TOMSK POLYTECHNIC UNIVERSITY ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

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
Специальность 09.04.01 Информатика и вычислительная техника
Отделение школы (НОЦ) Отделение информационных технологий

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

Тема работы
Разработка алгоритма детектирования и распознавания надписей на изображениях реальных сцен

УДК 004.492.3:004.932.72'1

Обучающийся

Группа	ФИО	Подпись	Дата
8ВМ03	Гао Жэньцзе		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Иванова Ю.А.	к.т.н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

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

По разделу «Социальная ответственность»

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

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП, должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын В.Г.	д.т.н.		

Томск – 2022 г.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП

по направлению 09.04.01 Информатика и вычислительная техника

Код компетенции	Наименование компетенции
Универсальные компетенции	
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
Общепрофессиональные компетенции	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями
ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований
ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования
ОПК(У)-7	Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
Профессиональные компетенции	
ПК(У)-1	Способен разрабатывать и администрировать системы управления базами данных
ПК(У)-2	Способен проектировать сложные пользовательские интерфейсы
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК(У)-4	Способен осуществлять руководство разработкой комплексных

	проектов на всех стадиях и этапах выполнения работ
ПК(У)-5	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий

TOMSK POLYTECHNIC UNIVERSITY ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

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 09.04.01 Информатика и вычислительная техника
Отделение школы (НОЦ) Отделение информационных технологий

УТВЕРЖДАЮ:
Руководитель ООП/ОПОП
_____ _____ Спицын В.Г.
(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

ВКР магистра

(ВКР бакалавра/ ВКР специалиста/ ВКР магистра)

Студенту:

Группа	**ФИО**
8ВМ03	Гао Жэньцзе

Тема работы:

Разработка алгоритма детектирования и распознавания надписей на изображениях реальных сцен	
Утверждена приказом директора (дата, номер)	№ 34-63/с от 03.02.2022

Срок сдачи студентом выполненной работы:	07.06.2022

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе *(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.).*	Объектом исследования является нейросетевой алгоритм детектирования и распознавания надписей на изображениях реальных сцен.

Перечень подлежащих исследованию, проектированию и разработке вопросов *(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).*	1. Исследование предметной области; 2. Разработка алгоритмов детектирования и распознавания надписей на изображениях реальных сцен; 3. Работа над разделом по финансовому менеджменту, ресурсоэффективности и ресурсосбережения; 4. Работа над разделом по социальной ответственности; 5. Работа над разделом на английском языке.
Перечень графического материала *(с точным указанием обязательных чертежей)*	1. Схемы структур сверточных нейронных сетей; 2. Таблицы сравнительных обучений сетей; 3. Предсказанные результаты для тестовых образцов; 4. Таблица трудоемкости выполнения работ; 5. Диаграмма Ганта.

Консультанты по разделам выпускной квалификационной работы
(с указанием разделов)

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Доцент ОСГН ШБИП, к.э.н., Былкова Т.В.
Социальная ответственность	Профессор ООД ШБИП, д.м.н., Федоренко О.Ю.
Английский язык	Старший преподаватель ОИЯ ШБИП, Ануфриева Т.Н.
Названия разделов, которые должны быть написаны на иностранном языке:	
Раздел 1 Text detection and recognition methods	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	01.03.2022

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Иванова Ю.А.	к.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ03	Гао Жэньцзе		

TOMSK POLYTECHNIC UNIVERSITY ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

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 09.04.01 Информатика и вычислительная техника
Уровень образования Магистратура
Отделение школы (НОЦ) Отделение информационных технологий
Период выполнения (осенний / весенний семестр 2021 /2022 учебного года)

Форма представления работы:

ВКР магистра

(ВКР бакалавра/ ВКР специалиста/ ВКР магистра)

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы

Обучающегося:

Группа	ФИО
8ВМ03	Гао Жэньцзе

Тема работы:

Разработка алгоритма детектирования и распознавания надписей на изображениях реальных сцен

Срок сдачи студентом выполненной работы:	07.06.2022

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
07.06.2022	Основная часть	70
07.06.2022	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
07.06.2022	Социальная ответственность	10
07.06.2022	Приложение на английском языке	10

СОСТАВИЛ:
Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Иванова Ю.А.	к.т.н.		

СОГЛАСОВАНО:
Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын В.Г.	д.т.н.		

Обучающийся

Группа	ФИО	Подпись	Дата
8ВМ03	Гао Жэньцзе		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8ВМ03	Гао Жэньцзе

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:	
1. *Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих*	Оклад инженера – 22 695 руб. Оклад руководителя – 35111,5 руб
2. *Нормы и нормативы расходования ресурсов*	Районный коэффициент 30%; Коэффициент дополнительной заработной платы 12%; Накладные расходы 16%.
3. *Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования*	Коэффициент отчислений на уплату во внебюджетные фонды 30%.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. *Оценка коммерческого и инновационного потенциала НТИ*	Анализ перспективности технических решений посредством Quad-анализа и SWOT-анализ. Оценка готовности научно-исследовательского проекта к коммерциализации.
2. *Разработка устава научно-технического проекта*	Определение цели научно-исследовательского проекта, требований к проекту, описание заинтересованных стороны проекта, рабочей группы.
3. *Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок*	Определение календарного графика проведения исследования. Расчет бюджета затрат на проведение исследовани.
Перечень графического материала *(с точным указанием обязательных чертежей):*	
1. Оценочная карта технологии QuaD 2. Матрица SWOT 3. Диаграмма Ганта, 4. Бюджет НТИ	

Дата выдачи задания для раздела по линейному графику	

Задание выдал консультант:

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

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ03	Гао Жэньцзе		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО		
8ВМ03	Гао Жэньцзе		
Школа	ИШИТР	**Отделение (НОЦ)**	ОИТ
Уровень образования	Магистратура	**Направление/специальность**	09.04.01 Информатика и вычислительная техника

Тема ВКР:

Разработка алгоритма детектирования и распознавания надписей на изображениях реальных сцен	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объект исследования: нейросетевой алгоритм. Область применения: распознавание надписей на изображениях реальных сцен. Рабочая зона: площадь помещения составляет 33,6м^2, есть искусственное и естественное освещение, а также отопление и вентиляция, присутствует рабочий стол и персональный компьютер, назначение офисного типа.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности при разработке проектного решения: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	Трудовой кодекс: Федеральный закон от 30 декабря 2001г. N 197-ФЗ. ГОСТ 12.0.003-2015 Опасные и вредные производственные факторы. Классификация. Перечень опасных и вредных факторов. ГОСТ 22269-76 «Рабочее место оператора. Взаимное расположение элементов рабочего места». ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования. ГОСТ Р 50923-96. Дисплеи. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения. ГОСТ 21889-76 «Система "Человек-машина". Кресло человека-оператора. Общие эргономические требования». СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания. ГОСТ 12.1.005-88 Система стандартов безопасности труда (ССБТ). Общие санитарно-гигиенические требования к воздуху рабочей зоны. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95. ГОСТ 12.1.003-83 Система стандартов безопасности труда (ССБТ). Шум. Общие

	требования безопасности. ГОСТ 12.1.030-81 Система стандартов безопасности труда (ССБТ). Электробезопасность. Защитное заземление. Зануление. ГОСТ 12.1.038-82 Система стандартов безопасности труда (ССБТ). Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов. ГОСТ 12.1.004-91 Система стандартов безопасности труда (ССБТ). Пожарная безопасность. Общие требования. ГОСТ 17.4.3.04-85 Охрана природы (ССОП). Почвы. Общие требования к контролю и охране от загрязнения.
2. Производственная безопасность при разработке проектного решения: – Анализ выявленных вредных и опасных производственных факторов – Расчет уровня опасного или вредного производственного фактора	Вредные факторы: - Недостаточная освещенность рабочей зоны; -Повышенный уровень электромагнитных излучений; - Отклонение показателей микроклимата; - Повышенный уровень шума на рабочем месте; - Умственное перенапряжение. Опасные факторы: - Поражение электрическим током. Требуемые средства коллективной и индивидуальной защиты от выявленных факторов: наушники, устройства для вентиляции и очистки воздуха, источники света, устройства очистки воздуха, защитные покрытия от электромагнитного излучения, устройства защитного заземления. Расчет: расчет системы искусственного освещения
3. Экологическая безопасность при разработке проектного решения	Воздействие на селитебную зону, гидросферу и атмосферу отсутствует. Воздействие на литосферу при утилизации компьютера и периферийных устройств (принтеры, МФУ, веб-камеры, наушники, колонки, телефоны);
4. Безопасность в чрезвычайных ситуациях при разработке проектного решения	Возможные ЧС: пожар, обрушение здания. Наиболее типичная ЧС: пожар.
Дата выдачи задания для раздела по линейному графику	

Задание выдал консультант:

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

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ03	Гао Жэньцзе		

РЕФЕРАТ

Выпускная квалификационная работа 96 страниц, 36 рисунка, 16 таблицы, 47 источников.

Ключевые слова: детектирование текста, распознавание текста, глубокое обучение, обработка изображений, машинное обучение.

Объектом исследования является нейросетевой алгоритм детектирования и распознавания надписей на изображениях реальных сцен.

Целью работы является разработка алгоритма детектирования и распознавания надписей на изображениях реальных сцен.

В процессе исследования были проведёны анализ и сравнение существующих методов детектирования и распознавания надписей на изображениях реальных сцен и определился с выбором метода. Был выбран метод соединения части обнаружения текста и части распознавания текста.

В результате исследования был разработан алгоритм детектирования и распознавания надписей на изображениях реальных сцен. Конфигурация и параметры обучения были выбраны для модели в процессе обучения. В результате конечного метода достигается точность детектирования текста 79,8% и точность распознавания текста 72,9%.

Область применения: компании, разрабатывающие возможности мгновенного перевода, поиска изображений, анализа сцен и геолокации.

