

Школа – Инженерная школа информационных технологий и робототехники  
 Направление подготовки – 15.04.06 Мехатроника и робототехника  
 Отделение школы (НОЦ) – Отделение автоматизации и робототехники

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы			
<b>Распознавание жестов рук с помощью нейронной сети с архитектурой Mask-RCNN</b>			
УДК 004.932.72'1:004.7.032.26			
Студент			
Группа	ФИО	Подпись	Дата
8ЕМ91	Булыгин Дмитрий Александрович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Мамонова Т.Е.	к.т.н., доцент		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская М.В.	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ООД ШБИП	Федоренко О.Ю.	д.м.н.		

По разделу на иностранном языке

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель ОИЯ ШБИП	Пичугова И.Л.	–		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОАР ИШИТР	Малышенко А.М.	д.т.н., профессор		

## ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП

Код компетенции	Наименование компетенции
<b>Универсальные компетенции</b>	
УК(У)-1	способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
<b>Общепрофессиональные компетенции</b>	
ОПК(У)-1	способен представлять адекватную современному уровню знаний научную картину мира на основе знания основных положений, законов и методов естественных наук и математики
ОПК(У)-2	владеет в полной мере основным физико-математическим аппаратом, необходимым для описания и исследования разрабатываемых систем и устройств
ОПК(У)-3	владеет современными информационными технологиями, готовностью применять современные и специализированные средства автоматизированного проектирования и машинной графики при проектировании систем и их отдельных модулей, знать и соблюдать основные требования информационной безопасности
ОПК(У)-4	готов собирать, обрабатывать, анализировать и систематизировать научно-техническую информацию по тематике исследования, использовать достижения отечественной и зарубежной науки, техники и технологии в своей профессиональной деятельности
ОПК(У)-5	способен использовать методы современной экономической теории при оценке эффективности разрабатываемых и исследуемых систем и устройств, а также результатов своей профессиональной деятельности
ОПК(У)-6	готов использовать на практике приобретенные умения и навыки в организации исследовательских и проектных работ, выполняемых малыми группами исполнителей
<b>Профессиональные компетенции</b>	
ПК(У)-1	способен составлять математические модели мехатронных и робототехнических систем, их подсистем, включая исполнительные, информационно-сенсорные и управляющие модули, с применением методов формальной логики, методов конечных автоматов, сетей Петри, методов искусственного интеллекта, нечеткой логики, генетических алгоритмов, искусственных нейронных и нейро-нечетких сетей
ПК(У)-2	способен использовать имеющиеся программные пакеты и, при необходимости, разрабатывать новое программное обеспечение,

	необходимое для обработки информации и управления в мехатронных и робототехнических системах, а также для их проектирования
ПК(У)-3	способен разрабатывать экспериментальные макеты управляющих, информационных и исполнительных модулей мехатронных и робототехнических систем и проводить их исследование с применением современных информационных технологий
ПК(У)-4	способен осуществлять анализ научно-технической информации, обобщать отечественный и зарубежный опыт в области мехатроники и робототехники, средств автоматизации и управления, проводить патентный поиск
ПК(У)-5	способен разрабатывать методики проведения экспериментов и проводить эксперименты на действующих макетах и образцах мехатронных и робототехнических систем и их подсистем, обрабатывать результаты с применением современных информационных технологий и технических средств
ПК(У)-6	готов к составлению аналитических обзоров и научно-технических отчетов по результатам выполненной работы, в подготовке публикаций по результатам исследований и разработок
ПК(У)-7	способен внедрять на практике результаты исследований и разработок, выполненных индивидуально и в составе группы исполнителей, обеспечивать защиту прав на объекты интеллектуальной собственности
ПК(У)-8	готов к руководству и участию в подготовке технико-экономического обоснования проектов создания мехатронных и робототехнических систем, их подсистем и отдельных модулей
ПК(У)-9	способен к подготовке технического задания на проектирование мехатронных и робототехнических систем их подсистем и отдельных устройств с использованием стандартных исполнительных и управляющих устройств, средств автоматики, измерительной и вычислительной техники, а также новых устройств и подсистем
ПК(У)-10	способен участвовать в разработке конструкторской и проектной документации мехатронных и робототехнических систем в соответствии с имеющимися стандартами и техническими условиями
ПК(У)-11	готов разрабатывать методику проведения экспериментальных исследований и испытаний мехатронной или робототехнической системы, способностью участвовать в проведении таких испытаний и обработке их результатов
<b>Дополнительно сформированная университетом профессиональная компетенция</b>	
ДПК(У)-12	готов к решению задач научно-педагогической деятельности в области профессионального образования, планированию и проведению учебных занятий и разработке учебно-методических пособий и указаний

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа – Инженерная школа информационных технологий и робототехники  
 Направление подготовки (специальность) – 15.04.06 Мехатроника и робототехника  
 Отделение школы (НОЦ) – Отделение автоматизации и робототехники

УТВЕРЖДАЮ:  
 Руководитель ООП  
 \_\_\_\_\_ Малышенко А.М.  
 (Подпись)    (Дата)    (Ф.И.О.)

**ЗАДАНИЕ**  
**на выполнение выпускной квалификационной работы**

В форме:

Магистерской диссертации
--------------------------

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8ЕМ91	Булыгину Дмитрию Александровичу

Тема работы:

Распознавание жестов рук с помощью нейронной сети с архитектурой Mask-RCNN	
Утверждена приказом директора (дата, номер)	№ 50-12/с от 19.02.2021

Срок сдачи студентом выполненной работы:	16.06.2021
--	------------

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<p><b>Исходные данные к работе</b></p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<p>Объектом исследования является нейронная сеть Mask-RCNN. Необходимо ее обучить и оценить способность распознавать жесты и обозначать их в кадре. Обученная нейронная сеть должна распознавать жесты различных классов при помощи камеры.</p>
---	---

<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b> <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> <li>1. Провести аналитический обзор существующих решений;</li> <li>2. Подготовить обучающий набор данных (изображения и маски);</li> <li>3. Сконфигурировать параметры нейронной сети;</li> <li>4. Обучить нейронную сеть Mask-RCNN;</li> <li>5. Провести тестирование нейронной сети.</li> </ol>
<p><b>Перечень графического материала</b> <i>(с точным указанием обязательных чертежей)</i></p>	<ol style="list-style-type: none"> <li>1. Блок-схема создания обучающих изображений и масок.</li> </ol>

**Консультанты по разделам выпускной квалификационной работы**

*(с указанием разделов)*

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Верховская М.В., доцент ОСГН ШБИП, к.э.н.
Социальная ответственность	Федоренко О.Ю., профессор ООД ШБИП, д.м.н.
Раздел на иностранном языке	Пичугова И.Л., старший преподаватель ОИЯ ШБИП

**Названия разделов, которые должны быть написаны на русском и иностранном языках:**

History of inventions for gesture recognition;  
 Devices being developed in our time;  
 Analysis of the methods for solving the gesture recognition problem;  
 Application of the Inventive Problem-Solving Algorithm

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	28.02.2021
---	------------

**Задание выдал руководитель / консультант (при наличии):**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Мамонова Т.Е.	к.т.н., доцент		28.02.2021

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ЕМ91	Булыгин Д.А.		28.02.2021

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа – Инженерная школа информационных технологий и робототехники  
 Направление подготовки (специальность) – 15.04.06 Мехатроника и робототехника  
 Уровень образования – Магистратура  
 Отделение школы (НОЦ) – Отделение автоматизации и робототехники  
 Период выполнения \_\_\_\_\_ (осенний / весенний семестр 2020 /2021 учебного года)

Форма представления работы:

Магистерская диссертация

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН**  
**выполнения выпускной квалификационной работы**

Срок сдачи студентом выполненной работы:	16.06.2021
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
13.05.2021	Основная часть	65
15.05.2021	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
17.05.2021	Социальная ответственность	10
18.05.2021	Приложение на иностранном языке	10

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Мамонова Т.Е.	к.т.н., доцент		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОАР ИШИТР	Малышенко А.М.	д.т.н., профессор		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ,  
РЕСУРСООБЪЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

<b>Группа</b>	<b>ФИО</b>
8ЕМ91	Булыгину Дмитрию Александровичу

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение</b>	<b>ОАР</b>
<b>Уровень образования</b>	Магистратура	<b>Направление</b>	15.04.06 Мехатроника и робототехника

**Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:**

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	- Оклад инженера – 22695,68 руб. в месяц; - Оклад руководителя проекта – 35111,5 руб. в месяц. - Человеческие ресурсы – 2 человека (руководитель и студент-дипломник).
2. Нормы и нормативы расходования ресурсов	- Тариф на электроэнергию – 6,59 руб./кВт·ч.; - Годовая норма амортизации составляет 40 %
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	На основании пункта 1 ст.58 закона №212-ФЗ для учреждений осуществляющих образовательную и научную деятельность вводится пониженная ставка – 27,1%

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения	- Методы коммерциализации результатов инженерных решений; - Морфологический анализ проекта.
2. Планирование и формирование бюджета научных исследований	- Определение трудоемкости выполнения работ; - Расчет материальных затрат НИИ; - Основная и дополнительная зарплата исполнителей темы; - Отчисления во внебюджетные фонды; - Накладные расходы; - Проведение анализа безубыточности проекта
3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	- - Расчет интегрального показателя финансовой эффективности.

**Перечень графического материала (с точным указанием обязательных чертежей):**

1. Оценка конкурентоспособности технических решений	
2. Альтернативы проведения НИ	
3. График проведения и бюджет НИ	
4. Оценка ресурсной, финансовой и экономической эффективности НИ	
<b>Дата выдачи задания для раздела по линейному графику</b>	01.03.2021

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская М.В.	К.Э.Н.		01.03.2021

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ЕМ91	Булыгин Дмитрий Александрович		01.03.2021

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»**

Студенту:

<b>Группа</b>	<b>ФИО</b>
8ЕМ91	Булыгину Дмитрию Александровичу

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение</b>	<b>ОАР</b>
<b>Уровень образования</b>	Магистратура	<b>Направление</b>	15.04.06 Мехатроника и робототехника

Тема ВКР:

<b>Распознавание жестов рук с помощью нейронной сети с архитектурой Mask-RCNN</b>	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	<p>Объектом исследования является алгоритм распознавания жестов рук на основе нейронной сети Mask-RCNN. Оборудованием, на котором производится работа, является компьютер с возможностью выхода в интернет, с подключённой к нему веб-камерой. Разработанная программа может использоваться для бесконтактного управления другими устройствами и механизмами.</p> <p>Рабочим местом является учебная аудитория №103 корпуса №10. Площадь помещения составляет 44,9 м<sup>2</sup>. Аудитория оборудована отоплением, вентиляцией, соответствует нормам освещенности и оборудована одноместными столами и персональными электронными вычислительными машинами.</p>
<b>Перечень вопросов, подлежащих исследованию, проектированию и разработке:</b>	
1. Правовые и организационные вопросы обеспечения безопасности: специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; организационные мероприятия при компоновке рабочей зоны.	<p>Трудовой кодекс Российской Федерации ГОСТ 12.2.032-78. Рабочее место при выполнении работ сидя ГОСТ 12.0.003-2015 ССБТ СанПиН 2.2.4.548-96 СанПиН 1.2.3685-21 СН 2.2.4/2.1.8.562 – 96 СНиП 23-05-95*. Естественное и искусственное освещение СП 52.13330.2016 ГОСТ 12.1.012-2004 ГОСТ 12.1.038-81 ССБТ ГОСТ 12.1.038-82 ССБТ ГОСТ 12.1.004-91 ССБТ СП 51.13330.2011 ГОСТ 17.4.3.04-85 ССОП</p>

<p><b>2. Производственная безопасность:</b>  2.1. Анализ выявленных вредных и опасных факторов  2.2. Обоснование мероприятий по снижению воздействия</p>	<p>Вредные факторы:  1. Отклонение показателей микроклимата от нормы;  2. Превышение уровня шума;  3. Отсутствие или недостаток освещения;  4. Вибрации;  5. Психофизиологические факторы (монотонность труда, нервно-психические перегрузки, перенапряжение зрительных анализаторов);  6. Электромагнитные излучения.</p> <p>Опасные факторы:  1. Поражение электрическим током;  2. Короткое замыкание;  3. Статическое электричество.</p>
<p><b>3. Экологическая безопасность:</b></p>	<p>Воздействие объекта на атмосферу, гидросферу отсутствует.  Воздействие на литосферу происходит при утилизации:  - компьютера и периферийных устройств (принтеры, МФУ, веб-камеры, наушники, колонки, телефоны);  - люминесцентных ламп;  - макулатуры.</p>
<p><b>4. Безопасность в чрезвычайных ситуациях:</b></p>	<p>Возможными чрезвычайными ситуациями при разработке устройства являются пожары, грозы, ураганы, оползни.  Наиболее типичная ЧС: пожар.</p>

Дата выдачи задания для раздела по линейному графику	01.03.2021
--	------------

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ООД ШБИП	Федоренко О.Ю.	д.м.н.		01.03.2021

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ЕМ91	Булугин Дмитрий Александрович		01.03.2021

## Реферат

Работа состоит из: 121 страницы, 40 рисунков, 34 таблиц, 36 источников.

Ключевые слова: свёрточные нейронные сети, распознавание жестов, нейронная сеть Mask-RCNN, сегментация, обучение с учителем, техническое зрение, Python, TensorFlow.

Настоящая выпускная квалификационная работа магистра содержит результаты анализа существующих решений по распознаванию жестов, конфигурацию и обучение нейронной сети Mask-RCNN, а также исследования по выяснению работоспособности нейронной сети. Для обучения этой сети был собран набор данных, состоящий из 300 синтезированных изображений, разделённых на 5 классов.

Также произведён расчёт стоимости конфигурации, обучения и оценки работы нейронной сети Mask-RCNN для задачи распознавания жестов, произведена оценка конкурентоспособности с другими методами, распознающими жесты. Произведена оценка безопасности при обучении и использовании данной нейронной сети, и ее работы для человека и окружающей среды.

## Оглавление

Введение	14
1 Обзор литературы	16
1.1 История изобретений, при помощи которых решалась данная задача	16
1.2 Компания, которая занимается подобными решениями в Томске: Rubius	19
1.3 Устройства, разрабатываемые в наше время	20
1.3.1 Распознавание жестов при помощи инфракрасных сенсоров на основе Intel RealSense	20
1.3.2 Динамическое распознавание жестов по направленным импульсным нейронным сетям	21
1.3.3 Распознавание жестов в реальном времени на основе сети перекалибровки функций с многомасштабной информацией	22
1.3.4 Обработка и управление при помощи биомедицинских сигналов	23
1.3.5 Система управления курсором мыши на основе жестов рук человека	24
1.4 Анализ приведенных методов решения задачи распознавания жестов	25
2 Исследование и анализ нейронной сети Mask-RCNN	27
2.1 Применение алгоритма решения изобретательской задачи АРИЗ-61	27
2.2 Задачи компьютерного зрения	29
2.3 Описание архитектуры Mask-RCNN	30
2.3.1 Основа модели	30
2.3.2 Область интересов (RoI)	31
2.3.3 Region Proposal Network	32
2.3.4 Наложение маски	34
2.3.5 Feature Pyramid Networks	35
3 Подготовка нейронной сети Mask-RCNN для распознавания жестов рук	37
3.1 Подготовка набора данных	37
3.2 Конфигурация Mask-RCNN	42

3.2.1	Выбор способа размещения обучающих и тестовых данных	42
3.2.2	Выбор backbone модели	44
3.2.3	Выбор параметра Learning Rate.	45
3.2.4	Выбор количества GPU	46
3.2.5	Выбор размера изображений	47
3.3	Итоги конфигурации Mask-RCNN	48
4	Оценка работы обученной нейронной сети на тестовой выборке	50
4.1	Зависимость точности распознавания жеста от расстояния между камерой и самим жестом	53
4.2	Оценка распознавания нейронной сетью каждого жеста	57
4.3	Оценка способности распознавания жестов, скрытых за преградой	59
5	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	61
5.1	Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения.	61
5.1.1	Потенциальные потребители результатов исследования	61
5.1.2	Анализ конкурентных технических решений	61
5.1.3	Морфологический анализ	64
5.2	Планирование научно-исследовательских работ	66
5.2.1	Структура работ в рамках научного исследования	66
5.2.2	Определение трудоемкости выполнения работ	67
5.2.3	Разработка графика проведения научного исследования	69
5.3.1	Расчет материальных затрат	71
5.3.2	Основная заработная плата исполнителям темы	72
5.3.3	Дополнительная заработная плата	74
5.3.4	Отчисления во внебюджетные фонды (страховые отчисления)	74
5.3.5	Накладные расходы	75
5.3.6	Формирование бюджета затрат научно-исследовательского проекта	76
5.4	Определение ресурсной, финансовой и экономической эффективности ресурсов	77

5.4.1 Оценка экономической эффективности проекта	77
5.4.2 Интегральный показатель ресурсоэффективности	77
6 Социальная ответственность	81
6.1 Правовые и организационные вопросы обеспечения безопасности	81
6.2 Производственная безопасность	83
6.3 Анализ вредных производственных факторов	84
6.3.1 Отклонения показателей микроклимата	84
6.3.2 Превышение уровня шума	86
6.3.4 Вибрации	89
6.3.5 Психофизиологические факторы	90
6.3.6 Перенапряжение зрительных анализаторов	91
6.4.2 Статическое электричество	93
6.5 Экологическая безопасность	94
6.6 Безопасность в чрезвычайных ситуациях	95
Заключение к главе “Социальная ответственность”	97
Заключение	98
Список используемых источников литературы	99
Список научных публикаций студента	103
Приложение А (справочное) Раздел на иностранном языке	104

## Введение

В настоящее время всё больше исследований направлено на решение задач с применением компьютерного зрения и искусственного интеллекта. Наиболее частыми являются решения и подходы с использованием распознавания жестов на основании инфракрасных сенсоров или нейронных сетей.

Актуальность распознавания жестов при помощи видеокамер обусловлена возможностью применения предлагаемого подхода для управления работой объектов без тактильного контакта и голосовой идентификации команд, а также своей простотой с точки зрения конечного пользователя.

В последнее время искусственный интеллект получил мощный толчок в своём развитии. Одним из направлений его развития является обработка и извлечение информации из изображений. Для этой задачи были специально разработаны свёрточные нейронные сети. На данной архитектуре и будет спроектирована работа.

Основной целью данной работы является обучение специальной архитектуры Mask-RCNN, способной одновременно распознавать несколько жестов рук человека без дополнительных устройств, а также обозначать их положение в кадре.

Данный тип управления имеет ряд преимуществ перед управлением при помощи голосовых команд или управления при помощи кнопок, а именно:

- 1) простота использования;
- 2) отсутствие физического контакта.

Данное устройство имеет множество применений:

- 1) командование мобильными роботами. Множественное распознавание жестов в кадре может позволить выдавать роботу несколько команд. Также есть возможность комбинировать жесты рук для увеличения числа возможных команд.

2) разработка с целью его внедрения в медицинские учреждения: поликлиники и больницы. Очень часто микробы и болезни передаются здоровому человеку от больного вследствие физического контакта с поверхностями. Чаще всего этими поверхностями являются обыкновенные двери. Также стоит отметить, что обычные датчики, стоящие в дверях у магазинов, не являются решением задачи, так как открывают двери всегда, когда приближается человек, что крайне неудобно в условиях поликлиники.

Для того, чтобы снизить риск заболевания стоит задача разработки бесконтактного способа взаимодействия с предметами и устройствами.

## **1 Обзор литературы**

### **1.1 История изобретений, при помощи которых решалась данная задача**

История распознавания жестов рук для компьютерного управления началась с изобретения интерфейсов управления на основе перчаток. Исследователи поняли, что жесты могут использоваться в качестве простых команд для взаимодействия с компьютером. Данное направление развивалось с разработкой очень точных акселерометров, инфракрасных камер и даже волоконно-оптических датчиков изгиба (оптических гониометров).

Первыми прототипами перчаток были: перчатка Sayre, перчатка LIT Массачусетского технологического института (MIT) и перчатка данных цифрового ввода [1]. В Sayre Glove, которая была разработана в 1977 году, использовались гибкие трубки с источником света на одном конце и фотоэлементом на другом, которые были установлены вдоль каждого пальца перчатки. Изгиб пальцев приводил к уменьшению количества света, проходящего между светодиодом и фотодиодом. Таким образом, система обнаруживала степень изгиба пальцев, используя напряжение, измеренное фотодиодом [2].

В течение 1980-х сенсорные технологии развивались быстрыми темпами отчасти из-за опасений холодной войны и естественного расширения промышленности во многих европейских странах. Эти сенсорные технологии проложили путь к быстрому развитию компьютерных технологий и периферийных устройств. Многие ведущие исследовательские группы по всему миру создали новую компьютерную периферию с ориентацией на рынок.

Первая коммерчески доступная Data Glove появилась в 1987 году. Это была улучшенная версия первого DataGlove, разработанного Циммерманом в 1982 году, который показан на рисунке 1 [3].



Рисунок 1 – Data Glove

Технология была похожа на ту, что использовалась в Sayre Glove в 1977 году. Однако в версии 1987 года вместо световых трубок использовалась волоконная оптика.

Также перчатка была оснащена датчиками в количестве от 5 до 15, увеличивающие его способность различать различные жесты. Многочисленные датчики, доступные на DataGlove, сделали его популярным среди исследователей в различных областях, и было разработано множество подобных устройств. Data Glove вдохновили на разработку Power Glove [4], который был продан Mattel Intellivision в качестве устройства управления для игровой консоли Nintendo в 1989 году.

Благодаря своим разработкам в начале 1980-х годов MIT Data Glove разительно развивалась, предлагая различные возможности для разных моделей. В настоящее время разработанная в рамках дочерней компании MIT AnthroTronix, asceleGlove, как показано на рисунке 2, представляет собой программируемую пользователем перчатку, которая регистрирует движения рук и пальцев в 3D. Другие модели, доступные от этой компании, включают Data Glove от 5DT для виртуальной реальности стоимостью от 1000 до 5000 долларов. Компания изначально разработала Data Gloves для обороны США для управления роботами. Их asceleGlove также используется в видеоиграх, спортивных тренировках или физической реабилитации.



Рисунок 2 – acceleGlove

Как показано на рис. 2, акселерометр лежит чуть ниже каждого кончика пальца и на тыльной стороне ладони. Акселерометры могут определять трехмерную ориентацию пальцев и ладони относительно силы тяжести, когда делается жест или любое движение. Точность этих измерений находится в пределах нескольких градусов, что позволяет программам различать небольшие изменения положения рук. Перчатка имеет отверстия для кончиков пальцев, которые позволяют пользователю печатать или писать во время ношения перчатки.

В 1996 году Iwai [5] предложили метод цветных перчаток, в котором были идентифицировано 10 областей пальцев. Они использовали несколько цветов для обозначения разных частей пальца и участков ладони, чтобы избежать проблемы окклюзии, от которой страдали многие подходы компьютерного зрения. В случае проблемы окклюзии некоторые части руки или пальцы закрыты окклюзией, и камера не может точно интерпретировать жест. Когда разные цветовые области обозначают разные участки руки (пальцы, ладонь), система может полагаться на цвет и границу для принятия обоснованных решений. Они использовали метод дерева решений для автоматического распознавания ограниченного количества жестов.

В последние годы все больше и больше исследований было сосредоточено на распознавании жестов на основе видения. По сравнению с распознаванием без зрения (проводные перчатки), распознавание на основе зрения более естественно и удобно, так как оно не ограничивает гибкость движений рук. На основе данных перчаток и электромагнитных волн была разработана цветная перчатка Ламберти, которую легко носить, не стесняя пользователя. Как показано на рисунке 3, цветная перчатка содержит отдельный цвет для отслеживания ладони, а пальцы помечены чередующимися цветами.

