дата государственной регистрации  
в Реестре программ для ЭВМ **24 июля 2020 г.**



*Руководитель Федеральной службы  
по интеллектуальной собственности*

*Г.П. Нилов*

РОССИЙСКАЯ ФЕДЕРАЦИЯ



## СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2022610262

**Endoscopy\_Polyps 1.0 – программа для детектирования и распознавания полипов в кишечнике с использованием эндоскопических изображений**

Правообладатели: *Андержанова Анастасия Сергеевна (RU), Казина Евгения Максимовна (RU), Лебедев Антон Александрович (RU), Среднякова Анастасия Сергеевна (RU)*

Авторы: *Андержанова Анастасия Сергеевна (RU), Казина Евгения Максимовна (RU), Лебедев Антон Александрович (RU), Среднякова Анастасия Сергеевна (RU)*

Заявка № **2021668310**

Дата поступления **15 ноября 2021 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **11 января 2022 г.**



*Руководитель Федеральной службы  
по интеллектуальной собственности*

*Г.П. Ильин*



РОССИЙСКАЯ ФЕДЕРАЦИЯ

**СВИДЕТЕЛЬСТВО**

о государственной регистрации базы данных

**№ 2022620355****Colonoscopy.DB - база данных эндоскопических  
изображений толстой кишки**

Правообладатели: *Гусейнова Сабина Руслановна (RU), Завьялов  
Дмитрий Вячеславович (RU), Лебедев Антон  
Александрович (RU)*

Авторы: *Гусейнова Сабина Руслановна (RU), Завьялов  
Дмитрий Вячеславович (RU), Лебедев Антон  
Александрович (RU)*

Заявка № **2021622960**Дата поступления **07 декабря 2021 г.**

Дата государственной регистрации

в Реестре баз данных **28 февраля 2022 г.**

*Руководитель Федеральной службы  
по интеллектуальной собственности*

*Ю.С. Зубов*

## ПРИЛОЖЕНИЕ В. Акты внедрения

Государственное бюджетное  
учреждение здравоохранения  
Ярославской области  
«Областная клиническая  
онкологическая больница»  
ИНН 7606028790 КПП 760601001  
150054, Ярославль, проспект  
Октября, 67

«УТВЕРЖДАЮ»  
Главный врач ГБУЗ ЯО ОКОБ  
П.В. Нестеров  
« 12 » мая 2022 г.

### АКТ

О внедрении результатов диссертационной работы Лебедева Антона Александровича на тему «Исследование нейросетевых алгоритмов обнаружения объектов на видеоизображениях в медицинских системах прикладного телевидения»

Комиссия в составе: председатель – заведующий отделением эндоскопии Кашин С.В., члены комиссии – ассистент кафедры гастроэнтерологии ФДПО РНИМУ им. Н.И. Пирогова, врач-эндоскопист, врач-гастроэнтеролог Куваев Р.О., доцент кафедры онкологии ЯГМУ, врач-эндоскопист Завьялов Д.В., рассмотрев диссертацию Лебедева А.А. составила настоящий акт о том, что ее результаты нашли применение в деятельности эндоскопического отделения ГБУЗ ЯО Областная клиническая онкологическая больница. Особый практический интерес представляют следующие результаты диссертации:

- разработанный алгоритм детектирования полипов на эндоскопических видеоданных слизистой оболочки толстой кишки на основе сверточной нейронной сети. При его тестировании на видеофрагментах колоноскопических исследований, содержащих 11 192 изображений (7057 содержали полипы), средняя точность алгоритма составила 83,3%, что свидетельствует о высокой степени совпадения экспертной врачебной разметки с результатами, предсказанными алгоритмом.

Данный алгоритм использован при создании системы поддержки принятия решений в колоноскопии, а также для обучения начинающих специалистов.

Заведующий отделением эндоскопии ЯОКОБ, руководитель учебного эндоскопического центра ЯГМУ, главный внештатный специалист по эндоскопии Департамента здравоохранения и фармации Ярославской области (председатель комиссии)

к.м.н., Кашин С.В.



Члены комиссии:  
к.м.н., Куваев Р.О.  
д.м.н., Завьялов Д.В.







**Общество с ограниченной  
ответственностью «Точка зрения»**

150018, Ярославская обл.,  
г. Ярославль, 3-й Норский пер.,  
д.7, корп.2, оф.43  
ИНН 7604288987, КПП 760401001

Исх 07-04/2022 от 27.04.2022

«УТВЕРЖДАЮ»

Генеральный директор  
ООО «Точка зрения»



к.т.н. Ганин А.Н.

«27» апреля 2022 г.

## АКТ

внедрения результатов диссертационной работы Лебедева Антона Александровича, выполненной в Ярославском государственном университете имени П.Г. Демидова (ЯрГУ), на тему «Исследование нейросетевых алгоритмов обнаружения объектов на видеоизображениях в медицинских системах прикладного телевидения».

Комиссия в составе: председатель комиссии – технический директор Назаровский А.Е., члены комиссии – руководитель проектов Волков Д.Б., ведущий программист Карелин П.В., рассмотрев диссертационную работу Лебедева А.А. составила настоящий акт о том, что ее результаты нашли применение в работе ООО «Точка зрения» и были протестированы в рамках Соглашения о сотрудничестве компании «Точка зрения» от 13.09.2021 № АП-П15 с ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России.