ОГЛАВЛЕНИЕ

ВЕДЕНИЕ 13
1 МЕТОДЫ ДЕТЕКТИРОВАНИЯ И РАСПОЗНАВАНИЯ ТЕКСТА 14
1.1 Методы детектирования текста 14
1.1.1 Традиционные методы детектирования текста 14
1.1.2 Методы глубокого обучения для детектирования текста 16
1.2 Метод распознавания текста 24
1.3 Сравнение методов детектирования текста 25
2 ВЫБОР МЕТОДА И РЕАЛИЗАЦИЯ АЛГОРИТМА 27
2.1 Используемые инструменты разработки 27
2.2 Реализация модели FCN 28
2.3 Реализация модели U-net 31
2.4 Реализация модели EAST 32
2.5 Сравнение результатов детектирования текста 35
2.6 Реализация финального метода 38
2.6.1 Ветка распознавания 39
3 РЕЗУЛЬТАТЫ МОДЕЛИ 43
3.1 Результаты детектирования текста для метода 43
3.2 Результат вырезания области текста 43
3.3 Результаты распознавания корейского и английского текста 44
3.4 Результаты распознавания текста для модели EAST 46
3.5 Результат точности метода распознавания текста 47
4 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ 49
4.1 Предпроектный анализ 49
4.2 Планирование управления научно-техническим проектом 53
4.3 Выводы по разделу 59
5 СОЦИАЛЬНАЯ ОТВЕТСВЕННОСТЬ 60
5.1 Правовые и организационные вопросы обеспечения безопасности 60
5.1.1 Специальные правовые нормы трудового законодательства 60

5.2 Организационные мероприятия при компоновке рабочей зоны 62
5.3 Производственная безопасность 63
5.3.1 Вредные производственные факторы 65
5.4 Экологическая безопасность 73
5.4.1 Анализ влияния объекта исследования на окружающую среду 73
5.4.2 Обоснование мероприятий по защите окружающей среды 74
5.5 Безопасность в чрезвычайных ситуациях 75
5.5.1 Анализ вероятных ЧС, которые может инициировать объект разработки 75
5.5.2 Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС 75
5.6 Выводы по разделу 77
ЗАКЛЮЧЕНИЕ 78
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ 79
ПРИЛОЖЕНИЕ А 84

ВЕДЕНИЕ

Тексты — это образы или символы, несущие язык, фиксирующие идеи и передающие знания. В современном обществе наши жизненные сцены полны различной текстовой информации. Текст с конкретной и четкой семантикой является чрезвычайно важным обобщением, описанием и выражением для реальных сцен. Детектирование текста реальной сцены является ключевой технологией для реализации интеллектуального восприятия сцены, имеет важные исследовательское значение. Однако из-за сложного и разнообразного фона, неоднородных текстовых шрифтов, несовместимых размеров и неопределенных направлений текста в реальных сценах текущая обработка этой задачи не достигла идеальных результатов.

Обработка текста обычно делится на два этапа: детектирование текста и распознавание текста. Традиционные методы детектирования текста обычно используют ручное извлечение признаков для детектирования текста, а затем эти тексты распознаются путем сопоставления с шаблоном или обучения модели машинного обучения. Текущие методы глубокого обучения используют сверточные нейронные сети вместо методов ручного извлечения признаков для детектирования текста, а затем для распознавания текста применяются нейронные сети.

Целью данной работы является разработка алгоритма детектирования и распознавания надписей на изображениях реальных сцен. Для достижения поставленной цели необходимо решить следующие задачи:

- анализ существующих моделей и алгоритмов детектирования и распознавания надписей на изображениях реальных сцен;
- поиск и анализ наборов данных;
- выбор моделей и алгоритмов для реализации;
- разработка алгоритма детектирования и распознавания надписей на изображениях реальных сцен;
- реализация алгоритма и проведение численных экспериментов.

1 Методы детектирования и распознавания текста

1.1 Методы детектирования текста

Детектирование текста можно рассматривать как особую форму задачи детектирования объектов компьютерного зрения. Входные данные для этой задачи — изображение, содержащее текст, а выходные — предсказательная информация в виде ограничивающих рамок или выделенные пиксели изображения, которые относятся к текстовой области (в случае попиксельной сегментации). Результатом общих задач детектирования объектов является положение и площадь таких объектов, как животные, мебель и автомобили на изображении, в то время как детектирование текста в основном фокусируется на точном позиционировании текста на изображении. По сравнению с обычным детектированием объектов текст в реальной сцене имеет характеристики разнонаправленности, неправильной формы, экстремального соотношения сторон, различных шрифтов, цветов и фона. Таким образом, алгоритмы, которые успешны в общем детектировании объектов, часто не могут быть напрямую перенесены на задачу детектирование текста.

Детектирование текста значительно продвинулось от детектирования горизонтального текста в начале к детектированию разнонаправленного текста, изогнутого текста и текста неправильной формы. Вообще говоря, методы детектирования текста можно разделить на две категории: традиционные методы детектирования текста и методы детектирования текста, основанные на глубоком обучении. А методы детектирования текста, основанные на глубоком обучении, в основном делятся на детектирование текста на основе сегментации и детектирование текста на основе регрессии.

1.1.1 Традиционные методы детектирования текста

Традиционные методы детектирования текста в основном делятся на следующие этапы:

- предварительная обработка изображения: изображение в оттенках серого, бинаризация, сглаживание, детектирование и коррекция наклона и т. д.;
- сегментирование изображение;
- извлечение признаков из изображения и их анализ;
- обучение моделей распознаванию текста;
- постобработка результатов распознавания.

Эти традиционные методы детектирования имеют плохую адаптируемость и защиту от помех для различных текстов из-за их ограниченных возможностей обработки изображений. Например, метод максимально стабильных экстремальных областей (MSER) [3], гистограмма направленных градиентов (HOG) [4] и признаки Хаара (Haar-like features) [5].

Метод MSER использует экстремальные области для распознавания текстовых областей. Экстремальные области на изображении определяются двумя условиями относительно множества пикселей внутри области [3]:

- область остаётся постоянной при преобразовании координат (поворот, растяжение);
- область остаётся постоянной при изменении яркости изображения.

Данные методы требуют малое количество памяти для реализации, могут работать в режиме реального времени, но плохо распознают текстовые области на размытых или неконтрастных изображениях, часто дают ложноположительные результаты [6].

Метод HOG [4] происходит разделение изображения на ячейки, затем для каждого пикселя внутри ячейки рассчитывается гистограмма направлений градиентов. Объединение полученных значений для каждой ячейки называется HOG [4] – дескриптором. Полученное градиентное изображение выделяет контуры на изображении и исключает из рассмотрения несущественную информацию [4]. Czarnek [5] в своей работе рассматривает отдельные цифры, а не полноценный текст. Такие данные легко визуализировать и интерпретировать в отличии от изображений реальных сцен, на которых помимо текста

присутствуют и другие объекты. Поэтому метод гистограмм направленных градиентов (HOG) больше подходит для предварительной обработки изображений, но не для самостоятельного применения.

Признаки Хаара в основном состоят из четырех категорий в шаблонах признаков: краевые признаки, линейные признаки, центральные признаки диагональные признаки (рисунок 1.1). В шаблоне признака есть два типа прямоугольников, белый и черный, и значение признака шаблона определяется как сумма пикселей белого прямоугольника минус сумма пикселей черного прямоугольника. Значение признака Хаара отражает изменение оттенков серого на изображении.

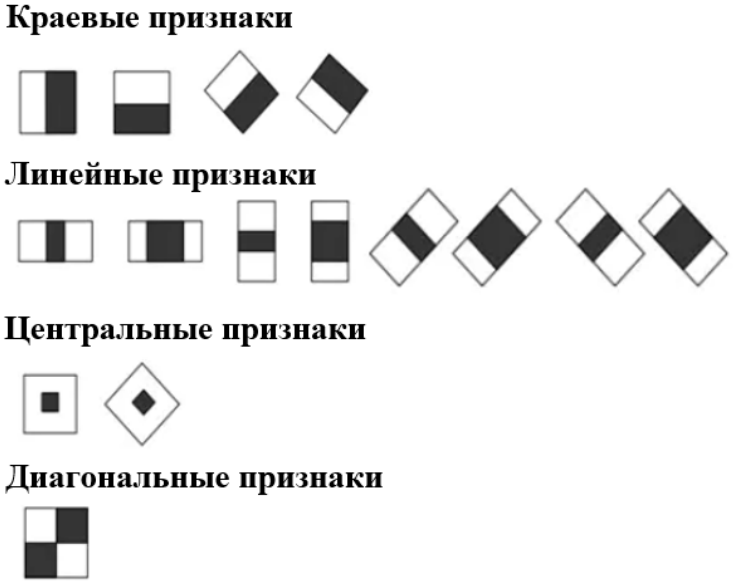

Рисунок 1.1 – Признаки Хаара

Поскольку шрифт, положение и т. д. текста на разных изображениях различаются, этот метод имеет большие недостатки для детектирования текста.

1.1.2 Методы глубокого обучения для детектирования текста

В традиционных методах детектирования текста на основе обучения в основном используются два этапа: «извлечение искусственных признаков» и «предсказание классификатора», а результаты детектирования ограничены возможностью представления искусственных признаков. Глубокие нейронные сети имеют возможность автоматически изучать признаки изображения.

В настоящее время появилось большое количество методов детектирования текста на изображениях реальных сцен, основанных на глубоком обучении, и они достигают лучших результатов, чем традиционные методы детектирования текста. Вообще говоря, методы детектирования текста на изображениях реальных сцен, основанные на глубоком обучении, в основном используют две стратегии обработки изображений с глубоким обучением: детектирование текста на основе регрессии и детектирование текста на основе сегментации.

1.1.2.1 Детектирование текста на основе регрессии

Этот класс методов использует сеть для детектирования общих объектов в качестве базовой модели, и на его основе алгоритм совершенствуется в сочетании с практическим применением детектирования текста. Например, метод Faster R-CNN (Faster region-based convolutional network) [2], метод SSD (Single shot multi-box detector) [1], метод R-FCN (Region-based fully convolutional networks) [13].