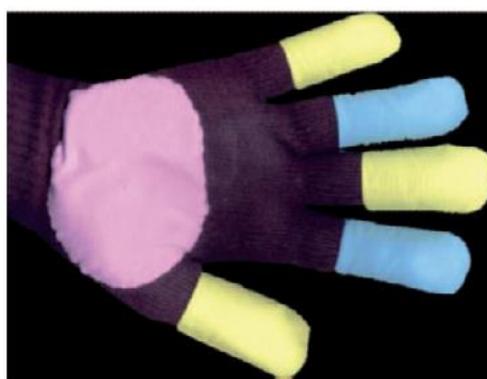


Рисунок 3 – Цветная перчатка Ламберти

Цель этого подхода состояла в том, чтобы разработать подход с очень низкими затратами по отношению к группе данных с гораздо большей гибкостью и очень низкими требованиями к вычислениям, чтобы пользователи с ограниченными возможностями могли использовать эту технологию в школьной среде.

## **1.2 Компания, которая занимается подобными решениями в Томске: Rubius**

Ей принадлежит разработка с названием “высокоточное распознавание жестов по облаку точек или цветному изображению (погрешность до 5 мм)”.

Алгоритмы компьютерного зрения и машинного обучения для устройств дополненной реальности распознают жесты по цветному изображению или

облаку точек, полученному с помощью 3D-сканера. Положение ключевых точек кисти определяется с точностью до 5 мм [6]. Ключевые точки изображены на рисунке 4.

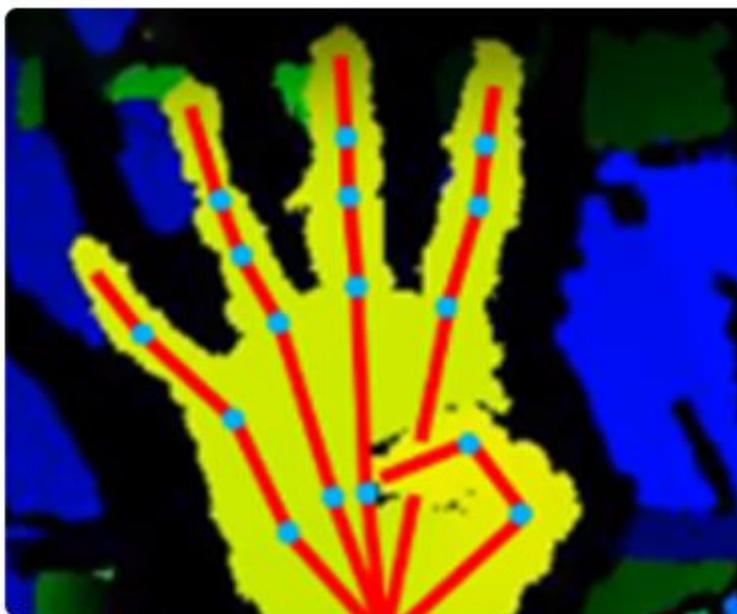


Рисунок 4 – Решение компании Rubius

### **1.3 Устройства, разрабатываемые в наше время**

#### **1.3.1 Распознавание жестов при помощи инфракрасных сенсоров на основе Intel RealSense**

Французскими учёными Квентином Смедтом, Хэземом Ванноусом и Джин-Филиппе Вандеборре была предложена идея распознавания жестов при помощи данных о скелете кисти в трёхмерном формате [7].

В данной работе используется устройство Intel RealSense. В основе данного устройства лежат инфракрасные сенсоры, при помощи которых можно получить информацию не только об изображении руки, но и ее положении в пространстве. Эта информация позволяет судить о том, насколько далеко находится рука от инфракрасного сенсора.

На рисунке 5 изображена информация о глубине и скелете кисти, полученная при помощи Intel RealSense. Сенсоры возвращают 22 соединения, по четыре на каждый палец, одно в центре ладони, и одно на запястье.

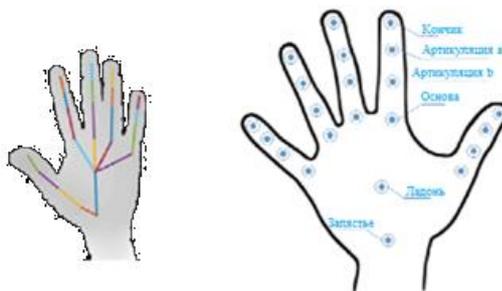


Рисунок 5 – Информация, полученная при помощи Intel RealSense

### **1.3.2 Динамическое распознавание жестов по направленным импульсным нейронным сетям**

Динамическое распознавание жестов по направленным импульсным нейронным сетям (PCNN) для взаимодействия человека с роботом в режиме реального времени предложено китайскими учёными и раскрыто в работе [8].

Здесь предлагается модифицированный алгоритм PCNN, называемый прямым PCNN (DPCNN), для распознавания жестов в реальном времени без большого количества обучающих данных. Трёхмерные координаты ключевых каркасных точек тела человека извлекаются в качестве входного объекта в виде последовательности из данных, захваченных Kinect, показанных на рисунке 6. Чтобы распознать жест ввода, сходство с ним вычисляется из матрицы объектов, созданной из элементов ввода и заданного шаблона последовательностей признаков.

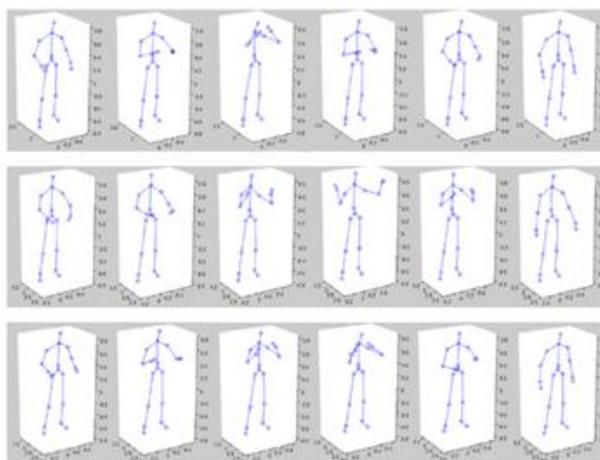


Рисунок 6 – Последовательность трехмерных данных

### 1.3.3 Распознавание жестов в реальном времени на основе сети перекалибровки функций с многомасштабной информацией

Последняя работа также была представлена китайскими учёными. Они использовали сверточные нейронные сети для решения задачи распознавания и классификации жестов [9].

Существуют две проблемы в процессе распознавания жеста, расположенного на большом расстоянии от камеры. Во-первых, жесты с различным соотношением размеров трудно идентифицировать; во-вторых, существует разная информация о признаках между признаками низкого уровня и признаками высокого уровня. Несмотря на то, что низкоуровневые признаки с высоким разрешением содержат больше деталей и информацию о позиции в кадре, это не способствует выявлению нужных признаков разных размеров при обнаружении. И, напротив, высокоуровневые признаки больше подходят для классификации по категориям, но имеют более низкое восприятие деталей из-за их более низкого разрешения.

Модуль в основном включает следующую стратегию:

Объединение признаков в разных масштабах. Чтобы лучше извлечь контекстную информацию разных масштабов, извлекается информация о характеристиках разных масштабов, при помощи сверточного ядра с размером шага 2 и размером  $3 \times 3$  и  $5 \times 5$ , соответственно.

Полная структура сверточной нейронной сети изображена на рисунке 7

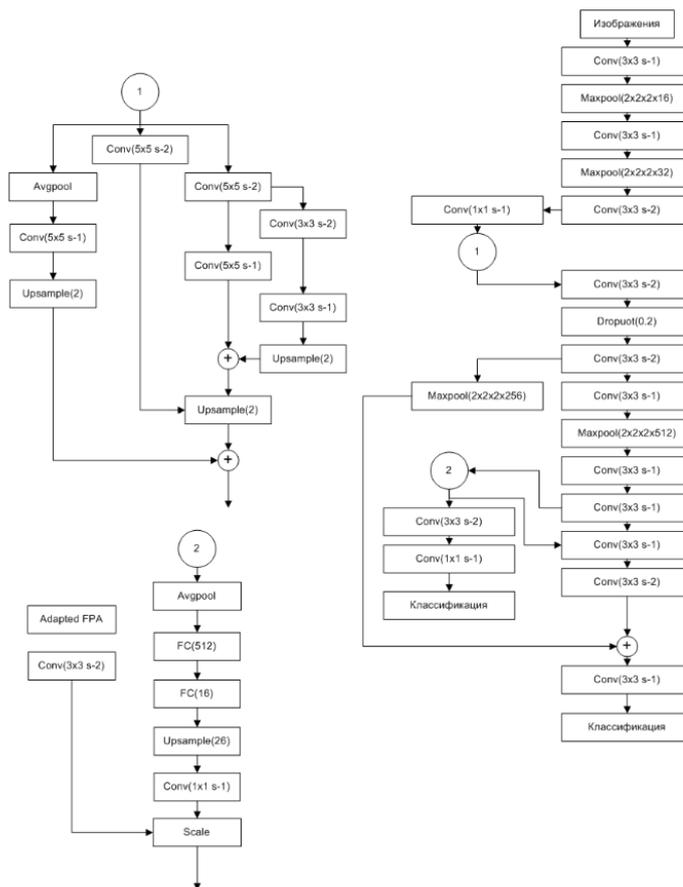


Рисунок 7 – Общая структура нейронной сети

### 1.3.4 Обработка и управление при помощи биомедицинских сигналов

В течение последнего десятилетия были разработаны методы разложения поверхностного электромиографического (ЭМГ) сигнала для извлечения нервной информации, передаваемой из спинного мозга в мышцы [10]. В работе оценивается точность идентификации действий рук по сигналам ЭМГ высокой плотности. Сигналы ЭМГ в автономном режиме были разложены на цепочки шипов двигателя (MUST) с использованием алгоритма слепого разделения источников. Был предложен подход распознавания жестов, основанный на классификации моторных единиц. Сначала MUST были объединены в группы, соответствующие 11 движениям. Выходной класс жестов был определен путем сравнения оцененного уровня активации каждого движения.

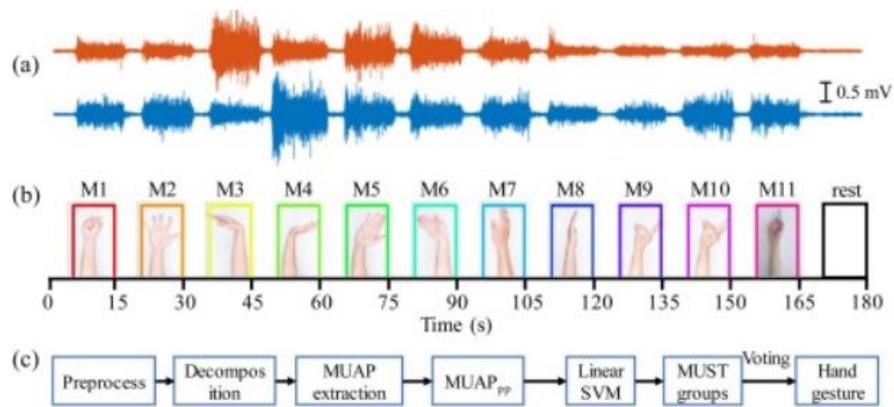


Рисунок 8 – Классификация жестов по ЭМГ

### 1.3.5 Система управления курсором мыши на основе жестов рук человека

Позиции рук, предложенные и использованные в статье для генерации событий мыши, показаны на рисунке 9. Положение рук «по умолчанию» (рисунок 9а) использовалось для движения мыши, а также для определения жестов «отпущена левая кнопка» или «отпущена правая кнопка». К событию «нажатие левой кнопки мыши» было прикреплено два положения: положение руки «нажата левая кнопка» № 1 (рисунок 9б) и положение руки № 2 «нажатие левой кнопки» (рисунок 9в). К событию нажатой правой кнопкой мыши было прикреплено положение руки «нажата левая кнопка» (рисунок 9г). Жест рассматривается как два положения руки [11]. Для отпущенной левой кнопки событие считалось последовательным: положение руки при нажатой левой кнопке следует за положением руки по умолчанию. Для отпущенной правой кнопкой мыши событие считалось последовательным: положение руки «нажатой правой кнопкой», за которым следует положение руки «по умолчанию».

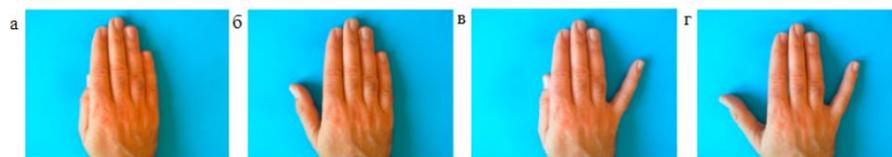


Рисунок 9 – Жесты для управления мышью

## 1.4 Анализ приведенных методов решения задачи распознавания жестов

Сравним существующие изобретения между собой, чтобы выявить преимущества и недостатки каждого из них.

Краткое описание рассмотренных алгоритмов, для решения задачи распознавания жестов рук приведено в таблице 1.

Таблица 1 – Сравнение существующих решений

№	Название	Оборудование	Входные величины	Алгоритм	Преимущества	Недостатки
1	Распознавание жестов при помощи инфракрасных сенсоров на основе Intel RealSense	Устройство Intel RealSense, вычислительная машина	Трехмерные данные о скелете руки	Использование инфракрасных сенсоров, получение соединений с последующей классификацией	Отсутствие влияния света на работу алгоритма, высокая точность распознавания	Необходимость наличия дополнительного устройства Intel RealSense, небольшая дальность работы, сложность алгоритма
2	Динамическое распознавание жестов по направленным импульсным нейронным сетям	Устройство Microsoft Kinect, вычислительная машина	Трехмерные данные о ключевых точках корпуса человека	Преобразование задачи распознавания в задачу кратчайшего пути при помощи PCNN	Отсутствие влияния света на работу алгоритма, скорость работы алгоритма	Сложность реализации, большие вычислительные затраты, необходимость наличия дополнительного устройства Microsoft Kinect
3	Распознавание жестов в реальном времени на основе сети перекалибровки и функций с многомасштабной информацией	Камера, вычислительная машина	Изображение, подающееся на вход сверточного слоя	Обработка изображения сверточными слоями для выделения признаков различных уровней	Высокая точность классификации жестов, хорошая дальность работы	Большая архитектура нейронной сети, невысокая скорость работы

Продолжение таблицы 1 – Сравнение существующих решений

4	Обработка и управление при помощи биомедицинских сигналов	Специальное оборудование для получения ЭКГ сигнала	Электромиографический (ЭМГ) сигнал для извлечения нервной информации, передаваемой из спинного мозга в мышцы	Получение ЭМГ сигнала и его классификация при помощи SVM-метода	Отсутствие влияния света на работу алгоритма, скорость работы алгоритма	Классификация ограниченного числа жестов, необходимость наличия проводов с сигналами
5	Система управления курсором мыши на основе жестов рук человека	Камера, вычислительная машина	Изображение, подающееся на вход сверточного слоя	Вычисление крайних точек, принадлежащих руке на изображении и последующая классификация жеста в зависимости от расстояния	Высокая точность классификации жестов, высокая скорость работы программы	Ограниченное пространство работы программы, возможность распознавания только одной руки

Исходя из таблицы 1, можно видеть, что рассмотренные методы обладают высокой точностью распознавания и классификации, но алгоритмы часто работают в ограниченной области, и не способны к распознаванию нескольких жестов одновременно.

Для решения задачи одновременного множественного распознавания жестов, а также их индикации на изображении и будет разрабатываться нейронная сеть.

В дальнейшем работа будет выполнена по следующим пунктам:

1. Получение изображения жеста;
2. Создание заготовок жестов для генерации набора данных;
3. Создание обучающего, оценочного и тестового набора данных;
4. Обучение нейронной сети;
5. Оценка точности распознавания на тестовой выборке.

## **2 Исследование и анализ нейронной сети Mask-RCNN**

### **2.1 Применение алгоритма решения изобретательской задачи АРИЗ-61**

#### **Аналитическая стадия**

Первый шаг. Поставить задачу. Одновременное множественное распознавание жестов рук.

Второй шаг. Представить себе идеальный конечный результат. Одновременное множественное распознавание нескольких жестов рук с высокой точностью и их индикация.

Третий шаг. Определить, что мешает достижению этого результата (т.е. найти противоречие). Отсутствие больших вычислительных мощностей, способных обучить нейронную сеть, созданную для данной задачи.

Четвертый шаг. Определить, почему мешает (найти причину противоречия). Из-за отсутствия мощного ЭВМ невозможно обучить сложную архитектуру нейронной сети.

Пятый шаг. Определить, при каких условиях не мешало бы (т.е. найти условия, при которых противоречие снимается). При высокой мощности вычислительных устройств и техническом совершенстве алгоритмов распознавания (Необходимы видеокарты с памятью от 11 Гбайт для обработки большого количества изображений).

#### **Оперативная стадия**

Первый шаг. Проверка возможных изменений в самом объекте (т.е. в данной машине, данном технологическом процессе)

1. Изменение размеров. Изменение размеров вычислительного устройства для увеличения мощности.

2. Изменение скорости. Изменение скорости работы алгоритма.

Второй шаг. Проверка возможности разделения объекта на независимые части.

1. Выделение «слабой» части. Вычислительной части не хватает мощности (Видеокарты с памятью до 8 Гбайт).

2. Выделение «необходимой и достаточной» части. Необходимой и достаточной частью является видеокамера с высоким разрешением (Разрешение видеокамеры Full HD).

Третий шаг. Проверка возможных изменений во внешней (для данного объекта) среде.

1. Измерение параметров среды. Изменение внешних источников света.

2. Использование внешней среды для выполнения полезных функций. Изменение внешней обстановки для более простого распознавания жеста на фоне.

Четвертый шаг. Исследование прообразов из других отраслей техники (поставить вопрос: как данное противоречие устраняется в других отраслях техники?) Установка более мощного оборудования, с более сложными алгоритмами (установка ЭВМ с параметрами: более 11 Гбайт видеопамяти, более 16 Гбайт оперативной памяти, видеокамера с разрешением 4к).

### **Синтетическая стадия**

Первый шаг. Внесение изменений в форму данного объекта (новой сущности машины должна соответствовать новая форма). Изменение внешнего вида для удобной комплектации более мощных устройств (увеличение объема до 0,001 м<sup>3</sup>).

Второй шаг. Внесение изменений в другие объекты, связанные с данными. Изменение путей получения и обработки информации при помощи камер.

Третий шаг. Внесение изменений в метод использования объекта. Создание алгоритмов с другими методами обработки информации (рекуррентные, полносвязные нейронные сети)

Четвертый шаг. На основе данного алгоритма возможно распознавания различных объектов, например, распознавание пешеходов для беспилотных автомобилей [12].

## 2.2 Задачи компьютерного зрения

Сегодня технологии технического зрения применяются повсеместно, их все чаще используют в самых разных сферах жизни человека, например, в камерах смартфонов для обнаружения лиц людей.

Чаще всего современные задачи компьютерного зрения бывают четырех видов:

- **Классификация** – классификация изображения по типу объекта, которое оно содержит;
- **Семантическая сегментация** – определение всех пикселей объектов определённого класса или фона на изображении. Если несколько объектов одного класса перекрываются, их пиксели никак не отделяются друг от друга;
- **Обнаружение объектов** – обнаружение всех объектов указанных классов и определение охватывающей рамки для каждого из них;
- **Сегментация объектов** – определение пикселей, принадлежащих каждому объекту каждого класса по отдельности;

Для наглядности примеры реализации классификации и семантической сегментации приведены на рисунке 10.



а)

б)

Рисунок 10 – Типы задач компьютерного зрения

а) Семантическая сегментация, б) Классификация с выделением объекта

Также приведены примеры обнаружения нескольких объектов и сегментации различных объектов на рисунке 11 [13].



а)

б)

Рисунок 11 – Примеры обнаружения объектов

а) Обнаружение объектов, б) Сегментация объектов

Для решения задачи сегментации жестов рук была использована архитектура Mask-RCNN (Сверточная нейронная сеть, накладывающая на обнаруженные объекты маски).

Данная архитектура была разработана исследователями компании Facebook в 2017г. Mask-RCNN это законченный вариант развития архитектуры Fast-RCNN, отличительной особенностью Mask-RCNN является более высокая скорость работы, а также возможность накладывать на обнаруженные объекты маски, что позволяет выделять эти объекты на общем изображении [14].

## 2.3 Описание архитектуры Mask-RCNN

### 2.3.1 Основа модели

В основе архитектуры Mask-RCNN находится обыкновенная сверточная нейронная сеть (CNN). Она извлекает признаки из изображения, для их дальнейшей обработки.

Свёрточная нейронная сеть — нейронная сеть, состоящая из одного или нескольких слоёв свёртки (convolutional layer). Также в СНС могут присутствовать слои пуллинга (pooling layer), стоящие после свёрточных слоёв. Замыкающими слоями являются полносвязные слои (fully connected layer) [15].

В СНС слои свёртки и пуллинга состоят из нескольких «уровней» нейронов, называемых картами признаков. Если обрабатывается изображение, карта признаков выглядит как матрица или двумерный массив нейронов. Полная топология свёрточных нейронных сетей представлена на рисунке 5.

Основой свёрточной нейронной сети являются слои свёртки [15]. Каждый слой включает в себя фильтры для каждого канала. Они обрабатывают предыдущий слой по частям (путём суммирования матричных фрагментов). Все веса ядра свёртки заранее неизвестны и изменяются в процессе обучения, в зависимости от входных данных. В конце слоя свёртки всегда стоит функция активации.

В качестве CNN была использована предобученная нейронная сеть ResNet 50. Все признаки, которые извлекаются ResNet 50 являются входом для последующих слоев [16].

### 2.3.2 Область интересов (RoI)

Регионы, полученные из RPN, могут иметь разные формы. Следовательно, был применен слой пуллинга, и все области преобразованы к одинаковой форме. Затем эти регионы проходят через полносвязные слои, которые нужны для прогноза метки класса и ограничивающей рамки.

Для этого сначала вычисляется интересующая область, чтобы можно было сократить время вычислений. Для всех прогнозируемых регионов вычисляется пересечение по объединению (IoU). IoU вычисляется по формуле:

$$IoU = \frac{S_{\Pi}}{S_o} \quad (1)$$

где  $S_{\Pi}$  – площадь пересечения;

$S_o$  – площадь объединения.

Площадь пересечения и площадь объединения состоят из пикселей, покрывающих объект на изображении. Если IoU больше или равно 0,5, то регион становится областью интереса. В противном случае этим конкретным

регионом пренебрегается. Данная операция производится для всех регионов, а затем выбирается только набор регионов, для которых IoU больше 0,5.

На рисунке 4 изображен поиск области интереса. После анализа изображения происходит выбор регионов, которые выделяют необходимый объект в рамку. Красным цветом изображена сама рамка – область интереса.