Разработанный Лебедевым А.А. алгоритм обнаружения патологий на основе сверточных нейронных сетей для систем прикладного телевидения в эндоскопии, продемонстрировал высокую чувствительность и точность определения колоректальных полипов в ходе колоноскопических исследований, проводимых для пациентов ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России.

Данный алгоритм использован при разработке программно-аппаратного комплекса анализа видеопотока – Точка.EndoPlus. Использование предложенного А.А. Лебедевым алгоритма позволило повысить точность обнаружения полипов эндоскопических видеоданных.

**Приложение.** Отзыв о сотрудничестве ООО «Точка зрения» с ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России

Председатель комиссии

Члены комиссии

к.т.н., Назаровский А.Е.

Волков Д.Б.

Карелин П.В.



**Приложение. Отзыв о сотрудничестве** ООО «Точка зрения» с ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное учреждение

**Национальный  
медико-хирургический Центр  
имени Н.И. Пирогова**

(ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России)

105203, г. Москва, ул. Пискаревская, дом 70  
тел. +7 (499) 464-49-54, +7 (495) 465-33-55,  
факс +7 (495) 465-09-52, e-mail: [info@nimgov-center.ru](mailto:info@nimgov-center.ru)  
ОКПО 55246260 ОГРН 1027700077668  
ИНН/КПП 7733108569/771901001

КОМУ:

Генеральному директору  
ООО «Точка зрения»  
Гашину А.Н.

Адрес:

150018, Ярославская обл., г. Ярославль,  
3-й Норский пер., д.7, корп.2, кв.43

27.04.2022 № БН  
Отзыв о сотрудничестве

ФГБУ «НМХЦ им. Н.И. Пирогова» Минздрава России (далее - Пироговский Центр) и ООО «Точка зрения» в рамках Соглашения о сотрудничестве от 13.09.2021 № АП-П5 работают над созданием решений и сервисов цифровой медицины в области анализа видеопотока и контроля качества колоноскопического исследования на основе методов искусственного интеллекта. В настоящий момент текущая версия программного обеспечения ООО «Точка зрения» по детектированию колоректальных полипов на видеоданных в пилотном режиме проходит экспертизу и доработку в отделении диагностической и оперативной эндоскопии Пироговского Центра при проведении колоноскопических исследований, демонстрируя потенциальные функциональные возможности применения в практическом здравоохранении.

Разработанный сотрудниками ООО «Точка зрения» алгоритм детектирования патологий на основе сверточных нейронных сетей для систем прикладного телевидения в эндоскопии продемонстрировал высокую чувствительность и точность определения колоректальных полипов при колоноскопических исследованиях.

Результаты проекта после его доработки на основе рекомендаций врачей и получения регистрационного удостоверения могут быть востребованы в широком спектре медицинских организаций, а также интегрированы в другие перспективные отечественные ИТ-разработки в области здравоохранения.

Заместитель генерального директора по  
научной и образовательной деятельности



А.А. Пулин

«УТВЕРЖДАЮ»

Ректор Ярославского

государственного университета

им. П.П. Демидова

А.И. Русаков

« 20 » июля 2022 г.



АКТ

внедрения результатов диссертационной работы Лебедева Антона Александровича на тему «Исследование нейросетевых алгоритмов обнаружения объектов на видеоизображениях в медицинских системах прикладного телевидения» в научно-исследовательские работы

Результаты диссертационной работы А.А. Лебедева, представленной на соискание ученой степени кандидата технических наук по специальности 2.2.13. Радиотехника, в том числе системы и устройства телевидения, использованы в научно-исследовательской работе:

– при выполнении НИР «Исследование применимости технологий глубокого обучения и сверточных нейронных сетей для распознавания патологий желудка по видеоизображениям эндоскопического исследования» (грант РФФИ 19-37-90153) внедрена методика полуавтоматической подготовки базы эндоскопических видеоданных с разметкой областей интереса для обучения нейросетевых моделей.

Заведующий кафедрой цифровых технологий и машинного обучения,  
доцент, к.ф.-м.н.

М.В. Чистяков

«УТВЕРЖДАЮ»  
 Ректор Ярославского  
 государственного университета  
 им. П.А. Демидова  
 А.И. Русаков  
 «20» \_\_\_\_\_ 2022 г.



### АКТ

внедрения результатов диссертационной работы Лебедева Антона Александровича на тему «Исследование нейросетевых алгоритмов обнаружения объектов на видеоназображениях в медицинских системах прикладного телевидения» в учебный процесс

Мы, нижеподписавшиеся, заведующий кафедрой цифровых технологий и машинного обучения, доцент, к.ф.-м.н. М.В. Чистяков и профессор кафедры цифровых технологий и машинного обучения, доцент, д.т.н. А.Л. Приоров составили настоящий акт о том, что результаты диссертационной работы А.А. Лебедева внедрены в учебный процесс на кафедре цифровых технологий и машинного обучения физического факультета ЯрГУ (направление «Радиотехника»):

- в курсе «Системы технического зрения» – нейросетевой алгоритм обнаружения объектов на видеоназображениях с усовершенствованными процедурами предобработки и постобработки;
- в курсе «Цифровая обработка изображений» – модифицированный алгоритм аугментации данных, учитывающий специфику видеоназображений в медицинских системах.

Заведующий кафедрой цифровых технологий и машинного обучения, к.ф.-м.н., доцент

 М.В. Чистяков

Профессор кафедры цифровых технологий и машинного обучения, д.т.н., доцент

 А.Л. Приоров