Faster R-CNN [2] состоит из четырех подмодулей: сверточный слой, сеть предложений регионов (Region proposal network, RPN), RoI (regions of interest) – пулинговый слой (RoI Pooling layer) и слой классификации и регрессии. Сверточные слои используются для извлечения признаков изображения, RPN используется для генерации нескольких блоков-кандидатов, связанных с текстовыми объектами, RoI – пулинговый слой – преобразует блоки-кандидаты разных размеров в фиксированный размер, слой классификации и регрессии делают прогнозы для области-кандидата и получают точное местоположение области-кандидата на изображении. Поскольку оригинальный Faster R-CNN может детектировать только горизонтально расположенный текст, Цзян и др. предложили алгоритм R2CNN (Rotational region CNN) [14] для детектирования текста, повернутого под любым углом в реальных сценах. Основанный на оригинальном Faster R-CNN, этот алгоритм использует сеть RPN для генерации

осесимметричных блоков-кандидатов в разных направлениях для текстовой области. После этого признаков областей-кандидатов в каждом направлении объединяются с различными размерами пула для объединения объектов， и определить осесимметричный блок-кандидат и наклонный блок-кандидат. Алгоритм достигает результата прогнозирования со значением F-меры, равной 82.54.

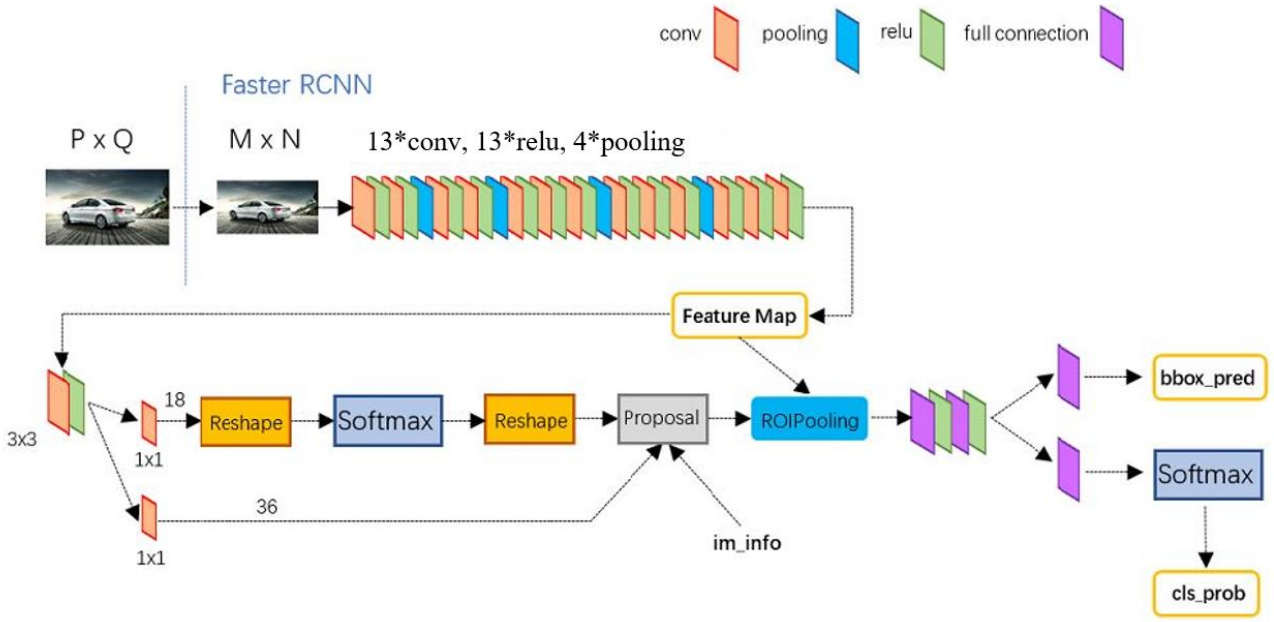

Рисунок 1.2 – Пример структуры Faster R-CNN [2]

SSD [1] вводит изображение в модифицированный VGG16 [15] для получения карт признаков разных размеров, а затем извлекает карты признаков шести сверточных слоев Conv4_3, Conv7, Conv8_2, Conv9_2, Conv10_2, Conv11_2, и строит эталонные поля разного масштаба в каждой точке этих карт признаков соответственно. Наконец, подавление «не-максимумов» (Non-Maximum Suppression, HMS) [16] используется для обработки всех эталонных полей, подавления неоптимальных эталонных полей и вывода окончательного результата детектирования.

Алгоритм SegLink [12], предложенный Ши и др., вводит изображение в сеть SSD и извлекает информацию о соединении между различными текстовыми полями. Вывод модели не для всей текстовой строки или слова, а для части текстовой строки или слова, называемой «сегментом», сегмент может состоять

из 1 или более символов или из 1 слова. Извлекая информацию о соединении из текстового поля, алгоритм принимает комбинацию различных сегментов в качестве конечного результата, избегая процесса постобработки соединения сегментов для построения текстовых строк. Стоит отметить, что текстовое поле, выводимое SegLink, содержит информацию об углах, для каждой точки на карте признаков выводится только одно поле, что значительно снижает вычислительную сложность.

1.1.2.2 Детектирование текста на основе сегментации

Этот тип метода использует семантическую сегментацию в качестве основного технического средства, обрабатывает изображения реальных сцен через сеть семантической сегментации с глубоким обучением и получает предсказание меток на уровне пикселей, эти выходные данные на уровне пикселей являются основой для построения текстовых строк. Сети сегментации, которые часто используются для детектирования текста, включают Mask R-CNN [7], полностью сверточную сеть (Fully сверточная сеть, FCN) [8], PixelLink [9], EAST [17] и др. и классическая сеть U-net [18] для сегментации медицинских изображений.

Маска R-CNN (рисунок 1.2) расширена от Faster R-CNN [2] и Fast RCNN. В дополнение к двум ветвям детектирования (классификация, регрессия ограничивающей рамки) исходной сети, в конце сети добавлена ветвь маски с функцией прогнозирования на уровне пикселей для семантической сегментации. Ветвь маски принимает среднюю бинарную кросс-энтропийную потерю и формирует функцию потерь сети вместе с потерями классификации и регрессионными потерями ограничивающей рамки. Лю и др. на конференции ECCV (Европейская конференция по компьютерному зрению) предложили сеть Mask TextSpotter [10] на основе Mask R-CNN. Его главное новшество состоит в том, чтобы изменить структуру вывода ветви Mask, включив в нее глобальную сегментацию текстовых экземпляров и функции сегментации символов. Mask

TextSpotter использует сегментацию и распознавание на уровне символов, поэтому он может обрабатывать текст любой неправильной формы (например, изогнутый текст), но его ограничение заключается в том, что для завершения обучения модели требуется аннотация на уровне символов.

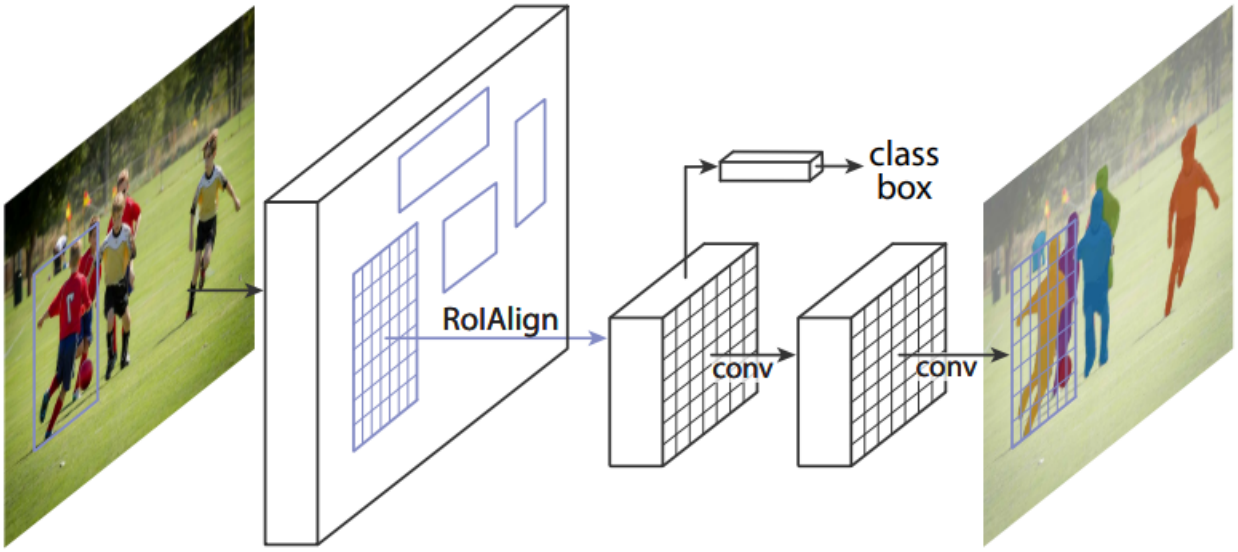

Рисунок 1.3 – Пример структуры Mask-R-CNN [7]

Полностью сверточная сеть FCN [8] – это сквозной метод семантической сегментации, который отличается от модуля классификации областей и модуля регрессии ограничивающей рамки в таких алгоритмах, как Mask R-CNN. В FCN вывод сети представляет собой прогнозирование всего изображения на уровне пикселей. В статье [11] используется FCN для обработки изображения, получения карты значимости (Salient map) текстовой области и анализа ее связанных компонентов для получения текстового блока. На этой основе область символов-кандидатов в текстовом блоке извлекается методом MSER, и каждая текстовая строка генерируется путем объединения ограничивающих рамок символов-кандидатов. При использовании полносвёрточной нейронной сети могут появляться ложные срабатывания или пропущенные символы при очень низкой контрастности, кривизне, сильном отражении света, плотном расположении строк текста или больших промежутках между символами. Еще одним недостатком данного типа сети является быстродействие.