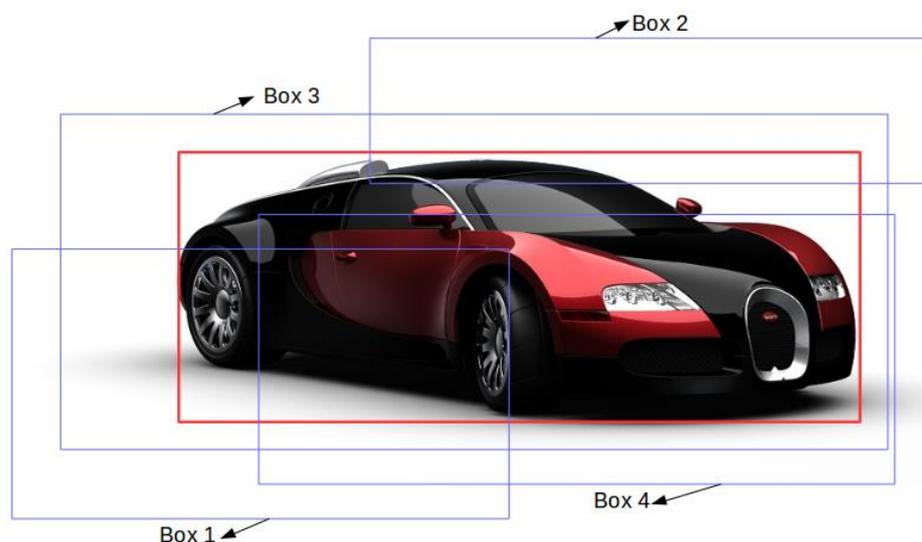


Рисунок 12 – Выбор области интереса

Здесь IoU области 1 и 2 меньше охватывают объект, следовательно, эти области будут игнорироваться. А области 3 и 4 намного больше охватывают объект. Следовательно, можно сказать, что области 3 и 4 являются областью интереса для этого конкретного изображения [17].

### 2.3.3 Region Proposal Network

Для того, чтобы быстрее обрабатывать изображения, было предложено вычислять регионы не по изначальному изображению, а по карте признаков, полученных из CNN. Для этого был добавлен модуль под названием Region Proposal Network (RPN). Новая архитектура целиком показана на рисунке 13.

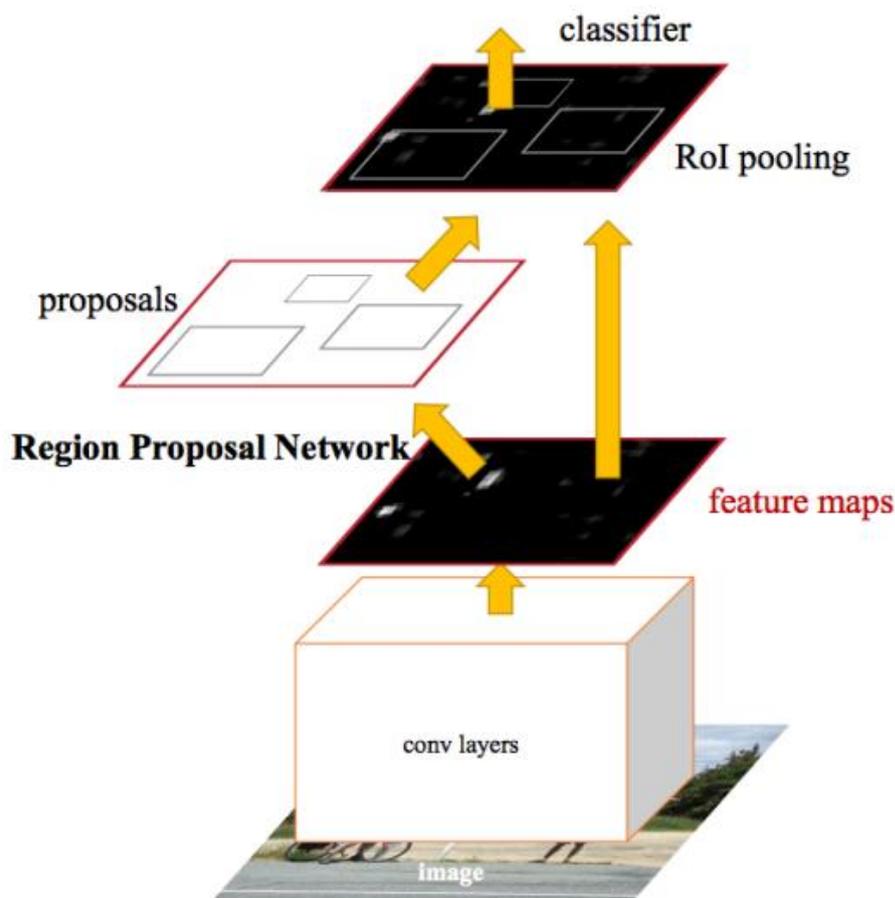


Рисунок 13 – Область RPN в модели Mask-RCNN

В рамках RPN, по извлечённым признакам сверточной нейронной сетью, скользящая маленькая нейросеть с ядром  $3 \times 3$ . Полученные с её помощью значения затем передаются в два полносвязанных слоя, которые расположены параллельно друг другу: регрессионный слой (reg) и классификационный слой (cls). На выходах этих слоёв образуются anchor-ы – это  $k$  рамок для каждого положения проходящего ядра, которые имеют различные соотношения сторон. Для каждого anchor-а регрессионный слой генерирует 4 координаты, корректирующие положение охватывающей рамки. Классификационный слой же вычисляет два числа – вероятности того, что рамка содержит хоть какой-то объект или что не содержит. В документе это иллюстрируется схемой, показанной на рисунке 14:

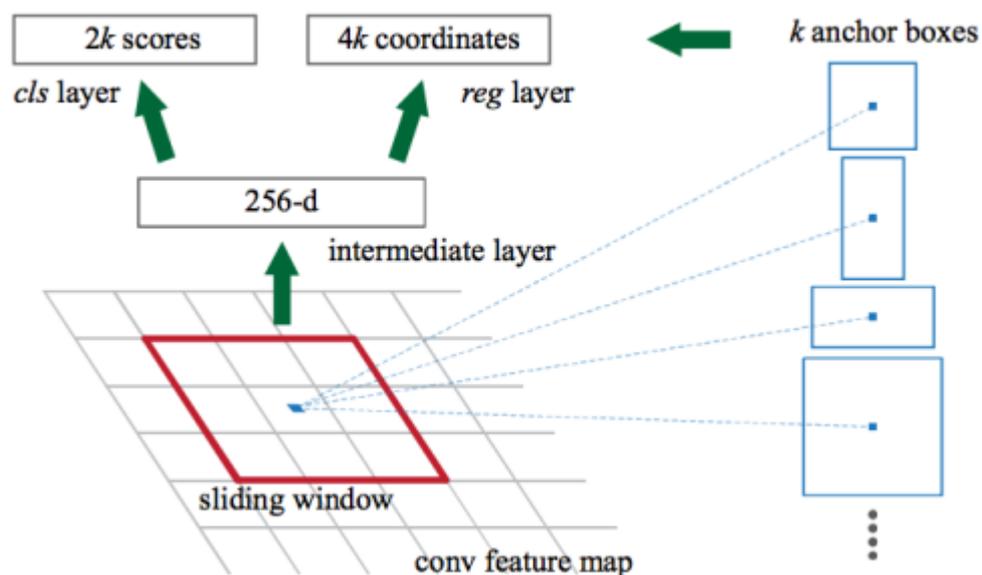


Рисунок 14 – Принцип работы регрессионного и классификационного слоев

Оба слоя RPN выдают только предположения для регионов-кандидатов. Те области, которые вероятнее всего содержат какой-либо объект, передаются дальше в модуль детектирования объектов и уточнения охватывающей рамки.

Для того, чтобы разделять признаки, получаемые в сверточной нейронной сети, между RPN и модулем детектирования, процесс обучения всей сети построен итерационно, с использованием нескольких шагов:

1. Инициализация и обучение на определение областей интереса RPN.
2. На основе предлагаемых RPN регионов происходит обучение сверточной нейронной сети.
3. Обученная сеть используется, чтобы инициализировать веса для RPN. Общие сверточные слои фиксируются и не изменяются. В это время производится донастройка только слоёв RPN.
4. С зафиксированными сверточными слоями окончательно донастраивается всей модели нейронной сети.

### 2.3.4 Наложение маски

В отличие от других сетей сверточного типа, в Mask R-CNN добавляется еще один “шаг”. На этом “шаге” происходит наложение маски, которая



Назначение Feature Pyramids, как и image pyramids, – улучшение качества детектирования объектов с учётом большого диапазона их возможных размеров.

В Feature Pyramid Network карты признаков, извлечённые последовательными слоями CNN с уменьшающейся размерностью, рассматриваются как некая иерархическая «пирамида», называемая bottom-up pathway. При этом карты признаков и нижних, и верхних уровней пирамиды обладают своими преимуществами и недостатками: первые имеют высокое разрешение, но низкую семантическую, обобщающую, способность; вторые – наоборот, как показано на рисунке 16.

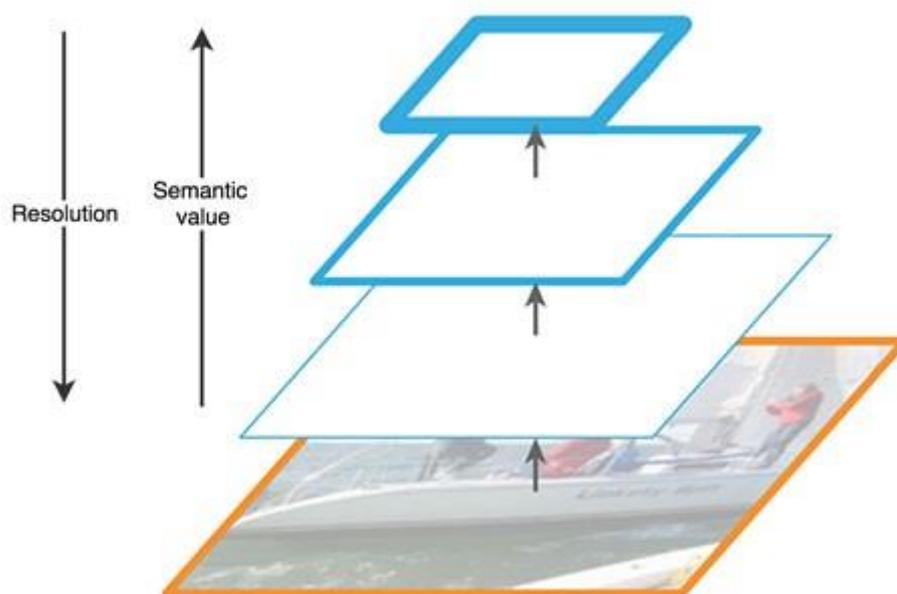


Рисунок 16 – FPN

Архитектура FPN позволяет объединить достоинства верхних и нижних слоёв при помощи добавления top-down pathway и lateral connections. Для этого карта каждого вышележащего слоя увеличивается до размера нижележащего и их содержимое поэлементно складывается. В итоговых предсказаниях используются результирующие карты всех уровней.

### 3 Подготовка нейронной сети Mask-RCNN для распознавания жестов рук

#### 3.1 Подготовка набора данных

Для того, чтобы обучить Mask-RCNN распознавать жесты необходимо подготовить обучающий набор данных. Для того, чтобы алгоритм распознавал жесты с высокой точностью (порядка от 90% до 95%) необходимо подготовить обучающую выборку с различными положениями рук на различном фоне.

Так как создание такого набора данных очень затратно по времени, было решено создать синтезированный набор данных. Его отличие от обычного состоит в том, что все изображения являются составными, а не полученными при помощи камер.

Также для каждого изображения необходима маска, чтобы обучать нейронную сеть в дальнейшем.

Полная последовательность действий для создания маски для одного жеста изображена на рисунке 17.



Рисунок 17 – Блок-схема процесса получения маски для одного жеста

Алгоритм создания синтезированного изображения выглядит следующим образом:

1. Получение исходной фотографии с одним из жестов рук (рисунок 18);



Рисунок 18 – Исходное изображение

2. “Вырезание” самого жеста из исходной фотографии (рисунок 19). Данная операция проводилась при помощи графической программы Gimp. На каждый класс было создано по три подобных изображения для создания разнообразия, таким образом сеть сможет лучше определять жест под углом, немного отличным от исходного.



Рисунок 19 – Изображения с жестом

3. Наложение вырезанного жеста на новое изображение, которое содержит только фон. На данном этапе происходит не только наложение, но и поворот и масштабирование накладываемого изображения жеста, как показано на рисунке 20. Это необходимо для того, чтобы нейронная сеть лучше

обучилась определять жесты, так как во время работы алгоритма жест может находиться на разном расстоянии от камеры и иметь другой угол поворота.



Рисунок 20 – Изображения с добавленным жестом

4. Создание изображения с масками, для дальнейшего обучения Mask-RCNN (рисунок 21). Это необходимо, чтобы обучить нейронную сеть правильно накладывать маски на искомые объекты.

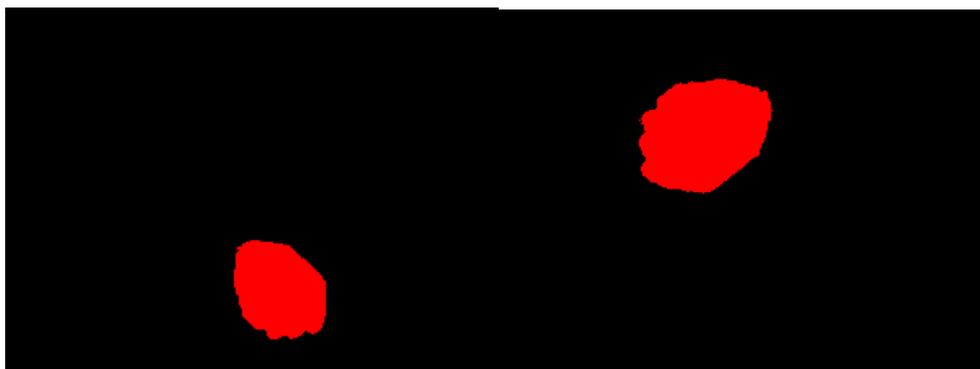


Рисунок 21 – Маски добавленных изображений

Также при работе возможен поиск нескольких жестов рук, как изображено на рисунке 22. Каждая маска имеет свой цвет, который сообщает при обучении нейронной сети класс жеста.

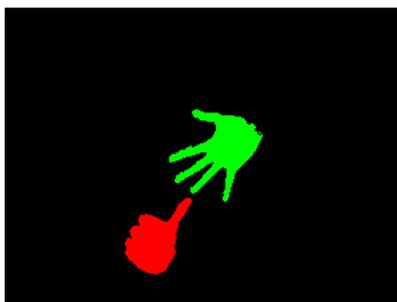


Рисунок 22 – Маски добавленных изображений

5. Добавление информации о классе каждой маски и ее цвете в отдельный json-файл. Пример json-файла изображен на рисунке 23.

```
{
  "masks": {
    "images/00000000.jpg": {
      "mask": "masks/00000000.png",
      "color_categories": {
        "(255, 0, 0)": {
          "category": "Victory",
          "super_category": "Hand",
          "(0, 255, 0)": {
            "category": "BigFinger",
            "super_category": "Hand"
          }
        }
      }
    },
    "images/00000001.jpg": {
      "mask": "masks/00000001.png",
      "color_categories": {
        "(255, 0, 0)": {
          "category": "Phone",
          "super_category": "Hand"
        }
      }
    },
    "images/00000002.jpg": {
      "mask": "masks/00000002.png",
      "color_categories": {
        "(255, 0, 0)": {
          "category": "Victory",
          "super_category": "Hand"
        }
      }
    },
    "images/00000003.jpg": {
      "mask": "masks/00000003.png",
      "color_categories": {
        "(255, 0, 0)": {
          "category": "BigFinger",
          "super_category": "Hand",
          "(0, 255, 0)": {
            "category": "Victory",
            "super_category": "Hand",
          },
          "(0, 0, 255)": {
            "category": "Fist",
            "super_category": "Hand"
          }
        }
      }
    }
  }
}
```

Рисунок 23 – json-файл

Для создания синтезированного набора данных необходимо сгенерировать изображения, на которых будут помещены руки.

На рисунке 24 изображена схема генерации синтезированного набора данных для обучения и проверки точности.



Рисунок 24 – Блок-схема алгоритма генерации изображений

Данная нейронная сеть обучена на распознавания пяти классов жестов рук:

1. Большой палец (Thumb Finger);
2. Кулак (Fist);
3. Открытая ладонь (Palm);
4. Телефон (Phone);
5. Победа (Victory).

Все существующие типы изображений приведены на рисунке 25.



Рисунок 25 – Пять классов жестов

При помощи написанного скрипта было создано около 500 изображений для обучения модели и 100 для проверки работоспособности (рисунок 26). Также сгенерированы маски (рисунок 27).

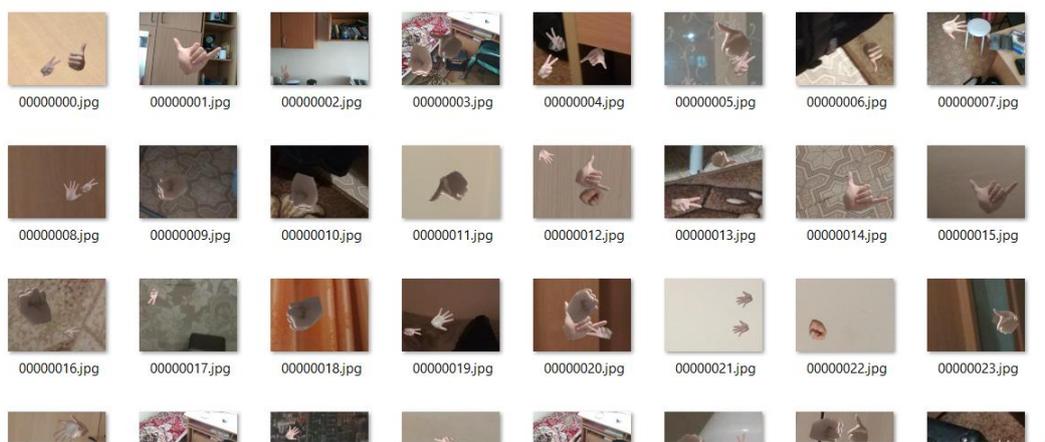


Рисунок 26 – Сгенерированные изображения

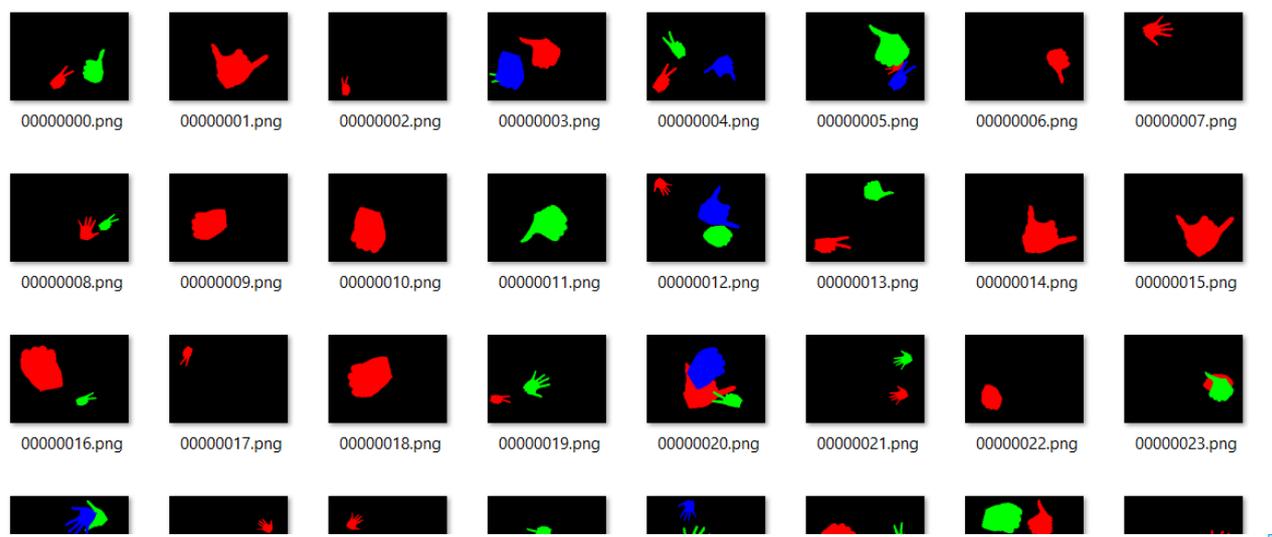


Рисунок 27 – Сгенерированные маски

### 3.2 Конфигурация Mask-RCNN

В основе обучения нейронной сети лежат библиотеки Keras и TensorFlow. При помощи первой строится архитектура СНС и устанавливается оптимизатор. TensorFlow же позволяет оптимально тренировать нейронную сеть для распознавания жестов.

Для корректной работы свёрточной нейронной сети необходимо правильно создать её архитектуру, а также верно взять следующие параметры:

- Выбор способа размещения обучающих и тестовых данных;
- Выбор backbone модели;
- Выбор параметра Learning Rate;
- Количество GPU;
- Размер изображений для обучения.

#### 3.2.1 Выбор способа размещения обучающих и тестовых данных

Для обучения нейронной сети Mask-RCNN необходимы большие мощности и большое количество времени. Поэтому было решено воспользоваться Google Colab.

Google Colab является бесплатным облачным сервисом, на котором можно обучить свою модель нейронной сети.

Для корректного обучения нейронной сети необходимо предустановить следующие библиотеки:

1. Keras 2.0.8. Данная библиотека позволяет использовать заранее созданные слои нейронных сетей, из которых в дальнейшем будет проектироваться архитектура Mask-RCNN.

2. TensorFlow-gpu 1.14.0. Библиотека TensorFlow позволяет конвертировать веса в удобную для библиотеки форму для дальнейшего обучения нейронной сети.

3. Cython. Данный модуль позволяет конвертировать код Python в код C для быстрого исполнения функций, что, следовательно, ускоряет работу нейронной сети в целом.

4. Rusocotools помогает работать с данными Mask-RCNN: парсить их, загружать, получать информацию для данных из json файлов и т.д.

5. Numpy необходим для представления данных и весов нейронной сети в особом формате, при котором повышается быстродействие ЭВМ для обучения нейронной сети, а так же ее работы.

6. PIL. Позволяет загружать, обрабатывать и хранить изображения при помощи языка Python.

Также для обучения все изображения были загружены на Google Drive. Размер всего набора данных составляет 100 Мбайт. В них входят 500 обучающих изображений и 500 изображений масок.

Также для оценки модели созданы 100 изображений и 100 изображений масок.

При помощи специальных функций можно подключить Google Drive с находящимися на нем изображениями непосредственно к Google Colab (рисунок 28).

```
from google.colab import drive

# This will prompt for authorization.
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

Рисунок 28 – Подключение диска

### 3.2.2 Выбор backbone модели

Далее необходимо выбрать backbone модель для самого обучения. Эта часть нейронной сети является основной, так как в ней происходит поиск признаков жестов.

Выбор наиболее подходящей модели будет осуществляться путем оценки ошибки между предсказанием модели и заранее заданными данными при обучении на одной эпохе. Результаты обучения приведены в таблице 2.

Таблица 2 – Результаты работы сети

Название backbone модели	Ошибка предсказания Класса жеста	Ошибка наложения маски	Ошибка выделения контура	Количество эпох обучения
Resnet50	0,207	0,209	0,231	1
Resnet101	0,354	0,182	0,429	1
imagenet	0,482	0,335	0,573	1

Исходя из полученных данных, можно сделать вывод, что количество нейронная сеть resnet50 быстрее обучается на одной эпохе. Resnet101 показывает лучший результат по наложению маски на изображение, но уступает по двум другим значениям ошибки.

По полученным данным было решено использовать в качестве backbone модели архитектуру resnet50.

### 3.2.3 Выбор параметра Learning Rate.

Коэффициент скорости обучения ( $\eta$ ) – это параметр алгоритмов обучения нейронной сети, корректирующей свои веса с помощью метода градиентного спуска. Данный коэффициент позволяет контролировать величину коррекции веса на каждой итерации, при обучении нейронной сети при помощи алгоритма обратного распространения ошибки.

$$\Delta\omega = -\eta \frac{dE}{d\omega} \quad (2)$$

где  $E$  – значение ошибки;

$\omega$  – значение веса.

Его значение выбирается в диапазоне от нуля до единицы. Если указать значение, равное нулю, то веса не будут изменяться.

При его выборе нужно быть осторожным. При больших значениях (0,7 - 1) шаг коррекции может быть слишком большим. Такой алгоритм будет работать быстрее (т.е. требуется меньше итераций для нахождения минимума функции ошибок). Однако может снизиться точность настройки модели на минимум функции ошибки, что потенциально увеличит ошибку обучения.

При малых значениях коэффициента (0,001 - 0,3) соответствуют меньшему шагу коррекции весов. При этом количество шагов (или эпох) обучения, необходимых для поиска экстремума, как правило, увеличивается, но точность алгоритма также увеличивается, что снижает ошибку обучения. На практике коэффициент скорости обучения обычно выбирается экспериментально.

Для получения наименьшего значения ошибки были протестированы несколько значений коэффициента при обучении нейронной сети на пяти эпохах.

Таблица 3 – Результаты работы сети

Learning Rate	Ошибка предсказания Класса жеста	Ошибка наложения маски	Ошибка выделения контура	Количество эпох обучения
0,01	0,529	0,462	0,684	5
0,1	0,223	0,284	0,301	5
0,5	0,476	0,291	0,342	5

По полученным данным видно, что самая большая ошибка у коэффициента со значением 0,1. Из-за низкого изменения значений весов, ошибка очень медленно снижается.

Коэффициент, равный 0,5 показал хороший результат, но значения ошибок перестали изменяться на третьей эпохе обучения. Это означает, что значение коэффициента слишком большое.

При значении 0,1 все значения ошибок стремились к нулю, что является доказательством правильного обучения нейронной сети.

### 3.2.4 Выбор количества GPU

Графические процессоры обладают высокой производительностью - благодаря архитектуре своего ядра графические процессоры эффективно справляются с большим количеством простых однотипных задач. Это одно из отличий видеокарт от процессоров.

Обучение нейронных сетей на CPU занимает месяцы, а видеокарты справятся с задачей за несколько дней. Однако они также потребляют меньше энергии из-за меньшей инфраструктуры.

Кроме того, графические процессоры используют преимущества параллелизма. В свою очередь, нейронные сети – это параллельные алгоритмы, поэтому графические процессоры хорошо подходят для машинного обучения.

Также видеокарты оптимизированы для матричных операций и ускоряют их - они необходимы нейронным сетям для получения результата.

Так как мощности серверов, предоставленных Google Colab ограничены, необходимо подобрать такое количество GPU, при котором будет возможно быстро обучать нейронную сеть.

Для оценки количества графических процессоров было решено засечь время обучения на одной эпохе.

Таблица 4 – Результаты работы сети

Количество GPU	Время, мин
2	45
4	26
8	12

Исходя из полученных данных, можно сделать вывод о том, что скорость обучения прямо пропорциональна количеству используемых при обучении графических процессоров. Максимально возможное количество GPU, которые можно использовать в Google Colab равно восьми.

### **3.2.5 Выбор размера изображений**

От размера изображений при обучении нейронной сети также зависит очень многое. Изображения с большим разрешением могут хранить много информации, но обучение может происходить очень долго вследствие затратной по времени их обработки.

И, наоборот, если использовать изображения с меньшим разрешением, то затраты вычислительных мощностей на их обработку будут меньше, но часть информации, хранящейся на картинках, может быть утрачена.

Для выбора наиболее подходящего разрешения, будут протестированы изображения с разным разрешением. В результате будет оцениваться ошибка распознавания жеста и время, затраченное на обучение на двух эпохах.

Таблица 5 – Результаты работы сети

Разрешение изображения	Ошибка предсказания Класса жеста	Ошибка наложения маски	Ошибка выделения контура	Время на обучение, мин
128×128	0,207	0,269	0,182	11
256×256	0,185	0,237	0,152	14
512×512	0,174	0,201	0,149	22

По полученным данным видно, что при использовании изображений с разрешением  $128 \times 128$ , нейронная сеть обучается быстрее всего, но точность такой сети уступает остальным исследуемым вариантам.

Разрешение изображений  $512 \times 512$  показало наивысшую точность при распознавании жестов, но у такой конфигурации заметно возросло время обучения нейронной сети. Также разница в точности распознавания и наложения маски у данного разрешения и варианта с разрешением  $256 \times 256$  довольно незначительна.

### 3.3 Итоги конфигурации Mask-RCNN

В результате конфигурации Mask-RCNN была получена следующая сверточная нейронная сеть с возможностью накладывать маски на найденные объекты:

1. Сформирован обучающий набор данных. Количество классов жестов рук равно пяти:
  - a. Большой палец (Thumb Finger);
  - b. Кулак (Fist);
  - c. Открытая ладонь (Palm);
  - d. Телефон (Phone);
  - e. Победа (Victory).

2. Количество изображений для тренировки нейронной сети равно 500. Количество для оценки работы нейронной сети (тестовая выборка) равно 100.

3. В качестве вычислительного устройства выступает облачный сервис Google Colab, на котором используется 8 GPU.

4. В качестве backbone модели выбрана модель ResNet50, на которой и будет обучаться основная часть нейронной сети.

5. Параметр скорости обучения равен 0,1 для избежания переобучения.

6. Разрешение изображений для обучающей выборки равно  $256 \times 256$ .

#### 4 Оценка работы обученной нейронной сети на тестовой выборке

Нейронная сеть Mask-RCNN обучалась на обучающей выборке пятнадцать эпох, в результате итоговые ошибки равны соответственно:

Таблица 6 – Ошибки разных составляющих нейронной сети

Ошибка предсказания класса жеста	Ошибка наложения маски	Ошибка выделения контура
0,207	0,0998	0,0464

Ошибка предсказания класса жеста считается по формуле (1) [18]:

$$L_{\text{клс}}(p, g) = -\log p_g$$
$$L(p, g) = \frac{1}{N_{\text{клс}}} \sum_i L_{\text{клс}}(p_i, p_i^*) \quad (3)$$

где  $p_g$  – это вероятность класса  $g$ ;

$N_{\text{клс}}$  – количество изображений с жестами, проходящих за одну итерацию обучения;

$p_i$  – вероятность принадлежности объекта к выделенному региону на изображении;

$p_i^*$  – метка истинности, является ли на самом деле выделенный регион искомым объектом.

Ошибка выделения контура считается следующим образом: сверточная нейронная сеть выводит четыре значения  $(r, c, h, w)$ , где  $(r, c)$  задают значения положения левого угла и  $(h, w)$  высоту и ширину окна. Ошибка выделения контура увеличивается, когда выходные данные нейронной сети сильно отличаются от помеченных  $(r, c, h, w)$  в обучающем наборе [19].

Mask-RCNN позволяет генерировать маски для каждого класса без конкуренции между классами. В данном случае, к каждому пикселю применяется функция сигмоиды, по значению которой находится вероятность принадлежности пикселя к искомому объекту. На основании разницы между

пикселями маски, данной в обучающей выборке, и пикселями маски, которую построила нейронная сеть, и получается ошибка наложения маски [20].

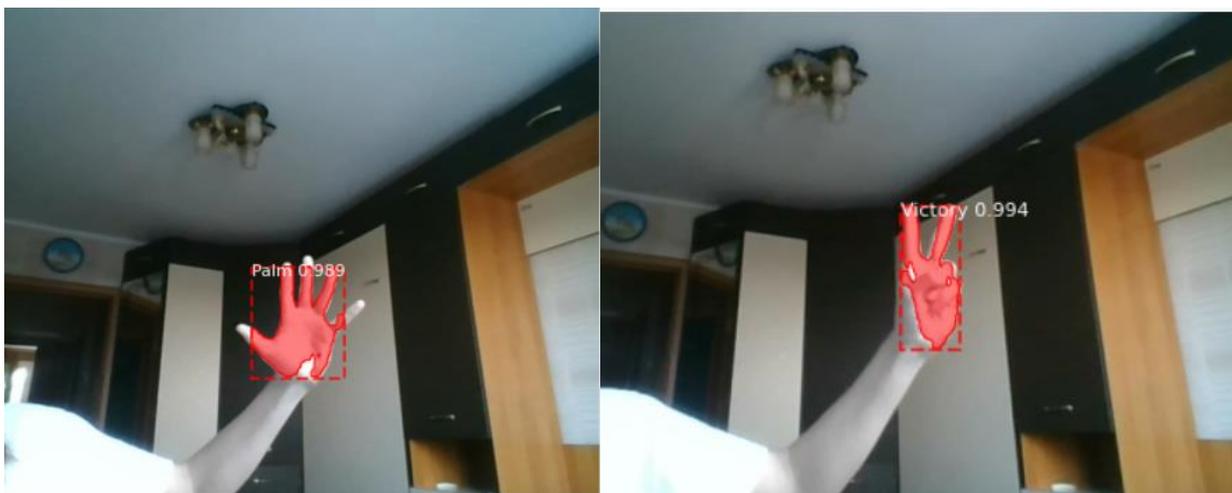
Результаты работы нейронной сети по обнаружению жестов на синтезированном наборе данных можно увидеть на рисунке 29.



Рисунок 29 – Результаты работы нейронной сети на синтезированном наборе данных

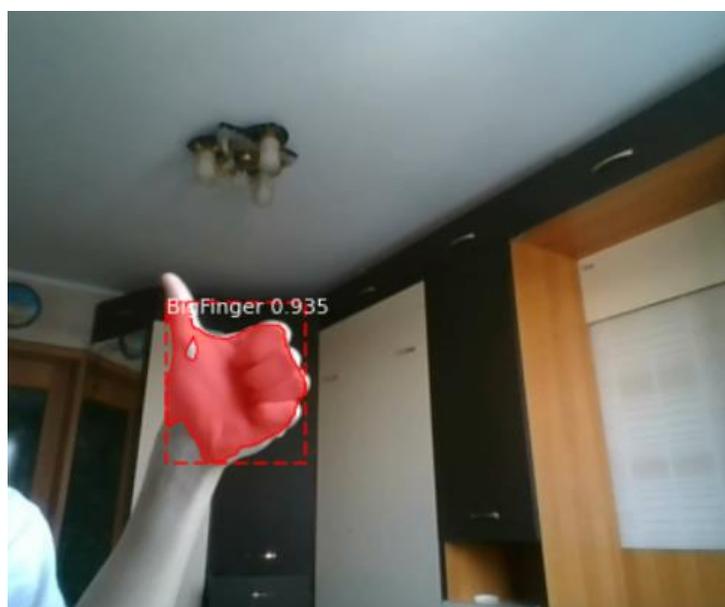
Нейронная сеть хорошо определяет класс жеста, но выделение и наложение маски происходит с погрешностями, как видно на рисунке 29.

Затем была проверена работа нейронной сети на реальных изображениях. Результат работы Mask-RCNN можно увидеть на рисунке 30.



а)

б)



в)

а) успешное распознавание жеста “Рука”, б) успешное распознавание жеста “Победа”, в) успешное распознавание жеста “Большой палец”

Рисунок 30 – Результаты работы нейронной сети

На реальных данных, выделение контуров жестов и наложение масок на жест менее точное, чем на синтезированном наборе данных, но класс жеста определяется правильно.

#### 4.1 Зависимость точности распознавания жеста от расстояния между камерой и самим жестом

Для оценки работы нейронной сети, в зависимости от расстояния до жеста было решено провести исследование следующим образом:

- 1) Подготовить 30 изображений жестов, по 6 изображений на каждый класс;
- 2) Проанализировать работу нейронной сети на расстоянии 30 см, 50 см, 60 см, 80 см и 100 см;
- 3) Оценить отношение верно и неверно распознанных жестов;
- 4) Оценить отношение верно распознанных жестов к общему количеству изображений.

В ходе исследования будет проведена проверка на двух различных веб-камерах: Logitech C210 и HP Wide Vision HD. Параметры камер отражены в таблице 7.

Таблица 7 – Характеристики веб-камер

Параметры	Название камеры	
	HP Wide Vision HD	Logitech C210
Количество мегапикселей	0,92 МП	0,3 МП
Разрешение	1280×720	640×480
Частота кадров, кадр/с	30	30

Результаты оценки работы нейронной сети с камерой HP Wide Vision HD можно увидеть в таблице 8.

Таблица 8 – Количество распознанных жестов

Расстояние между камерой и жестом	Количество верно распознанных жестов	Количество неверно распознанных жестов	Общее число изображений
30 см	14	8	30

Продолжение таблицы 8 – Количество распознанных жестов

50 см	21	8	30
60 см	27	3	30
80 см	26	5	30
100 см	19	10	30

Результаты оценки работы нейронной сети с камерой Logitech C210 можно увидеть в Таблице 9.

Таблица 9 – Количество распознанных жестов

Расстояние между камерой и жестом	Количество верно распознанных жестов	Количество неверно распознанных жестов	Общее число изображений
30 см	12	7	30
50 см	20	6	30
60 см	25	3	30
80 см	25	4	30
100 см	20	9	30

Отношение верно распознанных жестов к неверно распознанным будет вычисляться по формуле

$$M = \frac{C}{I}, \quad (4)$$

где С – количество верно распознанных жестов;

I – количество неверно распознанных жестов;

Отношение верно распознанных жестов к общему числу изображений будет вычисляться по формуле

$$V = \frac{C}{A} * 100\%, \quad (5)$$

где С – количество верно распознанных жестов;

A – количество всех изображений;

Результаты отношений можно увидеть в таблице 10.

Таблица 10 – Отношения

Расстояние между камерой и жестом	$M_1$	$V_1$ (%)	$M_2$	$V_2$ (%)
30 см	1,75	46	1,71	40
50 см	2,625	70	3,33	66
60 см	9	90	8,33	83
80 см	5,2	86	6,25	83
100 см	1,9	63	2,22	66

Графики отношений приведены на рисунках 31 и 32.

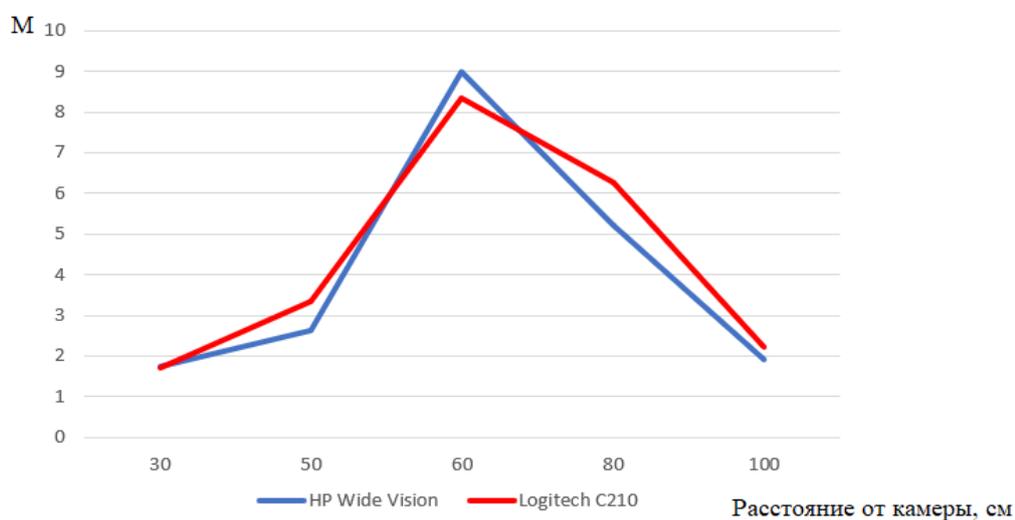


Рисунок 31 – Зависимость отношения  $M$  от расстояния до камеры

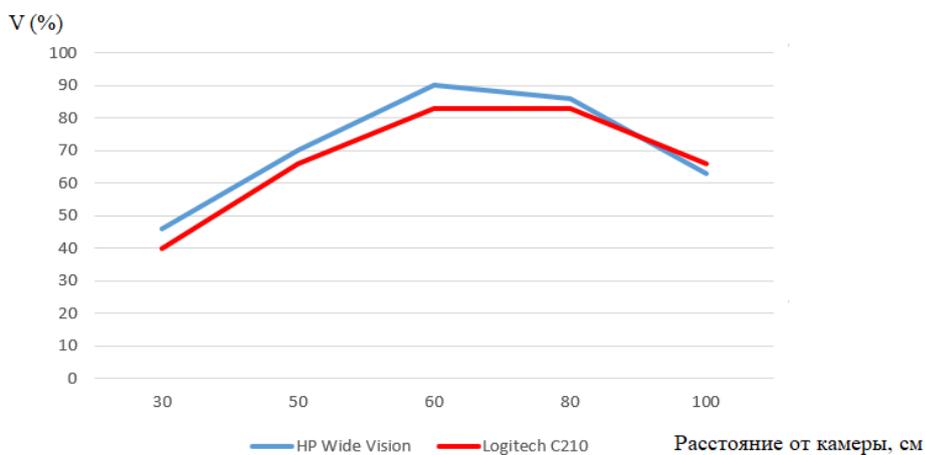


Рисунок 32 – Зависимость отношения  $V$  от расстояния до камеры

По полученным расчетам видно, что распознавание жестов, на дистанции менее 50 сантиметров, имеет довольно низкую точность. Количество верно распознанных жестов меньше количества неверно распознанных жестов почти в два раза. Также общая точность, по отношению к общему числу распознанных жестов меньше 50 %. Такие низкие показатели возможны из-за того, что нейронная сеть обучалась на изображениях с жестами небольшого размера, и, следовательно, не может различить жесты большого размера. Данная проблема изображена на рисунке 33.

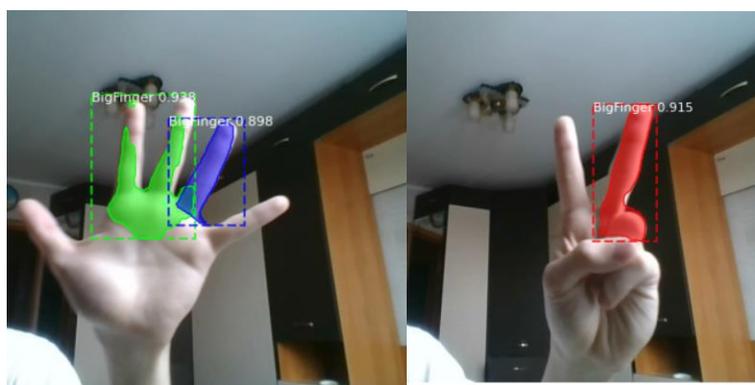


Рисунок 33 – Некорректная работа нейронной сети

На расстоянии 60 сантиметров, нейронная сеть показала наибольшее отношение верно распознанных жестов к неправильно распознанным. Также на этом расстоянии точность составляет 90%, что является хорошим показателем. Результаты работы нейронной сети на расстоянии 60 сантиметров показаны на рисунке 34.

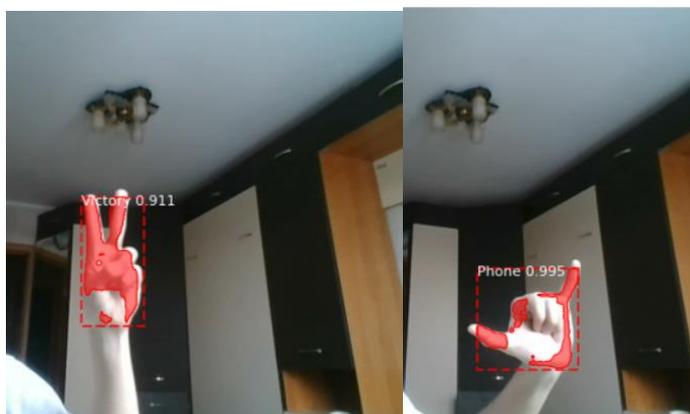


Рисунок 34 – Работа нейронной сети

Но на данном расстоянии все еще иногда происходят неверные распознавания жестов, как показано на рисунке 35.

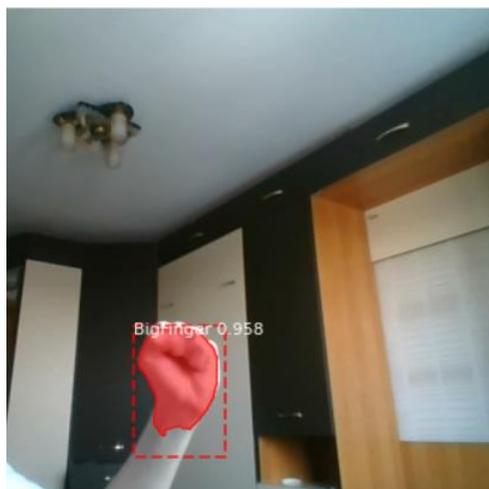


Рисунок 35 – Неверно распознанный жест “кулак”

#### 4.2 Оценка распознавания нейронной сетью каждого жеста

Для оценки работы нейронной сети, было решено провести исследование следующим образом:

- 1) Подготовить 150 изображений жестов, по 30 изображений на каждый класс;
- 2) Проанализировать способность нейронной сети распознавать каждый класс жестов на расстоянии 60 см;
- 3) Оценить отношение верно распознанных жестов к общему числу изображений одного класса;

Отношение верно распознанных жестов к общему числу изображений одного класса вычисляется по формуле:

$$T = \frac{C}{A} * 100\%, \quad (6)$$

где С – количество верно распознанных жестов;

А – количество неверно распознанных жестов;

Результаты оценки работы нейронной сети можно увидеть в таблице 11.

Таблица 11 – Количество распознанных жестов

Класс жеста	Количество верно распознанных жестов	Общее количество изображений данного класса	T (%)
Palm	29	30	96
Victory	27	30	90
Big Finger	24	30	80
Fist	22	30	73
Phone	26	30	86

График количества распознанных жестов приведен на рисунке 36.

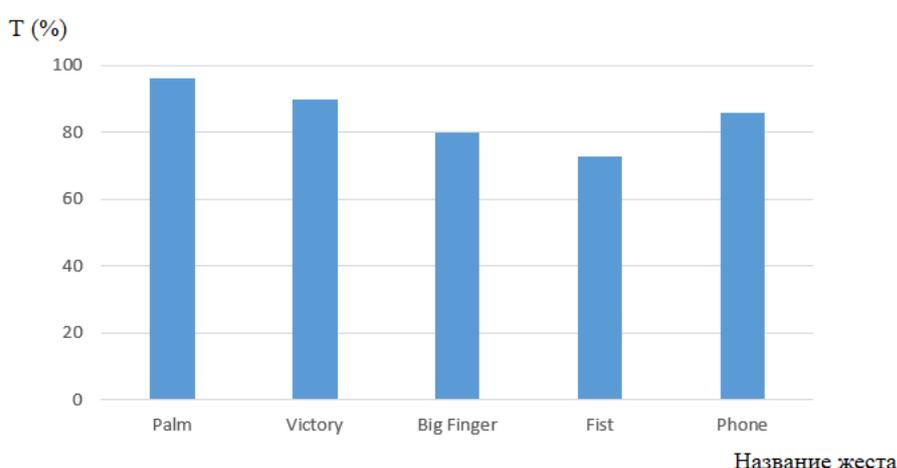


Рисунок 36 – Зависимость отношения T от класса жеста

По проведенному опыту можно сделать вывод, что классы “Кисть”, “Победа” и “Телефон” были успешно распознаны в более 80 % случаев. Классы “Большой палец” и “Кулак” были распознаны в довольно малом количестве (более 70 % случаев).

Для того, чтобы выяснить какой жест нейронная сеть чаще всего путает с другим, было решено провести опыт по подсчету количества неверно распознанных жестов. Результаты опыта отображены в таблице 12. Было взято 30 изображений исходного класса, расположенного в таблице по вертикали, и подсчитано количество раз, когда жест был спутан с другим, расположенным по горизонтали.

Таблица 12 – Подсчет жестов, которые были спутаны с другими

Название жеста	Кисть	Победа	Большой палец	Кулак	Телефон
Кисть	30	1	2	0	0
Победа	2	30	3	0	0
Большой палец	0	0	30	6	2
Кулак	0	0	5	30	4
Телефон	0	0	4	5	30

Как видно из таблицы 12, жесты “Большой палец”, “Кулак”, и “Телефон” наиболее часто были неверно распознаны. Нейронная сеть наиболее часто путает эти три класса, из-за общей визуальной схожести жестов.

#### **4.3 Оценка способности распознавания жестов, скрытых за преградой**

Иногда возникают случаи, когда жест частично скрыт за каким-то другим предметом. Целью данного опыта является выяснение способности нейронной сети Mask-RCNN распознавания жеста, который частично находится за преградой. Для оценки работы нейронной сети, было решено провести исследование следующим образом:

- 1) Подготовить 50 изображений жестов, по 10 изображений на каждый класс;
- 2) Проанализировать работу нейронной сети на расстоянии 60 см от камеры;
- 3) Получить количество верно и неверно распознанных жестов;
- 4) Оценить отношение верно распознанных жестов к общему количеству изображений.

Результаты оценки работы нейронной сети можно увидеть в таблице 13.

Таблица 13 – Количество распознанных жестов

Название жеста	Количество верно распознанных жестов	Количество неверно распознанных жестов	Количество нераспознанных жестов	Общее число изображений
Кисть	2	3	5	10
Победа	0	2	8	10
Большой палец	6	3	1	10
Кулак	2	8	0	10
Телефон	4	3	3	10

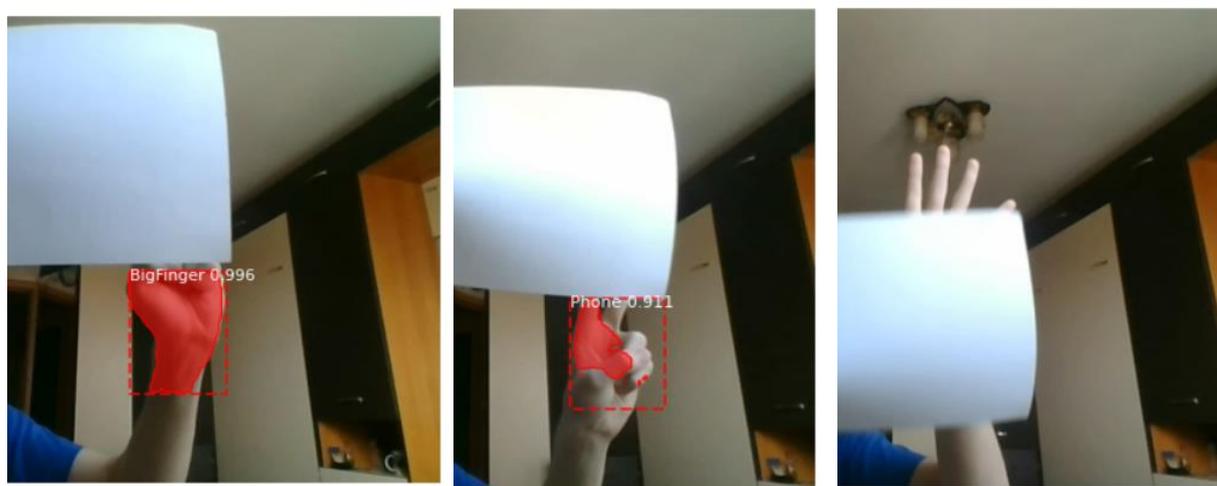


Рисунок 37 – Некорректная работа нейронной сети

По полученным результатам (рисунок 37) можно сделать вывод, что нейронная сеть Mask-RCNN не может распознавать жесты, находящиеся за какой-либо преградой. Количество верно распознанных жестов крайне мало. Нейронная сеть либо вообще их почти не распознает, как в случае с жестами “Кисть” и “Победа”, либо путает с другими классами жестов, из-за недостатка информации, как в случае с остальными жестами.

## **5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение**

### **5.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения.**

Цель экономического раздела - провести детальный анализ проекта по критериям конкурентоспособности и ресурсоэффективности. Оценить перспективность проекта, определить трудоемкость и график работ, а также рассчитать интегральный показатель ресурсоэффективности.

#### **5.1.1 Потенциальные потребители результатов исследования**

В ходе выпускной квалификационной работы были проведены исследования потребности рынка в устройстве распознавания жестов.

Для определения потенциальных потребителей требуется определить целевой рынок и произвести его сегментирование. В данном случае целевым рынком являются:

- Медицинские учреждения, а именно возможность внедрения устройства в двери;
- Предприятия, где возможно управление объектами небольшим количеством команд, например, передвижение кранов.

#### **5.1.2 Анализ конкурентных технических решений**

Устройства распознавания жестов являются узкоспециализированными, и не получили широкого применения. Известно только одно устройство, продающееся на рынке – Kinect компании Microsoft. Устройство Kinect позволяет распознавать жесты при помощи двух камер: обычной и камеры с инфракрасным излучением. При помощи двух камер, Kinect способен построить 3D модель человека, в том числе и его руки. Сейчас это устройство стоит 8500 рублей в среднем.

Недостатками является то, что данное устройство имеет много ненужных частей (две камеры), сложно в адаптации под задачу распознавания жестов, а также дорого стоит.

Из этого можно сделать вывод, что требуется разработать устройство, которое будет доступно для использования широкому кругу предприятий.

Проведем анализ конкурентного технического решения с помощью оценочной карты.

Позиция разработки и конкурентов оценивается по каждому показателю экспертным путем по пятибалльной шкале, где 1 – наиболее слабая позиция, а 5 – наиболее сильная. Веса показателей, определяемые экспертным путем, в сумме должны составлять 1.

Техническими критериями оценки являются следующие параметры:

1) Энергоэффективность. Характеристика проявляется в потреблении энергии устройствами. У Microsoft Kinect помимо самого устройства необходима ЭВМ, для обработки входящих сигналов. У разрабатываемого устройства для распознавания жестов необходим только ноутбук.

2) Удобство в эксплуатации. Данный параметр определяется простотой использования алгоритма. У разрабатываемого устройства достаточно использовать ноутбук и встроенную веб-камеру, тогда как для Kinect нужно устанавливать само устройство в определенное положение для успешного захвата жеста в кадре.

3) Помехоустойчивость. Характеристика определяется способностью справляться с шумами, создаваемыми окружающими источниками света. Оба устройства в целом справляются с этим, но иногда обильный поток света слишком зашумляет изображения.

4) Потребность в вычислительных ресурсах. Характеристика определяется необходимостью вычислительной мощности для решения задачи. Устройство Microsoft наиболее эффективно применяет свои алгоритмы.

5) Дальность работы. Оба устройства работают в достаточном

диапазоне, на котором распознаются жесты рук.

б) Способность распознавать частично скрытый жест. Оба устройства плохо справляются с данной задачей.

7) Функциональная мощность. Параметр определяется количеством распознаваемых жестов за раз. Разрабатываемое устройство за счет архитектуры нейронной сети Mask-RCNN способно распознавать более двух жестов в кадре, что не способен делать конкурент.

Экономическими критериями являются:

1) Цена. Параметр складывается из стоимости всех необходимых электронных компонентов для обеспечения работы алгоритма распознавания жестов.

2) Предполагаемый срок эксплуатации. Оба устройства могут работать достаточно продолжительное время, без выхода из строя.

3) Уровень проникновения на рынок. Этот параметр выше у устройства Kinect, так как компания Microsoft имеет отлаженную систему выпуска продукта на рынок. Оценочная карта приведена в таблице 14.

Таблица 14 – Оценочная карта

Критерии оценки	Вес критерия	Баллы		Конкурентоспособность	
		Б <sub>ф</sub>	Б <sub>к1</sub>	К <sub>ф</sub>	К <sub>к1</sub>
1	2	3	4	5	6
<b>Технические критерии оценки ресурсоэффективности</b>					
1.Энергоэффективность	0.12	5	4	0,6	0,48
2.Удобство в эксплуатации	0.12	5	4	0,6	0,48
3.Помехоустойчивость	0.08	4	4	0,32	0,32
4.Потребность в вычислительных ресурсах	0.12	3	4	0,36	0,48
5.Дальность работы	0.12	4	4	0,48	0,48
6. Способность распознавать частично скрытый жест	0.05	3	3	0,15	0,15

Продолжение таблицы 14 – Оценочная карта

7. Функциональная мощность	0.12	5	4	0,6	0,48
<b>Экономические критерии оценки ресурсоэффективности</b>					
8.Цена	0.08	5	3	0.4	0.24
9. Предполагаемый срок эксплуатации	0.09	4	4	0.36	0.36
10. Уровень проникновения на рынок	0.1	4	5	0.4	0.5
<b>Итого</b>	<b>1</b>	<b>44</b>	<b>39</b>	<b>4.27</b>	<b>3.97</b>

где  $B_{\phi}$  - разрабатываемое устройство,  $B_{K1}$  – устройство Kinect.

Конкурентоспособность рассчитываем по формуле:

$$K = \sum B_i * V_i \quad (7)$$

где  $K$  – конкурентоспособность научной разработки;

$V_i$  – вес показателя (в долях единицы);

$B_i$  – балл  $i$ -го показателя.

Недостатками перед конкурентом являются потребность в вычислительных ресурсах и уровень проникновения на рынок. Потребность в вычислительных ресурсах обусловлена более сложной архитектурой и вычислениями внутри нейронной сети, в следствии чего, в отличие от конкурента, необходимо больше времени на прогнозирование жеста.

Также у конкурента выше уровень проникновения на рынок, так как компания Microsoft уже обладает отлаженными инструментами для выпуска и распространения устройства Kinect на рынке.

### 5.1.3 Морфологический анализ

Морфологический подход основан на систематическом исследовании всех теоретически возможных вариантов, которые вытекают из

закономерностей объекта исследования. Анализ охватывает все возможные варианты. Путем комбинирования вариантов можно получить большое количество различных решений, ряд которых представляет практический интерес. Составим таблицу, в которой будут отражены возможные варианты исполнения по различным проблемам разработки (таблица 15).

Таблица 15 – Морфологический анализ

Характеристика	Варианты исполнения		
	1	2	3
Язык программирования	C++	Java	Python
Вид нейронных сетей	Сверточные	Рекуррентные	Полносвязные
Реализация	Стационарный компьютер	Одноплатный компьютер NanoPi	Одноплатный компьютер Raspberry Pi

Для создания устройства, способного распознавать жесты был выбран язык программирования Python, так как он имеет много библиотек для создания свёрточных нейронных сетей, а также прост в освоении и реализации. Свёрточные сети наиболее подходят для создания алгоритма распознавания жестов, так как прекрасно работают с изображениями и успешно способны выделять необходимые признаки.

Устройство разрабатывается на стационарном компьютере. Так как для работы нейронной сети необходимы значительные вычислительные мощности, остальные варианты не подходят ввиду их низких мощностей.

## 5.2 Планирование научно-исследовательских работ

### 5.2.1 Структура работ в рамках научного исследования

Трудоемкость выполнения проекта оценивается в человеко-часах и зависит от множества факторов, которые сложно учесть при разработке. Для реализации проекта необходимо 2 исполнителя – научный руководитель (НР), инженер (И). Этапы работы проекта представлены в таблице 16.

Таблица 16 – Этапы работы проекта

Основные этапы	№ этапа	Содержание работ	Исполнитель
Разработка задания	1	Постановка задачи	НР, И
Выбор направления исследования	2	Обзор научно-технической базы	НР, И
	3	Разработка и утверждение ТЗ	НР, И
	4	Составление календаря проекта	И
	5	Разработка вариантов исполнения проекта	НР, И
Теоретические исследования	6	Сбор и подготовка обучающих данных	И
	7	Поиск способа обучить модель нейронной сети	И
	8	Создание модели нейронной сети	И
	9	Обучение и конфигурация нейронной сети	И
Экспериментальные исследования	10	Проверка работы нейронной сети на тестовых данных	НР, И
	11	Оценка работоспособности в зависимости от типов жестов, расстояния до камеры и оценка способности распознавания частично скрытых жестов	И

### 5.2.2 Определение трудоемкости выполнения работ

Основная часть стоимости разработки зачастую приходится на заработную плату исполнителей, поэтому важно определить трудоемкость каждого из участников. Ожидаемая трудоемкость находится по формуле:

$$t_{ожi} = \frac{3 * t_{минi} + 2 * t_{маxi}}{5} \quad (8)$$

где  $t_{ожi}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы чел.-дн.;

$t_{минi}$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы. Оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств, чел.-дн.;

$t_{маxi}$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой работ. Пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств, чел.-дн.

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях  $T_{pi}$ , учитывающая параллельность выполнения работ несколькими исполнителями:

$$t_{pi} = \frac{t_{ожi}}{Ч_i} \quad (9)$$

где  $t_{pi}$  – продолжительность одной работы, раб.дн.;

$t_{ожi}$  – ожидаемая трудоемкость выполнения одной работы, чел.-дн.;

$Ч_i$  – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой. Для примера произведём расчёт первого этапа работы руководителя:

$$T_{ki} = T_{pi} * k_{кал} \quad (10)$$

где  $t_{ki}$  – продолжительность выполнения  $i$ -й работы в календарных днях;

$t_{pi}$  – продолжительность выполнения  $i$ -й работы в рабочих днях;

$k_{кал}$  – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$K_{\text{кал}} = \frac{t_{\text{кал}}}{t_{\text{кал}} - t_{\text{вых}} - t_{\text{пр}}} = \frac{365}{365 - 104 - 16} = 1,5 \quad (11)$$

где  $t_{\text{кал}}$  – количество календарных дней в году;

$t_{\text{вых}}$  – количество выходных дней в году;

$t_{\text{пр}}$  – количество праздничных дней в году.

В таблице 17 находятся расчеты этапов отдельных видов работ.

Таблица 17 – Временные показатели проведения научного исследования

Название работы	Грудоёмкость работ						Длительность работ в рабочих днях		Длительность работ в календарных днях	
	$t_{\text{min}}$ чел-дни		$t_{\text{max}}$ чел-дни		$t_{\text{ож}}$ чел-дни		$t_{\text{pi}}$	$t_{\text{ki}}$		
	Инженер	Научный руководитель	Инженер	Научный руководитель	Инженер	Научный руководитель	Одновременное выполнение работ		Одновременное выполнение работ	
							Инженер	Научный руководитель	Инженер	Научный руководитель
Постановка задачи	5	3	8	6	6,2	4,2	3,1	2,1	5	4
Обзор научно-технической базы	7	2	12	4	9	2,8	4,5	1,4	7	3
Разработка и утверждение ТЗ	7	1	12	2	9	1,4	4,5	0,7	7	2
Составление календаря проекта	3	0	5	0	3,8	0	3,8	0	6	0

Продолжение таблицы 17 – Временные показатели проведения научного исследования

Разработка вариантов исполнения проекта	9	4	16	7	11,8	5,2	5,9	2,6	9	4
Сбор и подготовка обучающих данных	15	0	27	0	19,8	0	19,8	0	10	0
Поиск способа обучить модель нейронной сети	6	0	12	0	8,4	0	8,4	0	13	0
Создание модели нейронной сети	8	0	14	0	10,4	0	10,4	0	11	0
Обучение и конфигурация нейронной сети	10	0	20	0	14	0	14	0	11	0
Проверка работы нейронной сети на тестовых данных	6	1	14	8	9,2	3,8	4,6	1,9	7	3
Оценка работоспособности в зависимости от типов жестов, расстояния до камеры и оценка способности распознавания частично скрытых жестов	10	0	15	0	12	0	12	0	18	0
Составление пояснительной записки	7	0	14	0	9,8	0	9,8	0	15	0
Итого							100,8	8,7	119	16

### 5.2.3 Разработка графика проведения научного исследования

По таблице 17 составим таблицу 18 – продолжительность работ проведения НИОКР.

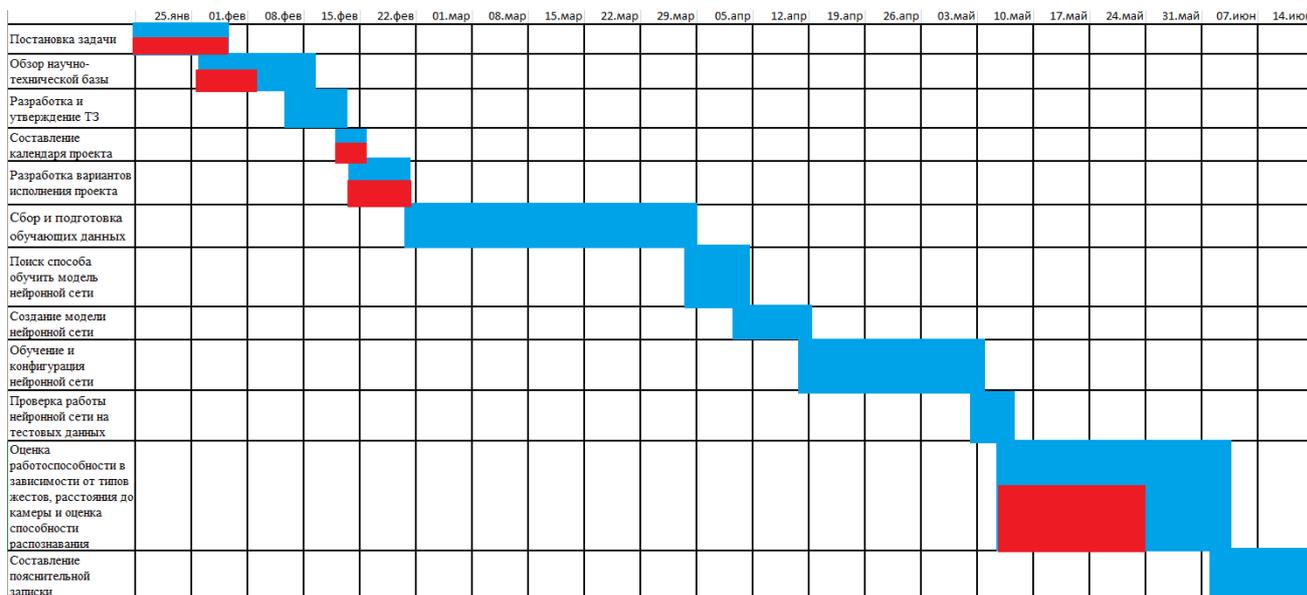
Таблица 18 – Продолжительность работ проведения НИОКР

№ этапа	Этап	Исполнитель	$T_{ki}$
1	Постановка задачи	НР	4
		С	5
2	Обзор научно-технической базы	НР	3
		С	7
3	Разработка и утверждение ТЗ	НР	2
		С	7
4	Составление календаря проекта	С	6
5	Разработка вариантов исполнения проекта	НР	4
		С	9
6	Сбор и подготовка обучающих данных	С	10
7	Поиск способа обучить модель нейронной сети	С	13
8	Создание модели нейронной сети	С	11
9	Обучение и конфигурация нейронной сети	С	11
10	Проверка работы нейронной сети на тестовых данных	НР	3
		С	7
11	Оценка работоспособности в зависимости от типов жестов, расстояния до камеры и оценка способности распознавания частично скрытых жестов	С	18
12	Составление пояснительной записки	С	15

По данным из таблицы 18 «Продолжительность работ проведения НИОКР» создадим диаграмму Ганта инженера и научного руководителя (таблица 19), которая строилась при максимальном количестве дней каждой

работы.

Таблица 19 – Диаграмма Ганта



### 5.3 Бюджет научно-технического исследования

Бюджет научно-технического исследования должен быть основан на достоверном отображении всех видов расходов, связанных выполнением проекта. В процессе формирования бюджета разработки используется следующая группировка затрат по статьям:

- материальные затраты разработки;
- основная заработная плата исполнителей темы;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- затраты на научные и производственные командировки;
- накладные расходы.

#### 5.3.1 Расчет материальных затрат

Материальные затраты — это затраты организации на приобретение сырья и материалов для создания готовой продукции. Все материальные затраты на создание устройство по распознаванию жестов приведены в таблице 20.

Таблица 20 – Материальные затраты

№	Название	Стоимость, руб
1	Ноутбук HP Pavilion 15	50000
Итого		50000

### 5.3.2 Основная заработная плата исполнителям темы

Данная статья расходов включает заработную плату научного руководителя и инженера (в его роли выступает исполнитель проекта), а также премии, входящие в фонд заработной платы. Расчет основной заработной платы выполняется на основе трудоемкости выполнения каждого этапа и величины месячного оклада исполнителя.

С учётом документа «Положение об оплате труда», доцент, к.т.н., работающий в ТПУ имеет оклад равный 35111,5 руб.

Инженер имеет оклад 22695,68 руб.

Данная статья расходов включает заработную плату научного руководителя и инженера, в его роли выступает исполнитель проекта. Расчет основной заработной платы сводится в таблице 20.

Статья включает основную заработную плату работников, непосредственно занятых выполнением НИИ, и дополнительную заработную плату:

$$Z_{зп} = Z_{осн} + Z_{доп}, \quad (12)$$

где  $Z_{осн}$  – основная заработная плата;

$Z_{доп}$  – дополнительная заработная плата (12-20 % от  $Z_{осн}$ ).

Основная заработная плата руководителя, лаборанта, инженера рассчитывается по следующей формуле:

$$Z_{осн} = Z_{дн} \cdot T_p, \quad (13)$$

где  $Z_{осн}$  – основная заработная плата одного работника;

$T_p$  – продолжительность работ, выполняемых научно-техническим работником, раб. дн.;

$Z_{дн}$  – среднедневная заработная плата работника, руб.

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{дн} = \frac{Z_m \times M}{F_d}, \quad (14)$$

где  $Z_m$  – месячный должностной оклад работника, руб.;

$M$  – количество месяцев работы без отпуска в течение года:

при отпуске в 24 раб. дня  $M=11,2$  месяца, 5-дневная неделя;

при отпуске в 56 раб. дней  $M=10,4$  месяца, 6-дневная неделя;

$F_d$  – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн.

Баланс рабочего времени показан в таблице 21.

Таблица 21 – Баланс рабочего времени

Показатели рабочего времени	Научный руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней	66	118
Потери рабочего времени на отпуск	56	24
Действительный годовой фонд рабочего времени	243	223

Месячный должностной оклад работника:

$$Z_m = Z_{окл} \cdot k_p, \quad (15)$$

где  $Z_{окл}$  – оклад, руб.;

$k_p$  – районный коэффициент, равный 1,3 (для Томска).

Научный руководитель имеет должность доцента и степень кандидата технических наук оклад на весну 2019 год составлял 33664 руб., затем осенью был проиндексирован на 4,3% и составил 35111,5 руб.

Оклад инженера на весну 2019 года составил 21760 руб., затем осенью был проиндексирован на 4,3% и составил 22695,68 руб.

В таблице 22 отражен расчет основной заработной платы.

Таблица 22 – Расчёт основной заработной платы

Исполнители	Разряд	$k_T$	$Z_{окл}$ , руб.	$k_p$	$Z_M$ , руб.	$Z_{дн}$ , руб.	$T_p$ , раб. дн.	$Z_{осн}$ , руб.
Научный руководитель	–	–	35111,5	1,3	45644,95	1953,53	8,7	16995,71
Инженер	–	–	22695,68		29504,5	1481,82	100,8	149367,46
Итого $Z_{осн}$								166363,17

### 5.3.3 Дополнительная заработная плата

Дополнительная заработная плата включает заработную плату за не отработанное рабочее время, но гарантированную действующим законодательством. Расчет дополнительной заработной платы ведется по формуле:

$$Z_{доп} = k_{доп} \times Z_{осн} \quad (16)$$

где  $k_{доп}$  – коэффициент дополнительной заработной платы, на стадии проектирования принимается равным 0,12.

Результаты по расчетам дополнительной заработной платы сведены в таблицу 23.

Таблица 23 – Затраты на дополнительную заработную плату

Исполнители	Основная зарплата (руб.)	Коэффициента дополнительной заработной платы ( $k_{доп}$ )	Дополнительная зарплата (руб.)
Руководитель	16995,71	0,12	2039,48
Студент	149367,46	0,12	17924,09
Итого:			19963,58

### 5.3.4 Отчисления во внебюджетные фонды (страховые отчисления)

Величина отчислений во внебюджетные фонды определяется исходя из

формулы:

$$З_{внеб} = k_{внеб} * (З_{осн} + З_{доп}) \quad (17)$$

где  $k_{внеб}$  – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

На 2018 г. в соответствии с Федеральным законом от 24.07.2009 №212-ФЗ установлен размер страховых взносов равный 30%. На основании пункта 1 ст.58 закона №212-ФЗ для учреждений, осуществляющих образовательную и научную деятельность в 2018 году пониженная ставка – 27,1%. Отчисления во внебюджетные фонды представлены в таблице 24.

Таблица 24 – Отчисления во внебюджетные фонды

Исполнитель	Основная заработная плата, руб.	Дополнительная заработная плата, руб.
Руководитель проекта	16995,71	2039,48
Студент	149367,46	17924,09
Коэффициент отчислений во внебюджетные фонды	27,1%	
<b>Итого</b>		
<b>Руководитель</b>	<b>5158,53</b>	
<b>Студент</b>	<b>45336,01</b>	
<b>Итого</b>	<b>50494,54</b>	

### 5.3.5 Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов, оплата услуг связи, электроэнергии и т.д. Расчет накладных расходов определяется по формуле:

$$З_{внеб} = \sum CT * k_{нр} \quad (18)$$

где  $k_{нр}$  – коэффициент, учитывающий накладные расходы,  $C_T$  – затраты по статьям накладных расходов

Величину коэффициента накладных расходов можно взять в размере 16%.

$$Z_{нак} = (50494,54 + 19035,19 + 167291,55) * 0.16 = 37891,4 \text{ руб.}$$

### 5.3.6 Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно-исследовательской работы является основой для формирования бюджета затрат проекта. Определение бюджета затрат на научно-исследовательский проект по каждому варианту исполнения приведен в таблице 25.

Таблица 25 – Расчет бюджета затрат НТИ

Наименование статьи	Сумма (руб.)
1. Материальные затраты НТИ	50000
2. Затраты на заработную плату научному руководителю	19035,19
3. Затраты на заработную плату инженеру	167291,55
4. Затраты на отчисления во внебюджетный фонд	50494,54
5. Накладные расходы	37891,4
Бюджет затрат НТИ	324712,68

ТПУ частично финансирует проект, предоставляя аудиторию для работы, заработную плату и иные ресурсы.

## **5.4 Определение ресурсной, финансовой и экономической эффективности ресурсов**

В результате исследования были определены затраты на проект по обучению и тестированию нейронной сети Mask-RCNN. Бюджет составляет 53466,09 рублей. Приняв во внимание все преимущества разрабатываемого ПО, можно предположить, что продукт будет иметь спрос на рынке и станет достойным аналогом своим конкурентам.

### **5.4.1 Оценка экономической эффективности проекта**

Актуальным аспектом качества выполненного проекта является экономическая эффективность его реализации, т.е. соотношение обусловленного ей экономического результата (эффекта) и затрат на разработку проекта. Так как последние являются единовременными, то мы имеем дело с частным случаем задачи оценки экономической эффективности инвестиций, т.е. вложением денежных средств в предприятие, организацию, отраслевую, региональную, социально-экономическую систему и т.п. (т.н. объекты инвестиций) с целью получения определенного результата в будущем.

Так как рынок устройств, распознающих жесты рук, на данный момент имеет только устройство Kinect, который более не производится, то есть возможность занять эту нишу при помощи разрабатываемого в рамках данной работы алгоритма. Исходя из этого, данную нейронную сеть можно будет использовать в дальнейшей разработке устройств для распознавания жестов и их реализации на рынке.

### **5.4.2 Интегральный показатель ресурсоэффективности**

В качестве вариантов исполнения были выбраны ближайшие аналоги:

1. Разработка инженером и научным руководителем.
2. Microsoft Kinect.
3. Распознавание жестов на основе технологии IntelRealsense.

В данном разделе необходимо произвести оценку ресурсоэффективности проекта, определяемую посредством расчета интегрального критерия, по следующей формуле:

$$I_{pi} = \sum a_i \cdot b_i, \quad (19)$$

где  $I_{pi}$  – интегральный показатель ресурсоэффективности для  $i$ -го варианта исполнения разработки;

$a_i$  – весовой коэффициент  $i$ -го варианта исполнения разработки;

$b_i^a, b_i^p$  – бальная оценка  $i$ -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания;

$n$  – число параметров сравнения.

Расчет интегрального показателя ресурсоэффективности представлен в форме таблицы (таблица 26).

Таблица 26 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии	Весовой коэффициент параметра	Исп.1	Исп.2	Исп.3
Энергоэффективность	0,2	4	4	5