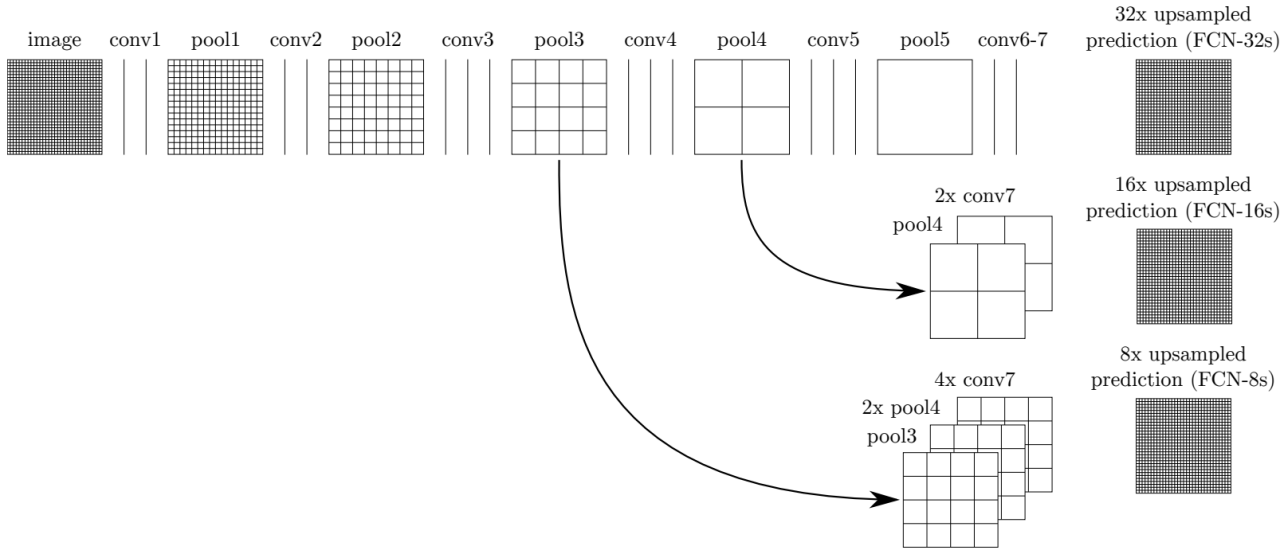

Рисунок 1.4 – Пример структуры FCN [8]

Метод PixelLink [9] предсказывает пиксели и отношения связи, связанные с текстом, через сеть глубокого обучения, применяет метод сегментации экземпляров для сегментации области текстовой строки, а затем непосредственно находит ограничивающий прямоугольник соответствующей текстовой строки. Весь процесс состоит из двух частей: предсказание и соединение «положительных пикселей» в соответствии с результатом предсказания «ссылка положительна», затем получение карты сегментации текстовых экземпляров и непосредственное извлечение ограничивающих рамок текстовых строк из карты сегментации. По сравнению с сегментацией изображения, определение местоположения текста при детектировании текста является более точным, и только метод сегментации не может точно определить местонахождение текста на небольшом расстоянии. Поэтому PixelLink принимает идею «link» в SegLink [12], В прогнозировании он не только предсказывает, являются ли пиксели текстом, но также предсказывает, могут ли эти пиксели быть соединены для формирования хорошего текстового поля и вывода более точного текстового поля.

U-net [18] — это широко используемая и простая модель сегментации изображений, которая эффективна, понятна, проста в построении и может быть обучена на небольших наборах данных. U-net усовершенствована на основе FCN. Первоначальным назначением сети было решение проблемы сегментации

медицинских изображений. Ее сетевая структура показана на следующем рисунке.

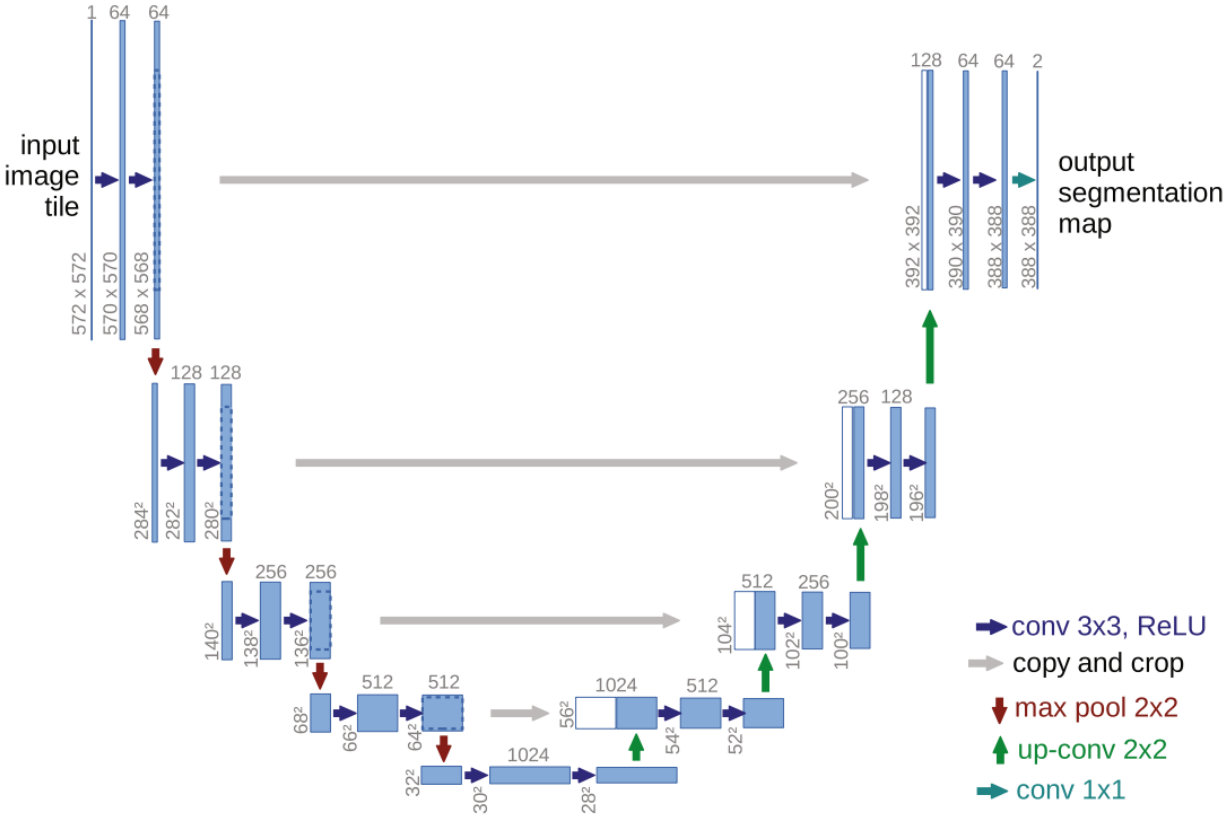

Рисунок 1.5 – Пример структуры U-net [8]

Сеть U-net относительно проста: первая половина (левая часть рисунка) используется для извлечения признаков, а вторая половина (правая часть рисунка) — для повышения частоты дискретизации. Его структура также называется структурой кодер-декодер. Поскольку общая структура этой сети похожа на заглавную букву U, она называется U-net. U-net сильно отличается от других распространенных сетей сегментации: U-net использует совершенно другой метод слияния признаков: сплайсинг, который использует сращивание признаков вместе в измерении канала для формирования более значимых признаков.

Семантика медицинских изображений относительно проста, структура фиксирована, поэтому нет необходимости фильтровывать бесполезную информацию. Все функции медицинских изображений важны, поэтому важны функции низкого уровня и семантические функции высокого уровня, поэтому U-

образные структуры показывают хорошие результаты в таких задачах. Существует много общего между сегментацией изображений текстовых надписей и сегментацией медицинских изображений.

EAST — это алгоритм детектирования текста сцены на основе пикселей, который может предсказывать два разных ограничивающих прямоугольника: одна — RBOX (Rotated Box), которая напрямую предсказывает расстояние от центральной точки до четырех сторон (Это аналогичен идее FCOS) и угол поворота текстового поля; другой QUAD, который предсказывает четыре вершины неправильного четырехугольника. Таким образом, EAST может детектировать повернутые или перекошенные текстовые области, но не криволинейные текстовые области. Кроме того, в статье [17] также предлагается новая NMS: Locality Aware NMS (LANMS), которая представляет собой стратегию, основанную на полях детектирования слияния строк, которая снижает сложность NMS с $O(n^2)$ до оптимальной $O(n)$, скорость постобработки детектирования значительно улучшена. Результаты, полученные EAST, полезны для последующих исследований по распознаванию текста. Его сетевая структура показана на рисунке 1.6.

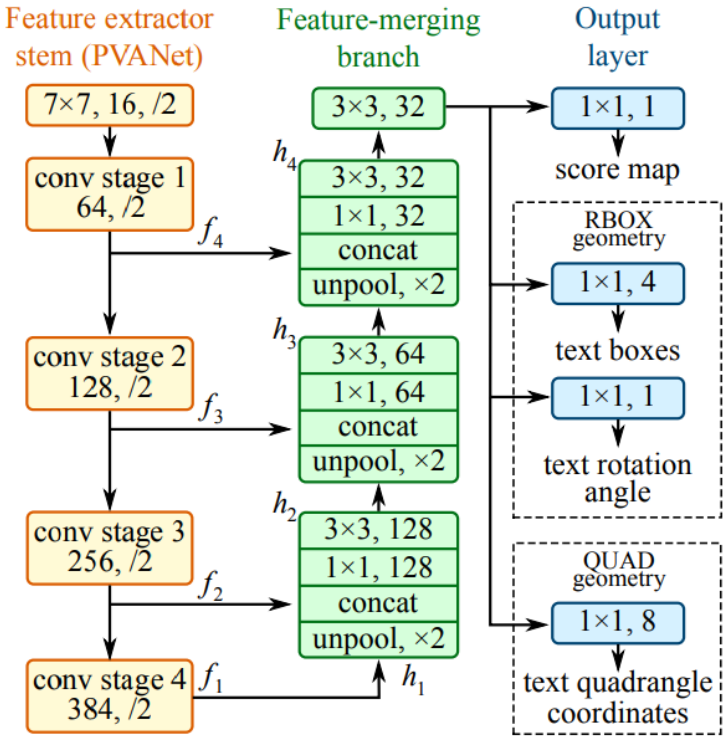

Рисунок 1.6 – Пример структуры EAST [17]

Сетевую структуру EAST можно разделить на три части:

Feature extractor: на рисунке используется PVANet [22], фактически можно использовать любую модель, обученную на больших наборах данных, например, ResNet [23], VGG [15] и т. Д.