Удобство в эксплуатации	0,1	5	4	4
Количество распознаваемых жестов	0,1	5	4	3
Помехоустойчивость	0,2	4	4	4
Потребность в вычислительных ресурсах	0,1	3	4	5
Дальность работы	0,05	5	5	3
Способность распознавать частично скрытый жест	0,1	4	4	3
ИТОГО	1	34	32	31

Расчет интегрального показателя для разрабатываемого проекта:

$$I_{p1} = 0,2 \cdot 4 + 0,1 \cdot 5 + 0,1 \cdot 5 + 0,2 \cdot 4 + 0,1 \cdot 3 + 0,05 \cdot 5 + 0,1 \cdot 4 = 3,55$$

$$I_{p2} = 0,2 \cdot 4 + 0,1 \cdot 4 + 0,1 \cdot 4 + 0,2 \cdot 4 + 0,1 \cdot 4 + 0,05 \cdot 5 + 0,1 \cdot 4 = 3,45$$

$$I_{p3} = 0,2 \cdot 5 + 0,1 \cdot 4 + 0,1 \cdot 3 + 0,2 \cdot 4 + 0,1 \cdot 5 + 0,05 \cdot 3 + 0,1 \cdot 3 = 3,45$$

Интегральный показатель эффективности вариантов исполнения разработки ( $I_{исп.1}$ ) определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{исп.1} = \frac{I_{р-исп.1}}{I_{финр}}, \quad (20)$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта (таблицу 27) и выбрать наиболее целесообразный вариант из предложенных. Сравнительная эффективность проекта ( $\mathcal{E}_{ср}$ ):

$$\mathcal{E}_{ср} = \frac{I_{исп.1}}{I_{исп.2}} \quad (21)$$

Таблица 27 – Эффективность разработки

№	Показатели	Исп.1	Исп.2	Исп.3
1	Интегральный финансовый показатель разработки	0,82	0,87	0,96
2	Интегральный показатель ресурсоэффективной разработки	4,4	4,0	3,7
3	Интегральный показатель эффективности	5,3	4,2	3,4
4	Сравнительная эффективность вариантов исполнения	1	0,87	0,74

В результате выполнения изначально сформулированных целей раздела, можно сделать следующие выводы:

1. При проведении планирования был разработан план-график выполнения этапов работ для руководителя и инженера, позволяющий оценить и спланировать рабочее время исполнителей. Были определены: общее количество календарных дней, в течение которых работал инженер и общее количество календарных дней, в течение которых работал руководитель;

2. Составлен бюджет проектирования, позволяющий оценить затраты на реализацию проекта, которые составляют 324712,68 руб;

3. По факту оценки эффективности ИР, можно сделать выводы:

- Значение интегрального финансового показателя ИР составляет

0,82, что является показателем того, что ИР является финансово выгодной, по сравнению с аналогами;

- Значение интегрального показателя ресурсоэффективности ИР составляет 4,4, по сравнению с 4,0 и 3,7;

Значение интегрального показателя эффективности ИР составляет 5,2, по сравнению с 4,2 и 3,4, и является наиболее высоким, что означает, что техническое решение, рассматриваемое в ИР, является наиболее эффективным вариантом исполнения.

## **6 Социальная ответственность**

Целью данной выпускной квалификационной работы является обученная нейронная сеть Mask-RCNN, способная распознавать жесты рук, а также выделять и классифицировать их в кадре.

В данном разделе ВКР исследованы меры по защите разработчика от возможного негативного воздействия среды, а также вредные и опасные факторы среды. Кроме того, были рассмотрены возможные чрезвычайные ситуации и действия, которые разработчик должен выполнить в случае возникновения чрезвычайной ситуации.

Были выделены и рассмотрены такие факторы, воздействующие на разработчика устройства, как: освещение, микроклимат, электромагнитное излучение, нервно-психические перегрузки. К опасным факторам при работе с персональным компьютером можно отнести высокое напряжение и возможность короткого замыкания, влекущего за собой опасность поражения разработчика электрическим током. Рассматриваются вопросы правового регулирования трудовых отношений, связанных с использованием разработанной системы.

Программные продукты не оказывают непосредственного негативного влияния на окружающую среду, однако их использование сопряжено с использованием ЭВМ, что оказывает негативное влияние на литосферу при их утилизации.

### **6.1 Правовые и организационные вопросы обеспечения безопасности**

Подготовка данных для обучения, и само обучение нейронной сети происходило при помощи ПК. При работе с ПК обязаны быть выполнены требования к: помещению; освещению; организации медицинского обследования пользователей.

Для студента-магистранта необходимо указать особенности трудового законодательства, а именно режим рабочего времени и защита персональных данных работника.

Согласно ТК РФ, 197 -ФЗ каждый работник имеет право:

- рабочее место, соответствующее требованиям охраны труда;
- обязательное социальное страхование от несчастных случаев на производстве и профессиональных заболеваний в соответствии с федеральным законом;
- получение достоверной информации от работодателя, соответствующих государственных органов и общественных организаций об условиях и охране труда на рабочем месте, о существующем риске повреждения здоровья, а также о мерах по защите от воздействия вредных и (или) опасных производственных факторов;
- отказ от выполнения работ в случае возникновения опасности для его жизни и здоровья вследствие нарушения требований охраны труда, за исключением случаев, предусмотренных федеральными законами, до устранения такой опасности;
- обеспечение средствами индивидуальной и коллективной защиты в соответствии с требованиями охраны труда за счет средств работодателя и т.д. Рабочее место должно соответствовать требованиям ГОСТ 12.2.032-78. Рабочий стол должен быть устойчивым, иметь однотонное неметаллическое покрытие, не обладающее способностью накапливать статическое электричество; рабочий стул должен иметь дизайн, исключающий онемение тела из-за нарушения кровообращения при продолжительной работе на рабочем месте; рабочее место должно соответствовать техническим требованиям и санитарным нормам. [21]

Устройство рабочего стола должна быть использовано для оптимального размещения используемого оборудования. Кроме того, форма рабочего стола должна быть удобна для поддержания рациональной позы пользователя, так что бы он мог менять положения своего тела для предупреждения утомления. На рисунках 38 и 39 приведена возможная схема рабочего места сотрудника. [22]

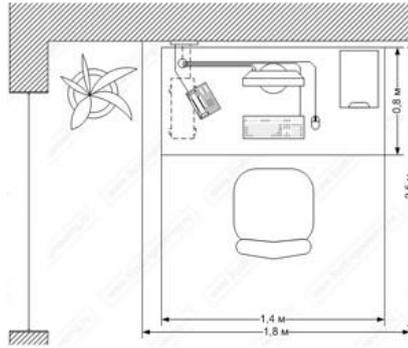


Рисунок 38 – Схема рабочего места (вид сверху)

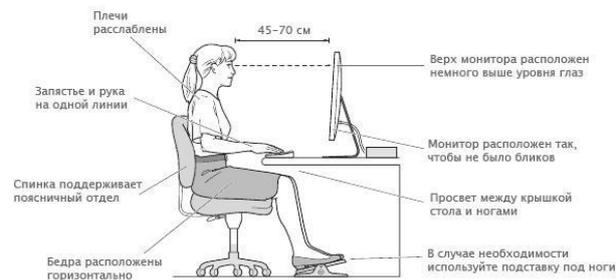


Рисунок 39 – Схема рабочего места (вид сбоку)

В соответствии с СН-245-71 в помещении должен быть организован воздухообмен.

По ТК РФ Работнику в течение рабочего дня предоставляется перерыв не более двух часов и не менее 30 минут, который в рабочее время не включается. Всем работникам предоставляются выходные дни, работа в выходные дни производится только с письменного согласия работника.

## 6.2 Производственная безопасность

Вредные и опасные факторы, воздействующие на сотрудника, устанавливаются согласно ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация». [23]

Опасные и вредные факторы при выполнении работ:

Таблица 28 – Вредные и опасные факторы

Источник фактора, наименование видов работ	Факторы (по ГОСТ 12.0.003-2015)		Нормативные документы
	Вредные	Опасные	
Подготовка обучающих данных, обучение и оценка работы нейронной сети Mask-RCNN	1. Отклонение показателей микроклимата от нормы; 2. Превышение уровня шума; 3. Отсутствие или недостаток освещения; 4. Вибрации; 5. Психофизические факторы (монотонность труда, нервно-психические перегрузки, перенапряжение зрительных анализаторов); 6. Электромагнитные излучения.	1. Поражение электрическим током; 2. Короткое замыкание; 3. Статическое электричество.	ГОСТ 12.0.003-2015 Опасные и вредные производственные факторы. Классификация. [23] СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания. [24] Микроклимат: СанПиН 2.2.4.548-96 [25] Шум, санитарные нормы: СН 2.2.4/ 2.1.8.562-96 [26] СП 51.13330.2011 [27]  Освещение: СНиП 23-05-95* [28][29][30] Вибрации: ГОСТ 12.1.012-2004 [31] Электрический ток, короткое замыкание: ГОСТ 12.1.033-81 [32] ГОСТ 12.1.038-82 ССБТ [33]

### 6.3 Анализ вредных производственных факторов

#### 6.3.1 Отклонения показателей микроклимата

В производственных помещениях, в которых работа с использованием

ПЭВМ является основной (диспетчерские, операторские, расчетные, кабины и посты управления, залы вычислительной техники и др.) и связана с нервно-эмоциональным напряжением, должны обеспечиваться оптимальные параметры микроклимата для категории работ 1а и 1б в соответствии с действующими санитарно-эпидемиологическими нормативами микроклимата производственных помещений. На других рабочих местах следует поддерживать параметры микроклимата на допустимом уровне, соответствующем требованиям указанных выше нормативов.

Уставленные гигиенические нормативы для помещений с ВДТ и ПЭВМ для категории работы 1б приведены в таблице 29. [24]

Таблица 29 – Оптимальные величины показателей микроклимата на рабочих местах производственных помещений (СанПиН 2.2.4.548-96 [25]).

Период года	Категория работ по уровню энергозатрат	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	1б (140-174 Вт)	21-23	20-24	60-40	0.1
Теплый		22-24	21-25	60-40	0.1

Работа оператора ПК относится к категории 1б, поскольку работа связана с минимальными физическими нагрузками. Также нужно учитывать нормы подачи свежего воздуха в помещении с ПК (таблица 30).

Таблица 30 – Нормы подачи свежего воздуха в помещении с ПК

Объем помещения, м <sup>3</sup>	Объемный расход подаваемого в помещение свежего воздуха, м <sup>3</sup> / на одного человека в час
До 20	Не менее 30
20 – 40	Не менее 20
Более 40	Естественная вентиляция

### 6.3.2 Превышение уровня шума

В производственных помещениях при выполнении основных или вспомогательных работ с использованием ПЭВМ уровни шума на рабочих местах не должны превышать предельно допустимых значений, установленных для данных видов работ в соответствии с действующими санитарно-эпидемиологическими нормативами.

Шумовой фон помещения создают десять одновременно работающих компьютеров. Также возникает шум, исходящий от принтера или телефонных аппаратов. Также источником шума является система вентиляции или шумы, поступающие извне помещения. Во избежание негативных последствий от производственного шума, его необходимо регулировать в соответствии с нормами, которые указаны в СН 2.2.4/ 2.1.8.562-96 Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки. Санитарные нормы. Нормативным эквивалентным уровнем звука на рабочих местах, является 80 дБА [26].

В СанПиН СН 2.2.4/ 2.1.8.562-96 установлены допустимые значения уровней звукового давления, создаваемого ПЭВМ (таблица 31).

Таблица 31 – Допустимые значения уровней звукового давления в октавных полосах частот и уровня звука, создаваемого ПЭВМ [27]

Уровни звукового давления в октавных полосах со среднегеометрическими частотами								
31,5 Гц	63 Гц	125 Гц	250 Гц	500 Гц	1000 Гц	2000 Гц	4000 Гц	8000 Гц
86 дБ	71 дБ	61 дБ	54 дБ	49 дБ	45 дБ	42 дБ	40 дБ	38 дБ

### 6.3.3 Отсутствие или недостаток освещения

Отсутствие хорошего освещения может привести к профессиональным заболеваниям, а также ухудшению концентрации работников. Работа инженера-программиста в основном проводится за дисплеем персонального компьютера, что вынуждает его работать с контрастным фоном, в случае недостаточной

освещённости рабочего места. В результате у работника может ухудшиться зрения, а также возникнуть переутомление. Тоже самое происходит и при избыточном освещении помещения.

В данном случае к негативным факторам относятся повышенные уровни излучения, а также увеличенная нагрузка на зрительные органы. Требования к освещению установлены в СНиП 23-05-95\* (таблица 32) [28].

Таблица 32 – Требования к освещению на рабочих местах, оборудованных ПЭВМ

Освещенность на рабочем столе	300-500 лк
Освещенность на экране ПЭВМ	Не выше 300 лк
Блики на экране	Не выше 40 кд/м <sup>2</sup>
Прямая блесккость источника света	200 кд/м <sup>2</sup>
Показатель ослепленности	Не более 20
Показатель дискомфорта	Не более 15
Отношение яркости между рабочими поверхностями	3:1-5:1 10:1
Коэффициент пульсации	Не более 5%

Для выявления соответствия лабораторной аудитории №103 10 корпуса ТПУ требованиям СНиП 23-05-95\* далее будут приведены расчеты искусственного освещения рабочей зоны работника согласно методическим указаниям к выполнению лабораторной работы «Расчет искусственного освещения» для студентов дневного и заочного обучения всех направлений и специальностей ТПУ [29].

Параметры аудитории №103 10 корпус ТПУ:

- площадь  $S = 44,9 \text{ м}^2$ ;
- длина  $A = 9,2 \text{ м}$ ;
- ширина  $B = 4,88 \text{ м}$ ;
- высота потолка  $H = 2,5 \text{ м}$ ;

- высота рабочей поверхности  $h_{rp} = 0,8$  м.;
- количество светильников  $n = 14$ ;
- тип светильников ЛВО 4×18 CSVТ люминесцентными лампами типа L18W/640 с потоком  $F = 1200$  лм.

Согласно отраслевым нормам освещенности уровень рабочей поверхности над полом составляет 0,8 м., а установлена минимальная норма освещенности  $E = 300$  лк [10].

Расчетная высота подвеса светильников над рабочей поверхностью ( $h$ ) определяется по формуле (22):

$$h = H - h_p - h_c \quad (22)$$

где  $H$  – высота потолка в помещении, м;

$h_p$  – расстояние от пола до рабочей поверхности стола, м;  $h_c$  – расстояние от потолка до светильника, м;

Так как светильники установлены в уровень с потолком, примем  $h_c$  равным 0. Таким образом, согласно формуле (1), высота подвеса светильников над рабочей поверхностью для лабораторной аудитории №103 корпуса №10 ТПУ равна:

$$h = 2,5 - 0,8 = 1,7 \text{ м}$$

Индекс помещения определяется по формуле (23) [26]:

$$i = \frac{S}{h \cdot (A + B)} \quad (23)$$

где  $S$  – площадь помещения, м<sup>2</sup>;  $A$  – длина комнаты, м;

$B$  – ширина комнаты, м;

$h$  – высота подвеса светильников, м.

Индекс помещения для лабораторной аудитории №101 10 корпус ТПУ:

$$i = \frac{44,9}{1,7 \cdot (9,2 + 4,88)} = 1,776$$

Исходя из того, что потолок в аудитории представлен светлыми потолочными панелями, поверхность стен окрашена белой краской и имеется два окна без штор, примем коэффициенты отражения от стен  $\rho_c = 50\%$  и потолка  $\rho_n = 50\%$ .

По таблице коэффициентов использования светового потока из пособия к МГСН 2.06-99 для соответствующих значений  $i$ ,  $\rho_c$ ,  $\rho_n$  примем  $\eta = 0,36$  (для потолочного светильника любого типа при  $i = 1,75$ ) [28].

Освещенность помещения рассчитывается по формуле (24):

$$E_{\Phi} = \frac{n \times \eta \times \Phi}{S \times k_3} \quad (24)$$

где  $\Phi$  – световой поток светильника, лм;

$S$  – площадь помещения,  $m^2$ ;

$k_3$  – коэффициент неравномерности освещения;

$n$  – число светильников;

$\eta$  – коэффициент использования светового потока.

Коэффициент запаса  $k$  учитывает запыленность светильников и их износ. Для помещений с малым выделением пыли  $k = 1,5$ . Поправочный коэффициент  $z$  это коэффициент неравномерности освещения. Для люминесцентных ламп  $z = 1,1$ . Каждый светильник состоит из 4 ламп.

Учитывая все параметры, освещенность аудитории №101 равна:

$$E_{\Phi} = \frac{14 \times 4 \times 0,36 \times 1200}{44,9 \times 1,5 \times 1,1} = 326,5 \text{ лк}$$

Освещенность компьютерной аудитории №103 корпуса №10 ТПУ удовлетворяет требованиям СНиП 23-05-95\*.

#### **6.3.4 Вибрации**

Шум и вибрация ухудшают условия труда, оказывая вредное воздействие на организм человека. Из-за шума работники испытывают более быстрое утомление, что приводит к снижению производительности на 10...15%,

увеличению количества ошибок при выполнении операций трудового процесса и, следовательно, к повышенному риску получения травмы.

Основные параметры вибрации: частота и амплитуда колебаний. Пороговая частота колебаний составляет 18 Гц, на более низкой частоте вибрация воспринимается как отдельные удары. Верхний порог частоты воспринимаемых колебаний находится на уровне 1500 Гц. При дальнейшем увеличении частоты вибрации возникает ощущение равномерного прикосновения определенной силы.

Допустимые уровни виброскорости по ГОСТ 12.1.012-2004 «ССБТ. Вибрационная безопасность. Общие требования.» приведены в таблице 33.

Таблица 33 – Гигиенические нормы уровней виброскорости (ГОСТ 12.1.012-2004) [31]

Вид вибрации (Технологическая)	Допустимый уровень виброскорости, дБ, в октавных полосах со среднегеометрическими частотами, Гц										
	1	2	4	8	16	31,5	63	125	250	500	1000
	–	108	99	93	92	92	92	92	–	–	–

### 6.3.5 Психофизиологические факторы

Работа с ПК сопряжена с воздействием вредных психофизиологических факторов, в частности, нервно-психических перегрузок. Для снижения воздействия вредных факторов, устанавливаются перерывы в работе для отдыха сотрудников. Суммарное время регламентированных перерывов при работе с ПК зависит от категории трудовой деятельности и уровня нагрузки за рабочую смену. В таблице 34 приведено суммарное время отдыха для каждой категории работ [34].

Таблица 34 – Суммарное время перерывов в зависимости от категории работы и нагрузки

Категория работы с ПЭВМ	Уровень нагрузки за рабочую смену при видах работ с ПЭВМ			Суммарное время регламентированных перерывов при 8-часовой смене, мин.
	группа А, количество знаков	группа Б, количество знаков	группа В, часов	
I	до 20 000	до 15 000	до 2	50
II	до 40 000	до 30 000	до 4	70
III	до 60 000	до 40 000	до 6	90

В данном случае уровень нагрузки относится к группе В, категория работы III. Согласно таблице, требуется установить перерывы, сумма которых за смену составит не менее 90 минут. По типовой инструкции по охране труда при работе на персональном компьютере ТОИ Р-45-084-01 для данной категории работ требуется установить перерывы по 15 минут каждый трудовой час.

### 6.3.6 Перенапряжение зрительных анализаторов

При работе за ПК оператор читает текст в прямом свете, его глаза смотрят на источник света.

При работе за дисплеем возникает следующая проблема: экран имеет высокую фоновую яркость, при ее понижении уменьшается контрастность изображения, что создает проблему для считывания информации. Поэтому пользователю приходится повышать яркость, что приводит к увеличению интенсивности вредных излучений и к утомлению глаз.

В настоящее время при работе за ПК визуальные параметры мониторов оказывают большую нагрузку на зрительные анализаторы, а именно на мышцы глаз. В результате чего мозгу сложно идентифицировать изображение. В связи с этим параметры безопасности компьютерной техники указаны в

соответствующих нормативных документах.

После длительной нагрузки необходим период восстановления и отдыха.

## **6.4 Анализ опасных производственных факторов**

### **6.4.1 Поражение электрическим током**

К опасностям использования электрического тока относятся возможность поражения электрическим током, а также воспламенения электронных устройств из-за воздействия различных условий – попадания влаги или нарушения изоляции

Требования безопасности при эксплуатации электрооборудования регламентируются следующими нормативными актами:

- Правилами устройства электроустановок (издание шестое с отдельными разделами и главами в издании седьмом), утвержденными Главтехуправлением, Госэнергонадзором Минэнерго СССР 05.10.1979 г.;

- Правилами технической эксплуатации электроустановок потребителей, утверждёнными Приказом Минэнерго России от 13.01.2003 г. №6;

- Межотраслевыми правилами охраны труда (правилами безопасности) при эксплуатации электроустановок (ПОТ РМ 016-2001), утвержденными Постановлением Минтруда России от 05.01.2001 г. №3.

Согласно им:

- электрооборудование, имеющее контакты для подключения заземления, должно быть заземлено, а помещения, где размещаются рабочие места с ПЭВМ, должны быть оборудованы защитным заземлением (занулением) в соответствии с техническими требованиями по эксплуатации оборудования;

- все крышки и защитные панели должны находиться на своих местах (при отсутствии крышки или защитной панели эксплуатация электрооборудования не допускается);

- при работе с электрооборудованием не допускать попадания влаги на поверхность электрооборудования, а также запрещается работать на электрооборудовании влажными руками;
- вентиляционные отверстия электрооборудования не должны быть перекрыты находящимися вплотную стенами, мебелью, посторонними предметами;
- выдергивание штепсельной вилки электроприбора необходимо осуществлять за корпус штепсельной вилки, при необходимости придерживая другой рукой корпус штепсельной розетки;
- подключение и отключение разъемов компьютеров и оргтехники должно производиться при отключенном питании (за исключением подключения и отключения USB-устройств);
- удаление пыли с электрооборудования должно производиться в отключенном от электрической цепи состоянии;
- перед использованием электроприборов необходимо проверить надёжность крепления электророзетки, свериться с номиналом используемого напряжения;
- корпуса штепсельных розеток и выключателей не должны содержать трещин, оплавлений и других дефектов, способных снизить защитные свойства или нарушить надёжность контакта;
- кабели (шнуры) электропитания не должны содержать повреждений изоляции, сильных изгибов и скручиваний. [32]

#### **6.4.2 Статическое электричество**

В помещениях, оборудованных ПЭВМ, токи статического электричества чаще всего возникают при прикосновении персонала к любому из элементов ПЭВМ. Такие разряды опасности для человека не представляют, однако кроме неприятных ощущений могут привести к выходу оборудования из строя.

Для предотвращения образования и защиты от статического

электричества в помещении используются нейтрализаторы и увлажнители, а полы имеют антистатическое покрытие в виде поливинилхлоридного антистатического линолеума.

Также в СанПиН 2.2.2/2.4.1340-03 установлен максимальный допустимый электростатический потенциал экрана видеомонитора – 500 В. [30]

В качестве мер уменьшения влияния вредных факторов на пользователя используются защитные фильтры для мониторов, увлажнители воздуха. Должны использоваться розетки с заземлением. Требуется проводить регулярную влажную уборку.

### **6.5 Экологическая безопасность**

Обращение с отходами регламентируется ГОСТ Р 53692-2009 «Ресурсосбережение. Обращение с отходами.» [34] Поэтому, при необходимости утилизировать вышедшую из употребления электронику наиболее безопасным для окружающей среды способом необходимо обращаться в специализированную компанию по утилизации. Такие компании действуют на всей территории Российской Федерации, в том числе и в Томской области. Необходимо отметить, что в целом при работе с компьютером существенного загрязнения окружающей среды не происходит и вредные выбросы не сравнимы с производственными.

Также существуют компании, занимающиеся утилизацией энергосберегающих ламп. Причина опасности данных ламп заключается в наличии ртути в их составе. Специализированные компании занимаются демеркуризацией и утилизацией ртутных отходов. Данные компании имеют специальные лицензии на сбор, использование, транспортировку отходов. В случае выхода из строя используемой электроники или ламп, отходы передаются в соответствующие компании [35].

К отходам, производимым в помещении, можно отнести бумажные отходы, то есть макулатуру. Бумажные отходы рекомендуется накапливать и

передавать их в пункты приема макулатуры для дальнейшей переработки. Дополнительным методом снижения отходов является увеличение доли электронного документооборота.

### **6.6 Безопасность в чрезвычайных ситуациях**

Наиболее вероятной чрезвычайной ситуацией при обучении нейронной сети является пожар на рабочем месте. В качестве противопожарных мероприятий должны быть применены следующие меры:

- в помещении должны находиться средства тушения пожара, средства связи;
- электрическая проводка электрооборудования и осветительных приборов должна быть исправна;
- все сотрудники должны знать место нахождения средств пожаротушения и уметь ими воспользоваться, средств связи и номера экстренных служб.

В связи с возможностью возникновения пожара разработан следующий план действий:

- в случае возникновения пожара сообщить о нем руководителю, постараться устранить очаг возгорания имеющимися силами при помощи первичных средств пожаротушения (огнетушитель порошковый, углекислотный О-1П0 (з)-АВСЕ);
- привести в действие ручной пожарный извещатель, если очаг возгорания потушить не удастся;
- сообщить о возгорании в службу пожарной охраны по телефону 01 или 010, сообщить адрес, место и причину возникновения пожара;
- принять меры по эвакуации людей и материальных ценностей;
- встретить пожарную охрану, при необходимости сообщить всю необходимую информацию и оказать помощь при выборе наилучшего подхода к очагу возгорания.

Возникновение пожара в помещении, где установлена вычислительная и оргтехника, приводит к большим материальным потерям и возникновению чрезвычайной ситуации. Чрезвычайные ситуации приводят к полной потере информации и большим трудностям восстановления всей информации в полном объёме.

Согласно нормам технологического проектирования [36], в зависимости от характеристики используемых в производстве веществ и их количества, по пожарной и взрывной опасности помещения подразделяются на категории А, Б, В, Г, Д.

В случае возникновения пожара необходимо отключить электропитание, вызвать по телефону пожарную команду, произвести эвакуацию и приступить к ликвидации пожара огнетушителями. При наличии небольшого очага пламени можно воспользоваться подручными средствами с целью прекращения доступа воздуха к объекту возгорания. Для тушения пожаров в помещении необходимо установить углекислотный огнетушитель типа ОУ-5. Покидать помещение необходимо согласно плану эвакуации, на рисунке 40.

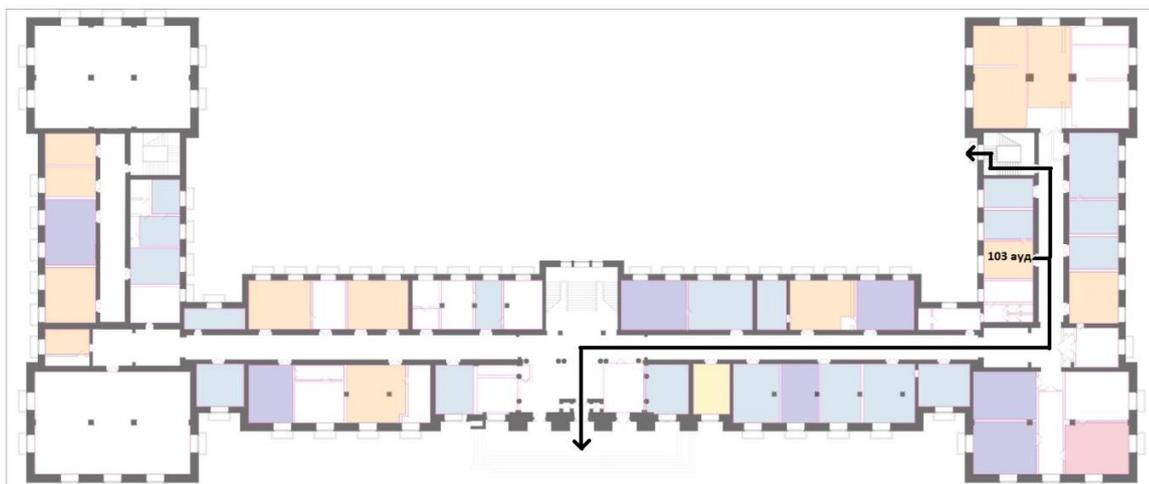


Рисунок 40– План эвакуации при пожаре и других ЧС

Рабочее помещение оборудовано в соответствии с требованиями пожарной безопасности. Имеется порошковый огнетушитель, а также пожарная сигнализация и средства связи.

### **Заключение к главе “Социальная ответственность”**

В ходе выполнения работы над разделом “Социальная ответственность” были выявлены опасные и вредные факторы, воздействию которых может подвергнуться человек, использующий Mask-RCNN для распознавания жестов рук.

Был проведён анализ нормативной документации. В целом, рабочее место удовлетворяет требованиям безопасности.

Освещение на рабочем месте соответствует нормам – используется несколько энергосберегающих ламп.

Уровни шума находятся в допустимых пределах – источником шума при эксплуатации ПК могут являться системы охлаждения, а также жесткий диск, однако уровень создаваемого ими шума невысок.

Микроклиматические условия соблюдаются за счет использования систем отопления и кондиционирования.

Защита от повреждений электроники статическим электричеством не обеспечивается, однако так как корпус ПК закрыт, вероятность поражения элементов или работника минимальна (если не прикасаться мокрыми руками к корпусу).

Во время работы делаются перерывы для снижения нагрузки и предотвращения нервно-психических перегрузок.

Помещение оборудовано согласно требованиям электробезопасности.

В случае выхода из строя используемой электроники или ламп, отходы передаются в соответствующие компании.

Рабочее помещение оборудовано в соответствии с требованиями пожарной безопасности. Имеется порошковый огнетушитель и пожарная сигнализация.

## **Заключение**

Результатом данной выпускной квалификационной работы магистра является обученная нейронная сеть Mask-RCNN для решения задачи распознавания жестов рук. Была проанализирована работа нейронной сети на синтезированном и реальном наборе данных.

В процессе разработки было выполнено следующее:

- изучена структура нейронной сети Mask-RCNN;
- собран обучающий набор данных в размере 300 изображений для обучения, поделенный на 5 класса: “Кисть”, “Победа”, “Кулак”, “Большой палец” и “Телефон”;
- получены навыки работы с библиотеками TensorFlow, PIL, OpenCV и Keras;
- получены навыки обучения нейронных сетей;

В результате работы обучена свёрточная нейронная сеть, распознающая жесты с точностью 90 %. Этой точности не всегда достаточно для успешного распознавания, из-за того, что нейронная сеть путает жесты из-за общей схожести. В дальнейшем планируется увеличить выборку данных для повышения точности работы нейронной сети.

Также в ходе работы была исследована зависимость наиболее успешного распознавания жеста в зависимости от расстояния до камеры. Экспериментальным путем было установлено, что наиболее высокую точность можно получить при расположении жеста на расстоянии 60 сантиметров от камеры. При таком расположении, общая точность стремится к значению 90%.

Помимо этого, исследована работа нейронной сети на способность распознавания каждого жеста. Наименее точно Mask-RCNN распознает жесты Fist и Big Finger из-за их общей схожести. Остальные жесты имеют вероятность распознавания выше 83%.

## Список используемых источников литературы

1. Digital Data Entry Glove. [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: <http://www.idemployee.id.tue.nl/g.w.m.rauterberg/presentations/HCI-history/sld066.htm> (Дата обращения: 05.03.2020).
2. Sayre Glove Final Project Report. [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: <https://www.evl.uic.edu/entry.php?id=2296> (Дата обращения: 05.03.2020).
3. Wired glove. [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: [https://en.wikipedia.org/wiki/Wired\\_glove](https://en.wikipedia.org/wiki/Wired_glove) (Дата обращения: 05.03.2020).
4. Interfacing the Nintendo Power Glove to a Macintosh Computer [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: <https://pdfs.semanticscholar.org/9555/ac04a85a8a5fe59ce932f47fe7e5d8d13af5.pdf> (Дата обращения: 06.03.2020).
5. Gesture recognition using colored gloves [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: <https://www.semanticscholar.org/paper/Gesture-recognition-using-colored-gloves-Iwai-Watanabe/03d1b16cbcc133d60bea4a2038662412ccabb3b9> (Дата обращения: 07.03.2020).
6. Высокоточное распознавание жестов по облаку точек или цветному изображению (погрешность до 5 мм) [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: <https://rubius.com/portfolio/ru/#eye> (Дата обращения: 16.03.2020).
7. Heterogeneous hand gesture recognition using 3D dynamic skeletal data Q. De Smedt, H. Wannous and J.-P. Vandeborre/ Computer Vision and Image Understanding 181 (2019) 60–72.
8. Dynamic gesture recognition by directional pulse coupled neural networks for human-robot interaction in real time/ J. Dong et al. / J. Vis. Commun. Image R. 63 (2019) 102583.

9. Real-time gesture recognition based on feature recalibration network with multi-scale information. Z. Cao, X. Xu and B. Hu et al. / *Neurocomputing* 347 (2019) 119–130.
10. Hand gesture recognition based on motor unit spike trains decoded from high-density electromyography. C. Chen, Y. Yu, S. Ma et al. / *Biomedical Signal Processing and Control* 55 (2020) 101637.
11. Human hand gesture based system for mouse cursor control. Horatiu-Stefan Grif et al. / *Procedia Manufacturing* 22 (2018) 1038–1042.
12. Альтшуллер Г. Как научиться изобретать. - Тамбов: Тамбовское книжное издательство, 1961. - С. 44,55,82.
13. Обзор компьютерного зрения [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: <https://www.machinelearningmastery.ru/an-overview-of-computer-vision-1f75c2ab1b66/> (Дата обращения: 16.03.2020).
14. Mask R-CNN: архитектура современной нейронной сети для сегментации объектов на изображениях [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: <https://habr.com/ru/post/421299/> (Дата обращения: 16.03.2020).
15. Булыгин Д.А., Мамонова Т.Е. Распознавание жестов рук в режиме реального времени. *Вестник НГТУ №78* (2020), с. 25 – 40.
16. Convolutional Neural Networks Backbones for Object Detection / *ICISP 2020: Image and Signal Processing* pp 282-289
17. Computer Vision Tutorial: Implementing Mask R-CNN for Image Segmentation [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: [https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/?source=post\\_page-----](https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/?source=post_page-----) (Дата обращения: 16.03.2020).
18. Ross Girshick. “Fast R-CNN.” In *Proc. IEEE Intl. Conf. on computer vision*, pp. 1440-1448. 2015.

19. Bounding box prediction. [Электронный ресурс]. Режим доступа: свободный. Ссылка на ресурс: <https://datascience.stackexchange.com/questions/30557/how-does-the-bounding-box-regressor-work-in-fast-r-cnn> (Дата обращения: 05.03.2021).

20. Ross Girshick. “Mask R-CNN.” In Proc. IEEE Intl. Conf. on computer vision, 2018.

21. Трудовой кодекс Российской Федерации от 30.12.2001 N 197–ФЗ.

22. ГОСТ 12.2.032-78 Рабочее место при выполнении работ сидя.

23. ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация.

24. СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания.

25. СанПиН 2.2.4.548 – 96. Гигиенические требования к микроклимату производственных помещений. М.: Минздрав России, 1997.

26. СН 2.2.4/ 2.1.8.562-96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки. Санитарные нормы.

27. СП 51.13330.2011. Свод правил. Защита от шума.

28. СНиП 23-05-95\*. Естественное и искусственное освещение.

29. Безопасность жизнедеятельности. Расчет искусственного освещения.

Методические указания к выполнению индивидуальных заданий для студентов дневного и заочного обучения всех направлений и специальностей ТПУ. – Томск: Издательство Томского политехнического университета, 2008.– 20 с.

30. СП 52.13330.2016 Свод правил. Естественное и искусственное освещение.

31. ГОСТ 12.1.012-2004 Система стандартов безопасности труда (ССБТ). Вибрационная безопасность. Общие требования.

32. ГОСТ 12.1.033-81. Система стандартов безопасности труда. Пожарная безопасность.

33. ГОСТ 12.1.038-82. Система стандартов безопасности труда. Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов.

34. ГОСТ Р 53692-2009. Ресурсосбережение. Обращение с отходами.

35. ГОСТ 17.4.3.04-85 «Охрана природы. Почвы. Общие требования к контролю и охране от загрязнения».

36. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования.

### **Список научных публикаций студента**

1. Булыгин Д.А., Мамонова Т.Е. Распознавание жестов рук в режиме реального времени. Вестник НГТУ №78 (2020), с. 25 – 40.
2. Булыгин Д.А., Мамонова Т.Е. Применение различных типов изображений для обучения сверточной нейронной сети. Молодежь и современные информационные технологии: сборник трудов XVII Международной научно-практической конференции студентов, аспирантов и молодых ученых (Томск, 17–20 февраля 2020 г.) / Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2020, с. 208 – 209.
3. Булыгин Д.А., Мамонова Т.Е. Решение задачи распознавания жестов рук на основе нейронной сети Mask-RCNN. Молодежь и современные информационные технологии: сборник трудов XVIII Международной научно-практической конференции студентов, аспирантов и молодых ученых (Томск, 22–26 марта 2021 г.) / Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2021, с. 251 – 252.

**Приложение А**  
(справочное)

**History of inventions for gesture recognition;**  
**Devices being developed in our time;**  
**Analysis of the methods for solving the gesture recognition problem;**  
**Application of the Inventive Problem-Solving Algorithm – 61**

Студент

Группа	ФИО	Подпись	Дата
8ЕМ91	Булыгин Дмитрий Александрович		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Мамонова Татьяна Егоровна	к.т.н. доцент		

Консультант-лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель ОИЯ ШБИП	Пичугова Инна Леонидовна	—		

## **Introduction**

Today, more and more researches aim at solving problems with computer vision and artificial intelligence. The most common are solutions and approaches using gesture recognition based on infrared sensors or neural networks.

The relevance of gesture recognition using video cameras is due to the possibility of using the proposed approach to control the operation of objects without tactile contact and voice identification of commands, as well as its simplicity from the point of view of the end user.

Recently, artificial intelligence has received a powerful impetus in its development. One of the directions is the processing and extraction information from images. Convolutional neural networks have been specially developed for this task. The work will be developed on this architecture.

The main goal of this work is to train the special architecture Mask-RCNN, capable of simultaneously recognizing several gestures of a person's hands without additional devices.

This type of control has many advantages over control using voice commands or control using buttons. They are as follows:

- 1) ease to use;
- 2) lack of physical contact.

This device has many uses:

- 1) Command of mobile robots. Multiple gesture recognition in the frame can allow using multiple commands. It is also possible to combine hand gestures to increase the number of possible commands.

- 2) Development for the implementation in medical institutions: polyclinics and hospitals. Germs and diseases are often transmitted to healthy person from a sick person through physical contact by surfaces. In many cases, these surfaces are doors. It is also worth noting that ordinary sensors located in the doors of shops are not a solution to the problem, since they open the doors whenever a person approaches, which is extremely inconvenient in a polyclinic.

In order to reduce the risk of disease, the task is to develop a contactless way of interacting with objects and devices.

### **History of inventions for gesture recognitions**

The history of hand gesture recognition for computer control began with the invention of glove-based control interfaces. The researchers realized that gestures could be used as simple commands to interact with a computer. This direction has evolved with the development of very accurate accelerometers, infrared cameras and even fiber optic bend sensors (optical goniometers).

The first prototypes of the gloves were the Sayre Glove, the MIT LIT Glove, and the Digital Input Data Glove. The Sayre Glove, which was developed in 1977, used flexible tubes with a light source at one end and a photocell at the other, which were installed along each finger of the glove. Flexing the fingers reduced the amount of light passing between the LED and the photodiode. Thus, the system detected the degree of bending of the fingers using the voltage measured by the photodiode.

During the 1980s, sensor technology developed at a rapid pace, in part due to fears of the Cold War and the natural expansion of industry in many European countries. These sensor technologies paved the way for the rapid development of computer technology and peripherals. Many leading research groups around the world created new market-oriented computer peripherals.

The first commercially available Data Glove appeared in 1987. This was an improved version of the first DataGlove developed by Zimmerman in 1982, which is shown in Figure 1.



Figure 1 – Data Glove

The technology was similar to that used by Sayre Glove in 1977. However, in the 1987 version, fiber optics was used instead of light tubes.