Feature merging имеет U-образную структуру: $f1$, $f2$, $f3$, $f4$ описывают информацию о каждой шкале, $h2$, $h3$, $h4$ объединяют информацию о каждой шкале.

Output layer: пунктирная рамка здесь не обязательна. Карта оценки геометрия RBOX образуют выходной слой, или карта оценок и QUAD образуют выходной слой. То есть сеть может выводить четырехугольники или прямоугольники с углами.

1.1 Метод распознавания текста

Методы распознавания текста также делятся на традиционные методы распознавания и методы распознавания, основанные на глубоком обучении. Среди них традиционный метод распознавания подобен методу детектирования, который обрабатывает и распознает изображение через скользящее окно. Поскольку традиционные методы могут распознавать текст только в простых сценах, исследование сосредоточено на методах распознавания, основанных на глубоком обучении.

CRNN+CTC [19 – 20] — классическая архитектура нейронной сети, которая объединяет извлечение признаков, моделирование последовательностей и транскрипцию в единую структуру [19]. Используемая архитектура может естественным образом обрабатывать последовательности любой длины и не требует сегментации символов или нормализации горизонтального размера. Он не ограничен каким-либо предопределенным лексиконом и генерирует эффективную, но гораздо меньшую модель. Его сетевая структура показана на рисунке 1.7.

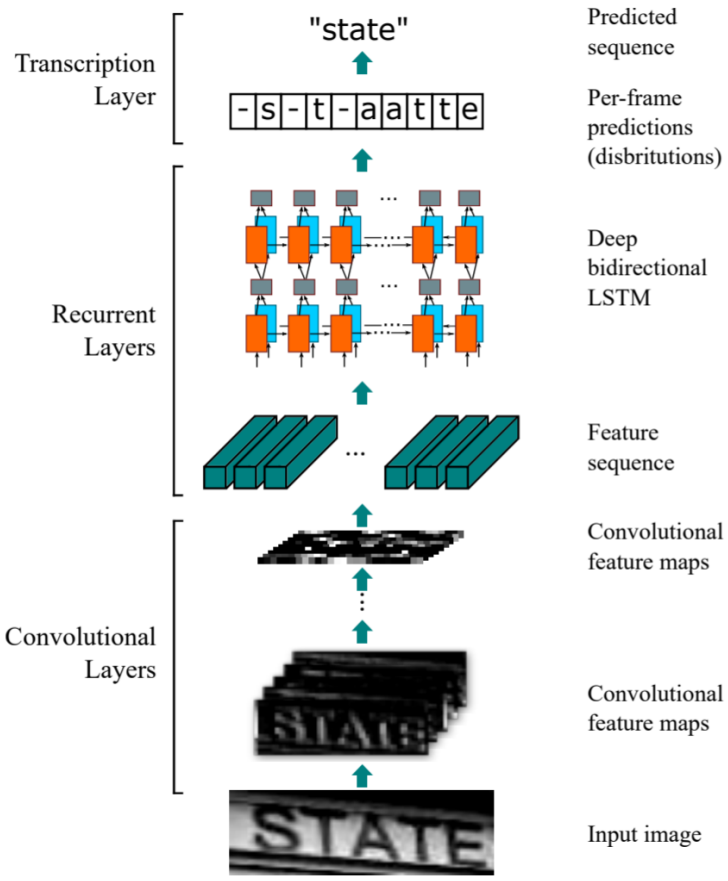

Рисунок 1.7 – структура сети CRNN [19]

Сетевая архитектура CRNN + CTC в основном использует сверточные слои и слои объединения для получения признаки изображении, а также использует LSTM [21] в качестве RNN для прогнозирования текстовой информации, а предсказанная текстовая информация обрабатывается CTC [20] для получения окончательного результата прогнозирования.

1.3 Сравнение методов детектирования текста

Из-за сложности надписей на изображениях реальных сцен традиционные методы детектирования не могут обеспечить достаточно хорошие результаты детектирования, поэтому рассмотренные методы, основанные на глубоком обучении, сравниваются согласно таблице 1. Были рассмотрены следующие методы сравнения:

- размер занимаемых вычислительных ресурсов;
- скорость результаты;

- Точность результатов;
- выполняемость программы метода.

Таблица 1.1 – Сравнение методов

Методы	Методы сравнения			
	Размер занимаемых вычислительных ресурсов	Скорость работы	Точность результатов	Осуществимость программы метода
Faster R-CNN	Средняя	Низкая	Высокая	Средняя
SSD	Средняя	Средняя	Средняя	Средняя
SegLink	Высокая	Средняя	Высокая	Средняя
Mask R-CNN	Высокая	Средняя	Средняя	Средняя
FCN	Средняя	Низкая	Высокая	Высокая
PixelLink	Высокая	Низкая	Средняя	Средняя
U-net	Средняя	Средняя	Высокая	Высокая
EAST	Средняя	Средняя	Высокая	Средняя

Как видно из таблицы, методы детектирования, основанные на глубоком обучении, в целом работают медленнее, но обычно достигают хороших результатов. А по размеру занимаемых вычислительных ресурсов и скорости выполнения программы были выбраны методы FCN, U-net и EAST для дальнейшей реализации и сопоставления.

2 Выбор метода и реализация алгоритма

2.1 Используемые инструменты разработки

Для реализации методов был выбран язык программирования высокого уровня Python. В качестве работающих устройств используются GPU и CPU, предоставляемые Google Colab pro. Для построения, обучения и тестирования модели использовались открытые библиотеки нейронных сетей Pytorch, TensorFlow и Keras.

Для обучения и тестирования сети FCN И U-net была выбрана база данных KAIST Scene Text [24], из которой было отобрано 1025 изображений, из которых 800 использовались для обучения и 225 – для валидации. Каждому изображению соответствует маска изображения, которая используется для определения положения текста. На входе сверточной нейронной сети каждое изображение обрезается от центра до размера 480*480 пикселей, а выходными данными является маска изображения с сегментированными текстовыми областями. Примеры изображений набора данных показаны на рисунке 2.1.

Рисунок 2.1 – Примеры изображений KAIST

ICDAR 2015 [25] был выбран для обучения и тестирования сети EAST. Набор данных ICDAR 2015 состоит из 1000 обучающих изображений и 500 тестовых изображений, размер каждого изображения 1280*720 пикселей. Пример набора данных показан на рисунке 1. Каждое изображение соответствует

файлу аннотаций, в котором каждая строка данных содержит восемь координат, соответствующих верхней левой, верхней правой, нижней правой и нижней левой вершинам текстовой области соответственно, а также текстовые аннотации в текстовой области. Для тех текстов, которые принадлежат текстовой области, но не могут быть распознаны из-за размытости изображения, они помечаются как «###». Примеры изображений набора данных ICDAR 2015 показаны на рисунке 2.2.

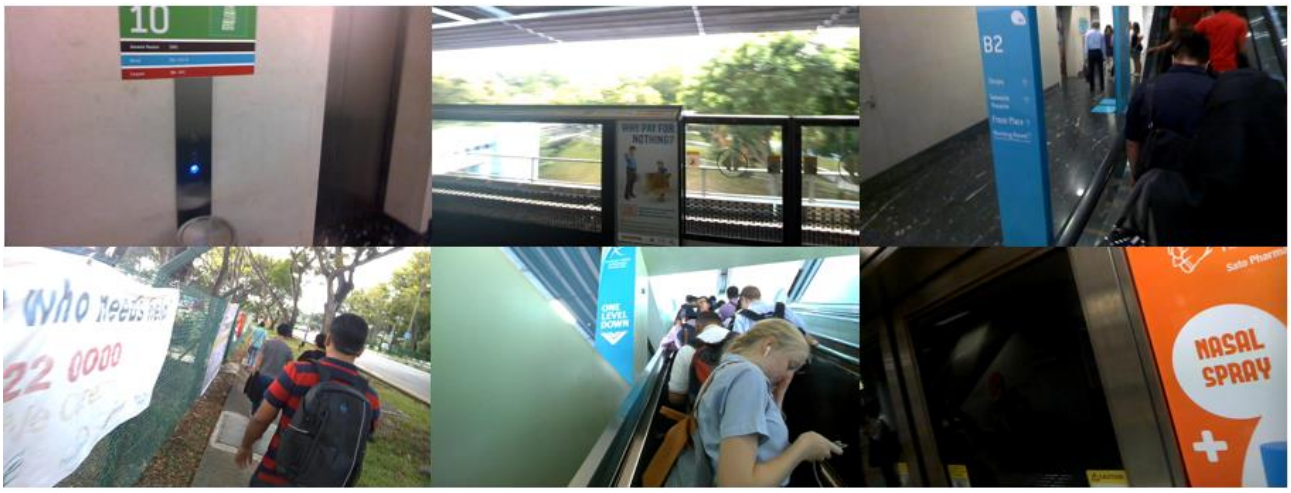

Рисунок 2.2 – Примеры изображений ICDAR 2015

MJSynth [26] был выбран для обучения и проверки сети CRNN. Набор данных MJSynth создан синтетически, состоит из 9 миллионов изображений, охватывающих 90 тысяч английских слов. Каждое из изображений имеет разный размер. В качестве входных данных для сети их необходимо настроить на один и тот же размер. С учетом производительности используемого оборудования было отобрано 135 000 изображений для обучения и 15 000 изображений для проверки. Пример набора данных показан на рисунке 2.3.

Рисунок 2.3 – Примеры изображений MJSynth

2.2 Реализация модели FCN

Общая структура сети разделена на две части: часть понижающей дискретизации и часть повышающей дискретизации. Для лучшей производительности сегментации часть исходного FCN с понижающей дискретизацией заменена глубокой остаточной сетью ResNet-34 [23]. Рисунок 2.4 представляет собой схему структуры сети FCN-ResNet 34.

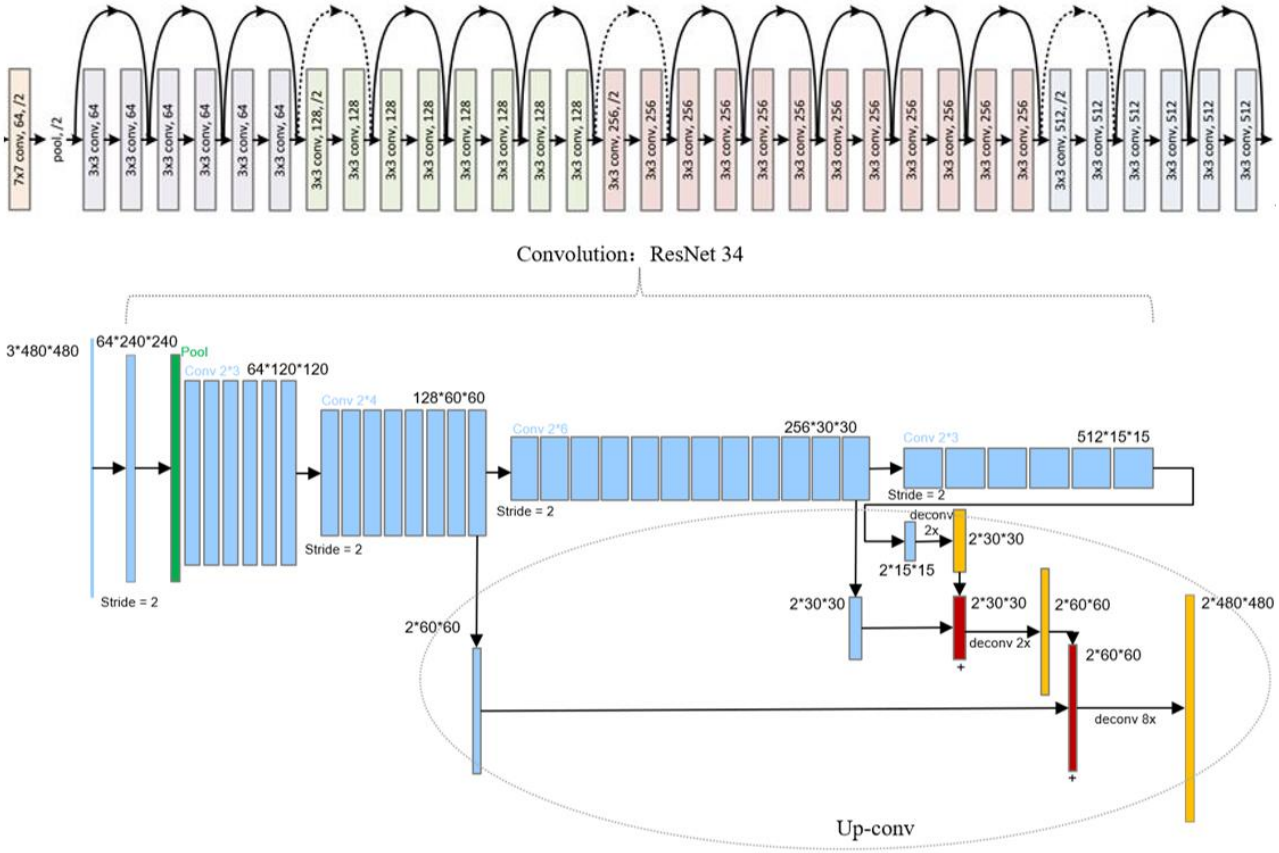

Рисунок 2.4 – структура сети FCN-ResNet34

Архитектуру сверточной нейронной сети ResNet-34 можно рассматривать как пять частей. Первая часть состоит в свертывании входного изображения с шагом 2 с использованием ядра свертки 7×7, уменьшении размера изображения вдвое и увеличении количества каналов до 64. Во второй части используется максимальный слой пула и остаточная сеть 2×3, чтобы вдвое уменьшить размер изображения, каждый слой использует ядро свертки 3×3. В третьей части используется остаточная сеть 2×4, а сеть первого уровня имеет шаг 2, что уменьшает размер изображения вдвое и увеличивает количество каналов до 128 и сохранить карту объектов. Четвертая и пятая части такие же, как и третья, увеличивают количество каналов до 256 и 512 соответственно, уменьшают

размер изображения вдвое и сохраняют карту объектов. Это приводит к трем картам объектов с пониженной дискретизацией и объединяет их количество каналов до 2.

Часть повышающей дискретизации инициализируется билинейным ядром свертки, и в общей сложности апсемплинг выполняется три раза. То есть третья карта объектов увеличивается в два раза с использованием ядра билинейной свертки и объединяется со второй картой объектов. Объединенная карта объектовснова удваивается и объединяется с первой картой объектов, а окончательная карта объектов расширяется в восемь раз и выводится, то есть получается изображение маски с 2 каналами и того же размера, что и входное изображение.

Использование остаточной сети ResNet-34 для субдискретизации имеет несколько преимуществ:

- Может увеличить количество слоев нейронной сети и улучшить производительность сети;
- ResNet имеет преимущество быстрой сходимости и уменьшить объем данных модели.

Поскольку это проблема бинарной классификации, была выбрана функция бинарной перекрестной потери энтропии BCELoss, рассчитывается по формуле (2.1) и (2.2).

$$BCELoss = \frac{1}{N}\sum_{n=1}^{N} l_n \ , \tag{2.1}$$

$$l_n = -w_n[y_n log x_n + (1 - y_n) lo\, g(1 - x_n)], \tag{2.2}$$

где x_n – значение, выводимое моделью, y_n– целевое значение (текс-нетекст), w_n – значение веса.

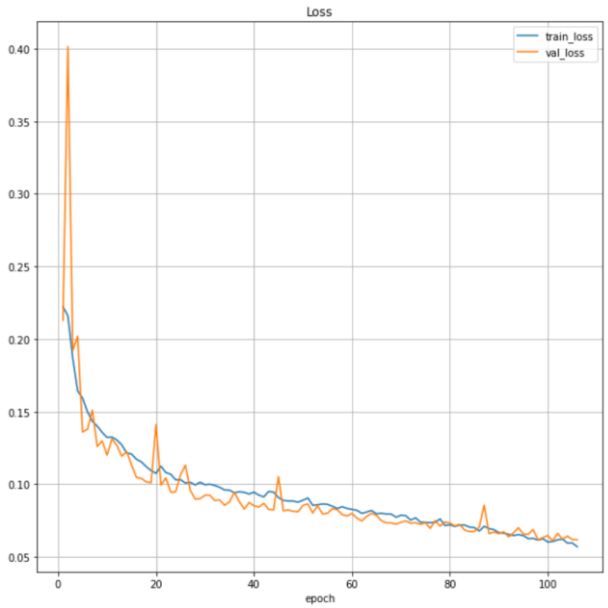

Рисунок 2.5 – Значение функции потерь на обучающей и валидационной выборке для модели FCN

На рисунке 2.5 показаны потери обучения и потери валидации FCN с использованием ResNet-34. Где абсцисса представляет количество эпох обучения, а ордината представляет функцию потерь.

2.3 Реализация модели U-net

U-net [18] — это широко используемая и простая модель сегментации изображений, которая эффективна, понятна, проста в построении и может быть обучена на небольших наборах данных. U-net усовершенствован на основе FCN, архитектура состоит из сужающего пути для захвата контекста и симметричного расширяющегося пути, обеспечивающего точную локализацию [18]. Первая половина используется для извлечения признаков, а вторая половина – для повышения частоты дискретизации. Такая архитектура называется структурой кодер-декодер. В сверточных слоях и слоях обратной свертки используется функция активации ReLU. Выходной слой обеих сверточных нейронных сетей использует сигмовидную функцию активации. Ее сетевая структура показана на рисунке 2.6.

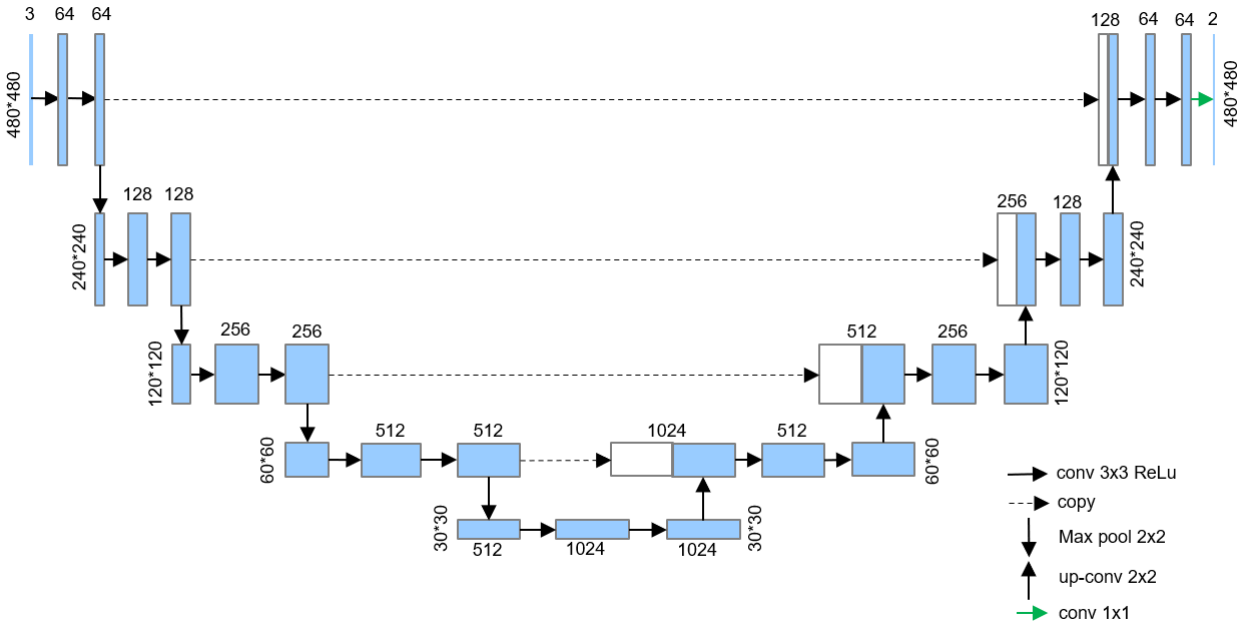

Рисунок 2.6 – структура сети U-net

Поскольку выход сети U-net такой же, как и FCN, это проблема бинарной классификации, поэтому была также выбрана функция бинарной перекрестной потери энтропии BCELoss. На рисунке 2.7 показаны потери обучения и потери валидации U-net.