In addition, the glove was equipped with 5-15 sensors, increasing its ability to distinguish between different gestures. The numerous sensors available on DataGlove made it popular with researchers in various fields, and many similar devices were developed. The Data Glove inspired the development of the Power Glove, which was sold by Mattel Intellivision as a controller for the Nintendo game console in 1989.

Through its developments in the early 1980s, the MIT Data Glove has grown dramatically, offering different capabilities for different models. Currently developed by MIT's AnthroTronix subsidiary, acceleGlove, as shown in Figure 2, is a user-programmable glove that records hand and finger movements in 3D. Other models available from them include 5DT's Data Glove for virtual reality ranging from \$ 1,000 to \$ 5,000. The company originally developed Data Gloves for the US Defense to control robots. Their acceleGlove is also used in video games, sports training, or physical rehabilitation.



Figure 2 – acceleGlove

As it is shown in Figure 2, the accelerometer lies just below each fingertip and on the back of the hand. Accelerometers can determine the three-dimensional orientation of the fingers and palm relative to gravity when a gesture or any movement is made. These measurements are accurate to within a few degrees, allowing the software to distinguish between small changes in hand position. The glove has fingertip openings that allow the user to type or write while wearing the glove.

In 1996, Iwai [5] proposed a method of colored gloves, with 10 areas of the fingers for identification. They used multiple colors to represent different parts of the finger and palm to avoid the occlusion problem. In the event of an occlusion problem, some parts of the hand or fingers are occluded and the camera cannot interpret the gesture accurately. When different color areas represent different areas of the hand (fingers, palm), the system can rely on color and border to make informed decisions. They used a decision tree method to automatically recognize a limited number of gestures.

In recent years, more and more research has focused on vision-based gesture recognition. Compared to non-vision recognition (wired gloves), vision-based recognition is more natural and convenient as it does not restrict the flexibility of hand movements. Based on gloves and electromagnetic waves, a colored Lamberti

glove was developed, which is easy to wear without embarrassing the user. As it is shown in Figure 3, the colored glove contains a separate color for palm tracking, and the fingers are marked with alternative colors.

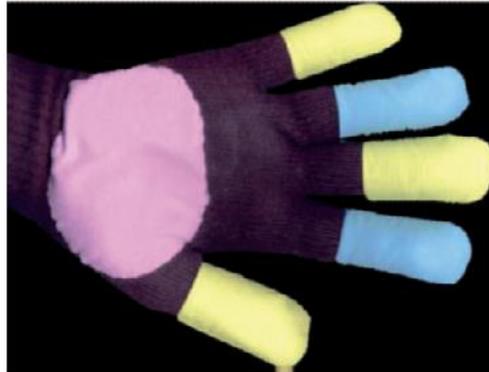


Figure 3 – Lamberti colored glove

The goal of this approach was to develop a low cost approach to a group of data with much more flexibility and very low computational requirements so that users with disabilities can use this technology in a school environment.

Recently, artificial intelligence has received a powerful impetus in its development. One of the directions of its development is to process and extract information from images. Convolutional neural networks have been specially developed for this task.

### **Devices being developed in our time**

#### **1. Gesture recognition using infrared sensors based on Intel RealSense**

French scientists Quentin Smedt, Hasem Vannous and Jean-Philippe Vandeborre proposed the idea of gesture recognition using data on the skeleton of a hand in three-dimensional format.

Intel RealSense device is based on infrared sensors, with the help of which it is possible to obtain information not only from the image of the hand, but also from its position in space. This information allows predicting how far the hand is from the infrared sensor, or the depth.

Figure 4 shows the depth and skeletal information of the hand using Intel RealSense. The sensors return 22 connections, four for each finger, one in the center of the palm, and one on the wrist.

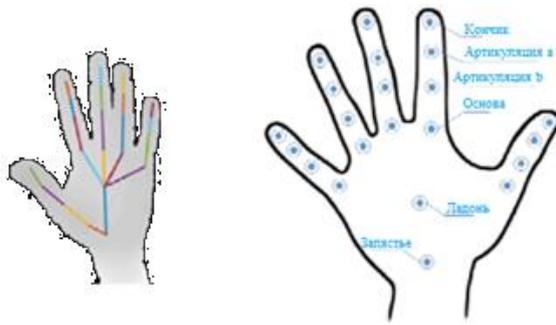


Figure 4 – Information obtained using Intel RealSense

## 2. Dynamic gesture recognition using directional impulse neural networks

Dynamic gesture recognition based on pulse coupled neural networks (PCNN) for human-robot interaction in real time was proposed by Chinese scientists.

It proposes a modified PCNN algorithm called deep PCNN (DPCNN) for real-time gesture recognition without a lot of training data. The 3D coordinates of the key wireframe points of the human body are extracted as an input. Data captured by the Kinect are shown in Figure 5. To recognize an input gesture, its similarity is computed from a matrix of objects created from the input and a given pattern of feature sequences.

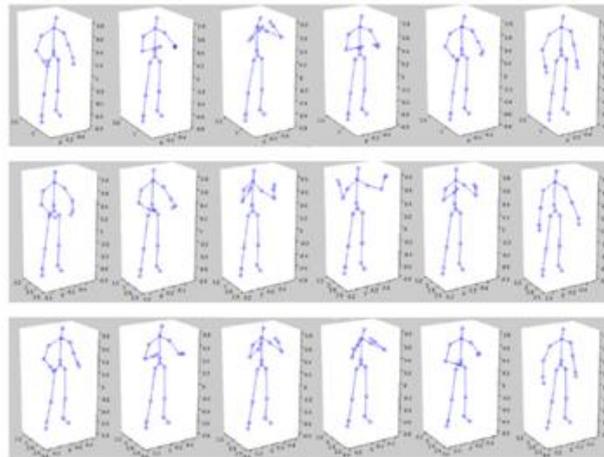


Figure 5 – Sequence of 3D data

### **3. Real-time gesture recognition based on network-based recalibration of functions with multi-scale information**

The later work was also presented by Chinese scientists. They used convolutional neural networks to solve the problem of gesture recognition and classification.

There are two problems with recognizing a gesture far away from the camera. First, gestures with different aspect ratios are difficult to identify. Secondly, there is different information about the signs between the low level signs and the high level signs. Although low-level, high-resolution features contain more detail and position information, this does not help to identify the desired features of different sizes when detected. In contrast, high-level features are more suitable for categorization, but they have a lower perception of detail due to their lower resolution.

The module mainly includes the following strategy: combining features at different scales. In order to better extract contextual information of different scales, information about the characteristics of different scales is extracted using a convolutional kernel with a step size of 2 and sizes of  $3 \times 3$  and  $5 \times 5$ .

The complete structure of a convolutional neural network is depicted in Figure 6.

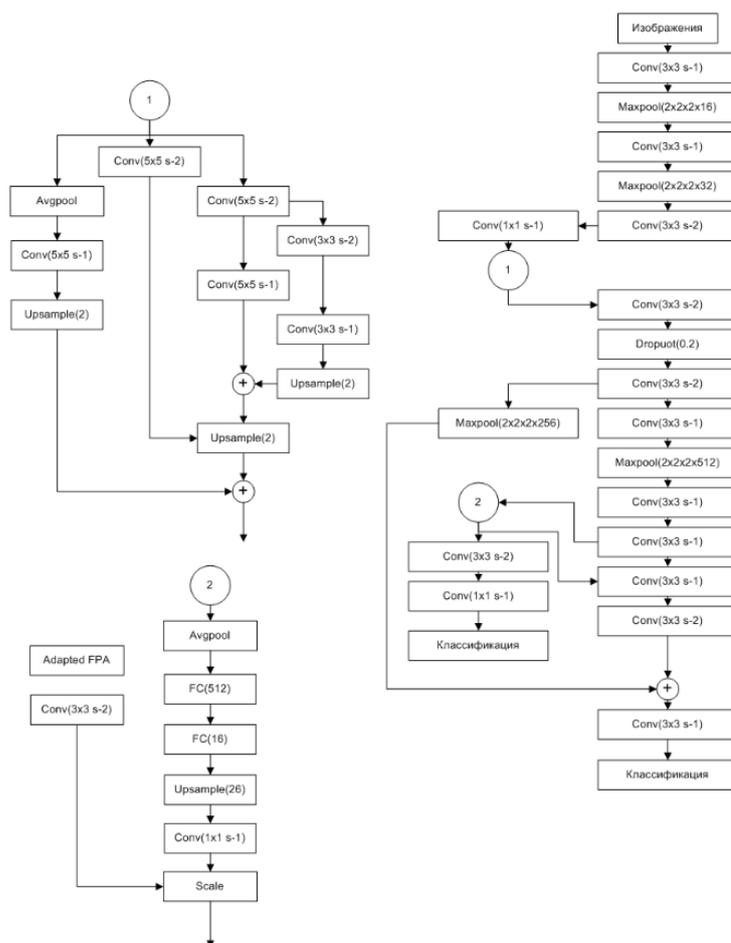


Figure 6 – Main structure of the neural network

#### 4. Processing and control using biomedical signals

During the last decade, methods of decomposition of the surface electromyographic (EMG) signal have been developed to extract neural information transmitted from the spinal cord to the muscles. The paper evaluates the accuracy of identification of hand actions based on high-density EMG signals. EMG signals have been decomposed into motor spike chains (MUST) using a blind source separation algorithm. A gesture recognition approach based on the classification of motor units has been proposed. The MUSTs were first grouped into 11 movements. The output class of gestures was determined by comparing the estimated activation level of each movement.

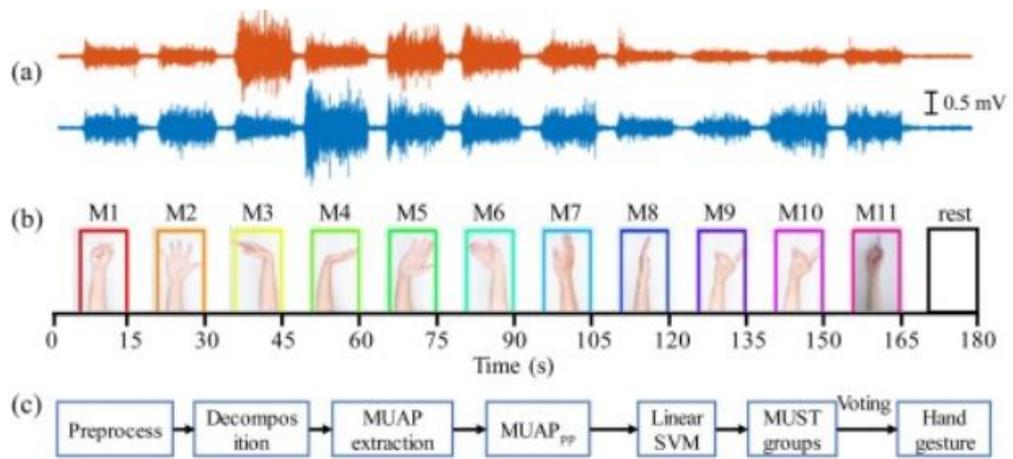


Figure 7 – Classification of gestures by EMG

### 5. Mouse cursor control system based on human hand gestures

In this paper hand positions were suggested and used to generate mouse events as it is shown in Figure 8. The “default” hand position (Fig. 8a) was used for mouse movement, as well as for determining the “left button released” or “right button released” gestures. Two positions were attached to the event “pressing the left mouse button”: the position of the hand “pressed the left button” No. 1 (Fig. 8b) and the position of hand No. 2 “pressing the left button” (Fig. 8c). The “left button pressed” hand position was attached to the event by pressing the right mouse button (Fig. 8d). The gesture is considered as two positions of the hand. For the left button released, the event was considered sequential: the hand position while the left button was pressed follows the default hand position.

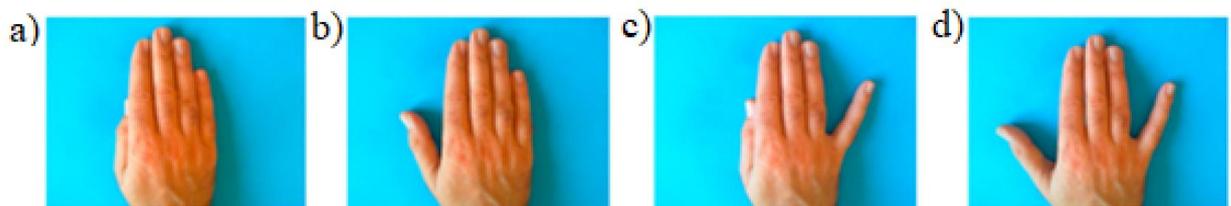


Figure 8 – Gestures for mouse control

## Analysis of the methods for solving the gesture recognition problem

Let us compare the existing inventions with each other in order to identify the advantages and disadvantages of each of them.

A brief description of the considered algorithms for solving the problem of recognizing hand gestures is given in Table 1.

Table 1 – Comparison of algorithms

№	Name	Equipment	Input	Algorithm	Advantages	Disadvantages
1	Gesture recognition with infrared sensors based on Intel RealSense	Intel RealSense device, computing machine	3D data of the skeleton of the hand	Using infrared sensors, obtaining compounds with subsequent classification	No influence of light on the operation of the algorithm, high recognition accuracy	The need for an additional Intel RealSense device, short range, algorithm complexity
2	Using infrared sensors, obtaining compounds with subsequent classification	Microsoft Kinect device, computing machine	3D data on key points of the human body	Converting a Recognition Problem to a Shortest Path Problem Using PCNN	Lack of influence of light on the operation of the algorithm, the speed of the algorithm	Complexity of implementation, high computational costs, the need for an additional Microsoft Kinect device
3	Real-time gesture recognition based on network-based recalibration of functions with multi-scale information	Camera, computing machine	Image fed to the input of a convolutional layer	Image processing with convolutional layers to isolate features of different levels	High accuracy of gesture classification, good working range	Large neural network architecture, low operating speed
4	Biomedical signal processing and control	Special equipment for obtaining an EMG signal	Electromyographic (EMG) signal to extract neural information transmitted from the spinal cord to the muscles	Receiving an EMG signal and its classification using the SVM method	Lack of influence of light on the operation of the algorithm, the speed of the algorithm	Classification of a limited number of gestures, the need for wires with signals
5	Mouse Cursor	Camera, computing machine	Image fed to the input of a	Calculation of the	High accuracy of	Limited work space of the

	Control System Based on Human Hand Gestures	g machine	convolutional layer	extreme points belonging to the hand in the image and the subsequent classification of the gesture depending on the distance	gesture classification , high speed of the program	program, the ability to recognize only one hand
--	---	-----------	---------------------	--	--	---

Based on Table 1, it can be seen that the considered methods have high recognition and classification accuracy, but the algorithms often work in a limited area, and are not capable of recognizing several gestures at the same time.

To solve the problem of simultaneous multiple recognition of gestures, as well as their indication in the image, a neural network will be developed.

The work will be performed on the following points:

1. Getting an image of a gesture;
2. Creation of gesture blanks for generating a synthesized data set;
3. Creation of training, evaluation and test dataset;
4. Neural network training;
5. Evaluation of the recognition accuracy on the test sample.

### **Application of the Inventive Problem-Solving Algorithm – 61**

#### **1. Analytical stage**

First step. Set a task. Simultaneous multiple hand gesture recognition.

Second step. Imagine the perfect end result. Simultaneous multiple recognition of multiple hand gestures with high accuracy and indication.

Thirst step. Determine what prevents the achievement of this result (i.e. find a contradiction). Lack of large computing power capable of training a neural network created for a given task.

Fourth step. Determine why it interferes (find the reason for the contradiction). Due to the lack of a powerful computer, it is impossible to train the complex architecture of a neural network.

Fifth step. Determine under what conditions it would not hurt (that is, find the conditions under which the contradiction is removed). With high power of computing devices and technical perfection of recognition algorithms (video cards with memory of 11 GB or more are required to process a large number of images).

## **2. Operational stage**

First step. Checking possible changes in the object itself (i.e. in this machine, this technological process)

1. Resizing. Resizing the computing device to increase power.
2. Change of speed. Changing the speed of the algorithm.

Second step. Checking the possibility of dividing an object into independent parts.

1. Isolation of the "weak" part. The computing part lacks power (video cards with memory up to 8 GB).
2. Highlighting the "necessary and sufficient" part. A necessary and sufficient part is a high-definition camcorder (Full HD camcorder resolution).

Step three. Checking for possible changes in the external (for a given object) environment.

1. Measurement of environmental parameters. Changing external light sources.
2. Using the external environment to perform useful functions. Changing the environment for easier recognition of the gesture in the background.

Fifth step. Research of prototypes from other branches of technology (ask the question: how is this contradiction eliminated in other branches of technology?) Installation of more powerful equipment, with more complex algorithms (installation of a computer with parameters: more than 11 GB of video memory, more than 16 GB of RAM, a video camera with a resolution of 4k).

### **3. Synthetic stage**

First step. Making changes to the form of this object (a new form must correspond to the new entity of the machine). Changing the appearance for a convenient assembly of more powerful devices (increase in volume up to 0.001 m<sup>3</sup>).

Second step. Making changes to other data-related objects. Changing the ways of receiving and processing information using cameras.

Step three. Making changes to the method of using the object. Creation of algorithms with other methods of information processing (recurrent, fully connected neural networks)

Fourth step. Checking the applicability of the found principle of the invention to solving other technical problems. Based on this algorithm, it is possible to recognize various objects, for example, the recognition of pedestrians for unmanned vehicles [11].

### **Computer vision tasks**

Today, technical vision technologies are applied everywhere, they are increasingly used in various areas of human life, for example, in smartphones cameras to detect human faces.

Modern computer vision tasks are:

- Classification – classification of an image by the type of object which it contains;
- Semantic segmentation – determination of all pixels of objects of a certain class or background in the image. If several objects of the same class overlap, their pixels are not separated from each other in any way;
- Object detection – detection of all objects of the specified classes and defining an enclosing frame for each of them;
- Segmentation of objects – definition of pixels belonging to each object of each class separately.

For visibility, examples of the implementation of classification and semantic segmentation are shown in Figure 9.

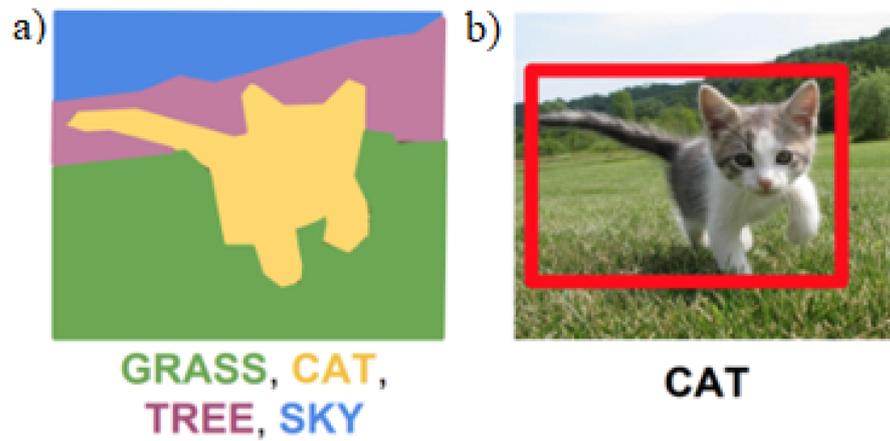


Figure 9 – a) Semantic segmentation, b) Classification with object detection

There are also examples of detecting multiple objects and segmentation of various objects in Figure 10 [12].

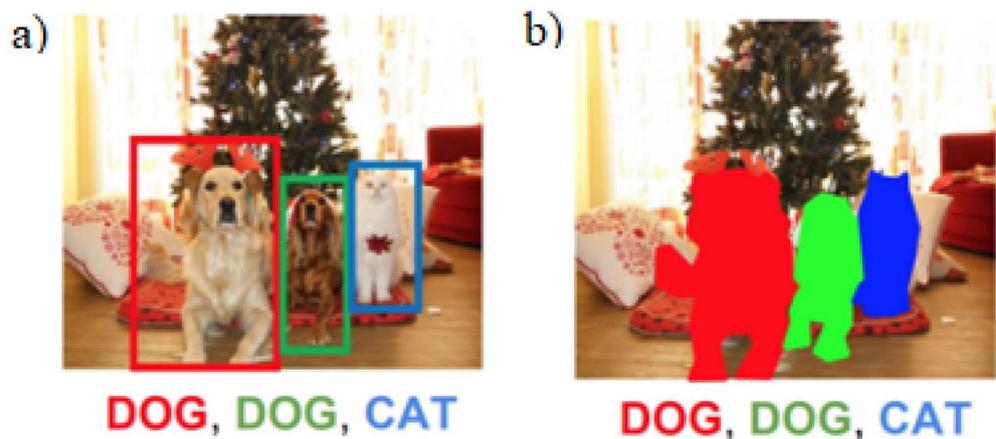


Figure 10 – a) Object detection, b) Object segmentation

To solve the problem of hand gesture segmentation, the Mask-RCNN architecture (Convolutional neural network that imposes masks on the detected objects) was used.

This architecture was developed by researchers at Facebook in 2017. Mask-RCNN is a complete version of the Fast-RCNN architecture development. A distinctive feature of Mask-RCNN is a higher operating speed, as well as the ability to apply masks to detected objects, which allows you to select these objects in the general image [13].

## **Conclusion**

As a result of this work, the history of the development of devices and algorithms for recognizing hand gestures has been analyzed: from gloves for recognition, working with strain gauges, to modern methods of solving the problem. Moreover, in this part of dissertation, brief descriptions of the work of modern algorithms have been given.

Gesture recognition is essential for human-machine interaction. Modern methods for solving gesture recognition have several disadvantages, such as low recognition speed, low speed and low performance when recognizing multiple targets or targets at a great distance in difficult conditions. In view of the aforementioned problems, the Inventive Problem-Solving Algorithm – 61 method has been proposed to describe a workable device that will avoid many disadvantages.

## References

1. Digital Data Entry Glove. [Electronic resource]. Access mode: free. Resource link: <http://www.idemployee.id.tue.nl/g.w.m.rauterberg/presentations/HCI-history/sld066.htm> (Access date: 05.03.2020).
2. Sayre Glove Final Project Report. [Electronic resource]. Access mode: free. Resource link: <https://www.evl.uic.edu/entry.php?id=2296> (Access date: 05.03.2020).
3. Wired glove. [Electronic resource]. Access mode: free. Resource link: [https://en.wikipedia.org/wiki/Wired\\_glove](https://en.wikipedia.org/wiki/Wired_glove) (Access date: 05.03.2020).
4. Interfacing the Nintendo Power Glove to a Macintosh Computer [Electronic resource]. Access mode: free. Resource link: <https://pdfs.semanticscholar.org/9555/ac04a85a8a5fe59ce932f47fe7e5d8d13af5.pdf> (Access date: 06.03.2020).
5. Gesture recognition using colored gloves [Electronic resource]. Access mode: free. Resource link: <https://www.semanticscholar.org/paper/Gesture-recognition-using-colored-gloves-Iwai-Watanabe/03d1b16cbcc133d60bea4a2038662412ccabb3b9> (Access date: 07.03.2020).
6. Heterogeneous hand gesture recognition using 3D dynamic skeletal data Q. De Smedt, H. Wannous and J.-P. Vandeborre/ Computer Vision and Image Understanding 181 (2019) 60–72.
7. Dynamic gesture recognition by directional pulse coupled neural networks for human-robot interaction in real time/ J. Dong et al. / J. Vis. Commun. Image R. 63 (2019) 102583.
8. Real-time gesture recognition based on feature recalibration network with multi-scale information. Z. Cao, X. Xu and B. Hu et al. / Neurocomputing 347 (2019) 119–130.

9. Hand gesture recognition based on motor unit spike trains decoded from high-density electromyography. C. Chen, Y. Yu, S. Ma et al. / Biomedical Signal Processing and Control 55 (2020) 101637.
10. Human hand gesture based system for mouse cursor control. Horatiu-Stefan Grif et al. / Procedia Manufacturing 22 (2018) 1038–1042.
11. Altshuller G. How to learn to invent. - Tambov: Tambov book publishing house, 1961. - P. 44,55,82.
12. Review of computer vision [Electronic resource]. Access mode: free. Resource link: <https://www.machinelearningmastery.ru/an-overview-of-computer-vision-1f75c2ab1b66/> (Access date: 14.03.2020).
13. Mask R-CNN: the architecture of a modern neural network for the segmentation of objects in images [Electronic resource]. Access mode: free. Resource link: <https://habr.com/ru/post/421299/> (Access date: 16.03.2020).