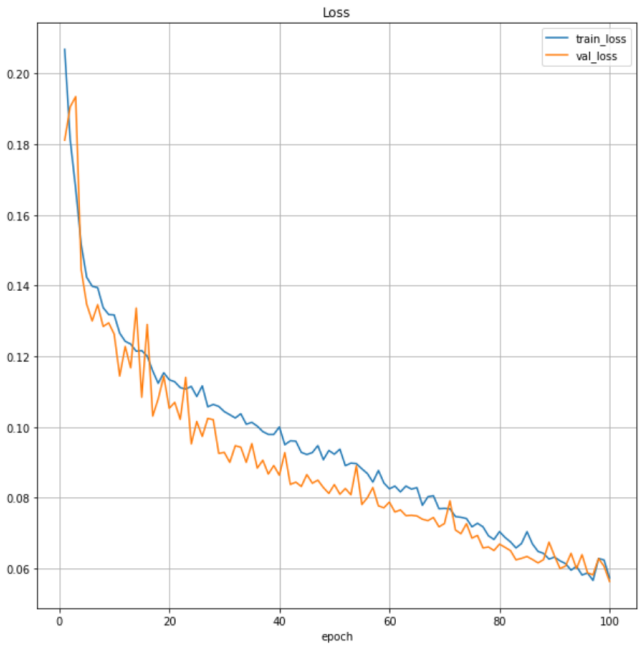

Рисунок 2.7 – Значение функции потерь на обучающей и валидационной выборке для модели U-net

2.4 Реализация модели EAST

EAST [17] использует метод объединения признаков, аналогичный U-net [18], который объединяет признаки нижнего уровня и признаки верхнего уровня. Его сеть извлечения признаков была заменена на ResNet50 [23], и размер входного изображения сети 3×512×512, изображение подвергается понижающей дискретизации сетью ResNet50 для получения четырех карт признаков: *f4*, *f3*, *f2* и *f1* с размерами 256×128×128, 512×64×64, 1024×32×32 и 2048×16×16 соответственно (рисунок 2,8).

Затем *f1* подвергается повышающей дискретизации и объединяется с *f2*, объединенная карта признаков свертывается и получается *h2*; *f2* подвергается повышающей дискретизации и объединяется с *f3*, объединенная карта признаков свертывается и получается *h3*; *f3* подвергается повышающей дискретизации и объединяется с *f4*, объединенная карта признаков свертывается и получается *h4*; *h4* входит в три ветки предсказания после свертки 3×3: карта оценок (1×128×128), геометрическая карта (4×128×128) и угловая карта (1×128×128).

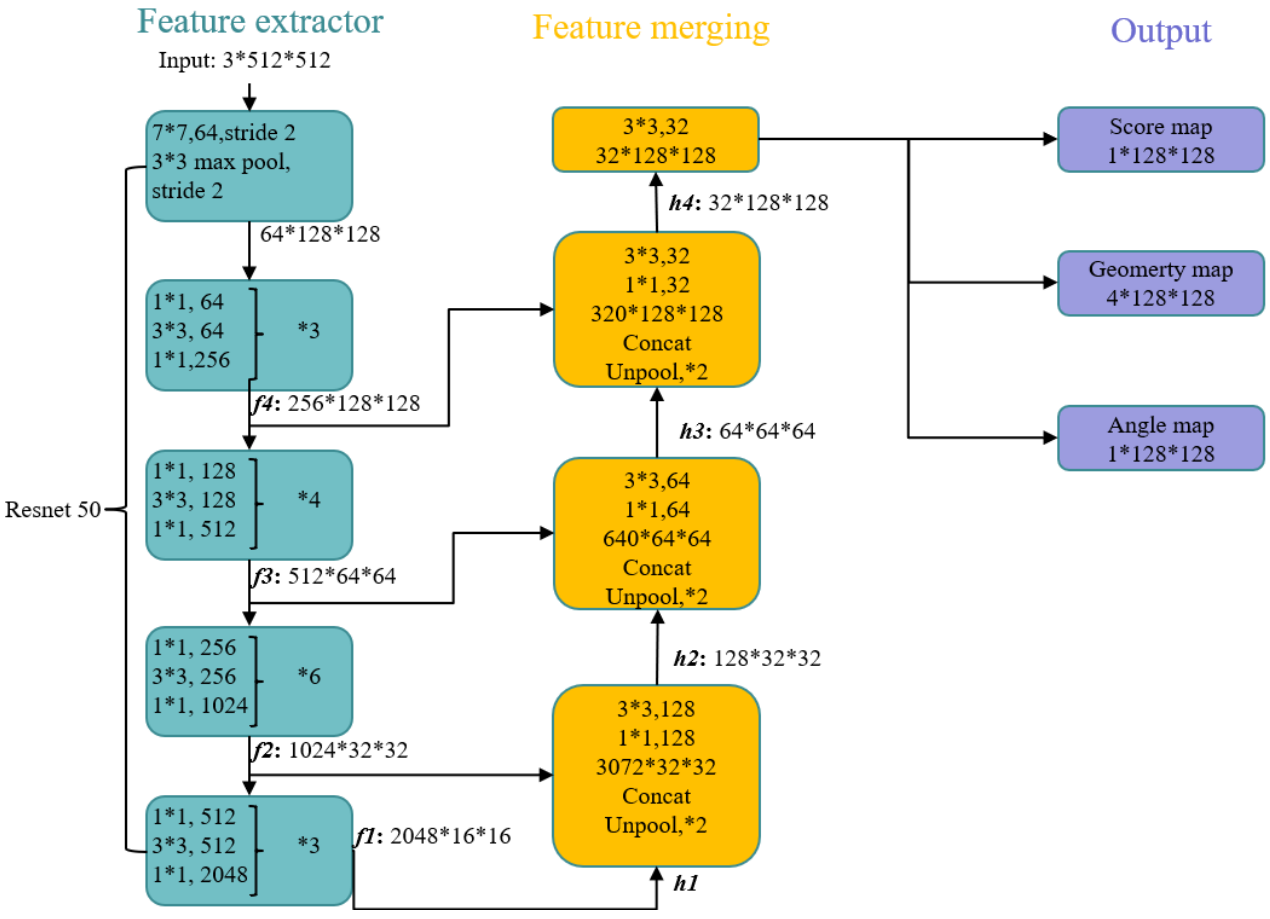

Рисунок 2.8 – структура сети EAST

Карты оценок (Score map) – это предсказание на уровне пикселей для текстовой области. Значение пикселя области текстового поля карты оценок равно 1, а значение других областей, не являющихся текстовыми, равно 0, как показано в 2.9(b). Каждый пиксель в текстовой области геометрической карты (Geometry map) содержит 4 значения, то есть расстояние от пикселя до верхней, нижней, левой и правой части текстового поля, как показано на рисунке 2.9(d). Красный, зеленый, синий и желтый представляют собой расстояние от пикселя до верхней, нижней, левой и правой части текстового поля соответственно. Рисунок 2.9€ представляет собой угловую карту (Angle map), которая задает угол поворота текстового поля. А значения расстояния и угла пикселей в других нетекстовых областях равны 0.

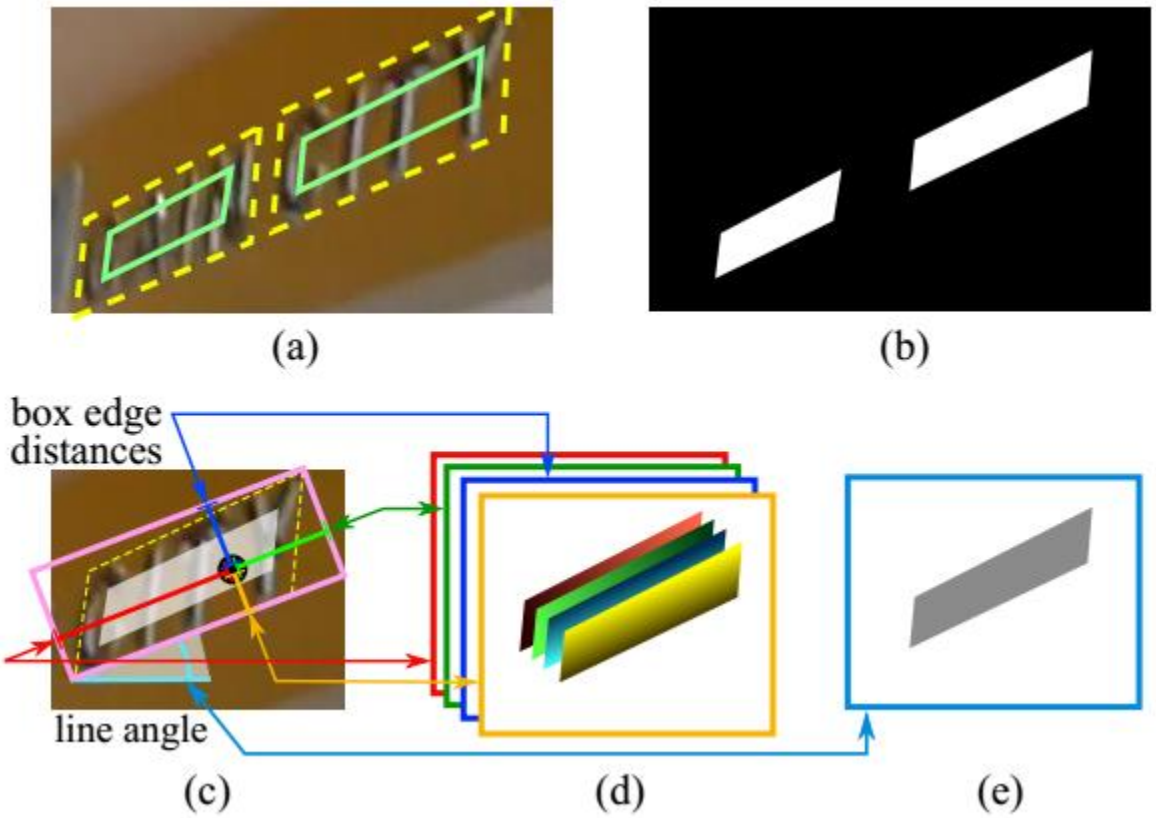

Рисунок 2.9 – процесс генерации этикетки [17]

Функция потерь состоит из двух частей: потери классификации карты оценок и регрессионной потери геометрической карты и угловой карты.

Для потери классификации выбирается функция потерь DiceLoss, которая обычно используется для вычисления подобия двух выборок. Расчет по формуле

(2.3):

$$L_s = 1 - \frac{2|X \cap Y|}{|X| + |Y|}, \tag{2.3}$$

где X — предсказанные значения; Y — фактические значения.

Расчет регрессионного проигрыша включает проигрыш положения и угла наклона прямоугольника. Для расчета убытка по позиции прямоугольника была выбрана функция потерь IoULoss [27], то есть отношение пересечения предсказанных прямоугольников и реальных. Расчет по формуле (2.4).

$$L_{\mathrm{IoU}} = -\log\frac{|X \cap Y|}{|X \cup Y|}, \tag{2.4}$$

где X — предсказанные значения;

Y — фактические значения.

Для расчета угловых потерь были выбраны потери косинусной разности углов. Расчет по формуле (2.5).

$$\mathrm{L}_\theta\left(\hat{\theta}, \theta^*\right) = 1 - \cos\left(\hat{\theta} - \theta^*\right), \tag{2.5}$$

где $\hat{\theta}$ — предсказанный угол;θ^* — фактический угол.

Общие потери прямоугольника расчет по формуле (2.6).

$$\mathrm{L}_g = L_{\mathrm{IoU}} + \lambda_\theta \mathrm{L}_\theta, \tag{2.6}$$

где λ_θ взвешивает важность между двумя потерями и устанавливается равным 1.

Общие потери расчет по формуле (2.7).

$$L = L_{\mathrm{s}} + \lambda_g \mathrm{L}_g, \tag{2.7}$$

где L_{s} и L_g представляют потери для карты оценок и геометрии соответственно, а λ_g взвешивает важность между двумя потерями. В этой практике λ_g устанавливается равным 10.

2.5 Сравнение результатов детектирования текста

Поскольку методы детектирования текста модели FCN и модели U-net отличаются от модели EAST, используются разные формулы оценки. Но оба

требуют вычислительной точности P (*precision*), полноте R (*recall*) и f-мере F (*f-measure*). Оценка для сетей FCN и U-net рассчитывается по формулам (2.8 – 2.10).

$$P = \frac{TP}{TP + FP}, \tag{2.8}$$

$$R = \frac{TP}{TP + FN}, \tag{2.9}$$

$$F = 2 \times \frac{P \times R}{P + R}, \tag{2.10}$$

где

TP (true positive) – сумма всех пикселей изображения, которые модель отнесла к положительному классу («текст») и угадала;

FP (false positive) — сумма всех пикселей изображения, которые модель отнесла к положительному классу («текст») и ошиблась;

TN (true negative) – сумма всех пикселей изображения, которые модель отнесла к отрицательному классу («нетекст») и угадала;

FN (false negative) – сумма всех пикселей изображения, которые модель отнесла к отрицательному классу («нетекст») и ошиблась.

Определить, являются ли два две прямоугольные области одинаковыми — непростая задача для сети EAST. Есть три метода сопоставления (рисунок 2.10).

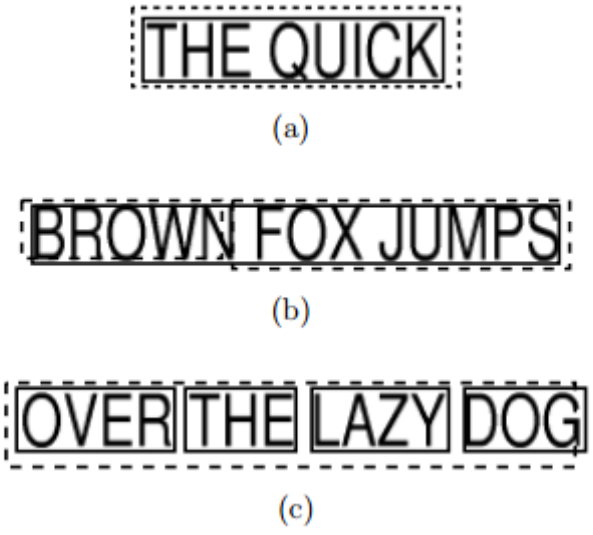

Рисунок 2.10 – метод соответствия прямоугольников [17]

Поле сплошной линии представляет реальную текстовую область, а поле пунктирной линии представляет результат алгоритма детектирования. Для полноты R (*recall*) и точности P (*precision*), рассчитываются по формуле (2.11) и (2.12).

$$R(G_i, D_i) = \frac{Area(G_i \cap D_i)}{Area(G_i)}, \tag{2.11}$$

$$P(G_i, D_i) = \frac{Area(G_i \cap D_i)}{Area(D_i)}, \tag{2.12}$$

где D — набор прямоугольных областей, предсказанный алгоритмом детектирования; G — коллекция истинных ответов.

Значение F-мера расчет по формуле (2.13).

$$F(G_i, D_i) = \frac{2 * R(G_i, D_i)P(G_i, D_i)}{R(G_i, D_i) + P(G_i, D_i)}, \tag{2.13}$$

В таблице 2.1 показаны результаты детектирования текста для моделей FCN, U-net и EAST.

Таблица 2.1 – Сравнение методов

Методы	**Точность**	**Полнота**	**F-мера**
FCN(ResNet-34)	0,794	0,739	0,765
U-net	0,868	0,825	0,846
EAST(ResNet-50)	0,798	0,729	0,762

Как видно из приведенной выше таблицы, среди трех моделей детектирования текста модель U-net имеет лучший эффект детектирования текста, а модель FCN (ResNet-34) и модель EAST (ResNet-50) имеют сопоставимые результаты детектирования. Если рассматривать только проблему детектирования текста, модель U-net может быть лучше, но выходные данные как модели U-net, так и модели FCN представляют собой замаскированные изображения текста. Если эти изображения маски должны использоваться в качестве входных данных для сети распознавания текста, требуется сложная обработка после детектирования (например, положение и угол текста должны определяться изображением маски). А модели FCN и U-net имеют низкую точность детектирования мелкого текста и подходят только для детектирования

четких и крупных текстов.

Поэтому, учитывая последующую задачу распознавания надписей на изображениях реальных сцен, для детектирования текста был окончательно выбран метод EAST.

2.6 Реализация финального метода

Архитектура предлагаемого метода состоит из двух основных частей: ветка детектирования текста и ветка распознавания текста. Общая архитектура показана на рисунке 2.11. Базовая сеть для детектирования ответвлений использует EAST (ResNet50), описанную в разделе 2.4, и ее сеть выводит карту оценок, геометрическую карту и угловую карту. Эти результаты также необходимо обработать и получить правильную текстовую область. После того, как ветвь детектирования текста получит несколько детектирования прямоугольника, ей необходимо объединить сильно повторяющиеся детектирования прямоугольника, но для EAST количество прямоугольников детектирования, которые ему необходимо объединить, часто исчисляется тысячами. Временная сложность традиционной NMS составляет $O(n^2)$, что требует очень много времени для EAST.

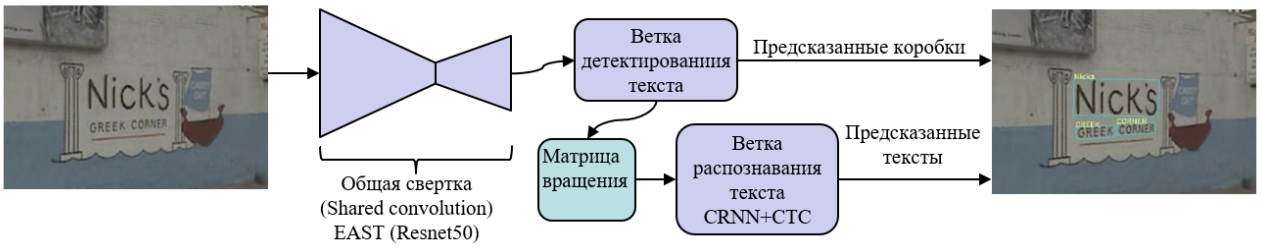

Рисунок 2.11 – общая архитектура

Для решения этой проблемы в статье предлагается Locality Aware NMS (LANMS) [17] с временной сложностью $O(n)$. Это связано с тем, что при детектировании текста смежная геометрия сильно коррелирует. Его основные этапы следующие:

1. Объедините набор всех предсказанных прямоугольников в соответствии с порогом, чтобы получить результирующее

предсказание;

2. Выполнить стандартные немаксимальные операции подавления в результирующем предсказании.

Его псевдокод показан на рисунке 2.12.

Algorithm 1 Locality-Aware NMS

```
function NMSLocality(geometries)
    S ← ∅, p ← ∅
    for g ∈ geometries in row first order do
        if p ≠ ∅ ∧ ShouldMerge(g, p) then
            p ← WeightedMerge(g, p)
        else
            if p ≠ ∅ then
                S ← S ∪ {p}
            end if
            p ← g
        end if
    end for
    if p ≠ ∅ then
        S ← S ∪ {p}
    end if
    return StandardNMS(S)
end function
```

Рисунок 2.12 – псевдокод для LANMS

2.6.1 Ветка распознавания

2.6.1.1 Поворот прямоугольника

Результат текстовой области, полученный ветвью детектирования, обрезается и обрабатывается как входное изображение сети CRNN. Поскольку получившиеся прямоугольники не все горизонтальны, их необходимо повернуть на координаты и углы прямоугольника.

Аффинное преобразование используется для реализации этого шага. Аффинное преобразование — это линейное преобразование из двумерных координат в двумерные, при котором сохраняется прямолинейность двумерного изображения (то есть преобразованная линия остается прямой линией, а дуга — дугой окружности) и параллельность (то есть относительное позиционное

отношение между двумерными изображениями остается неизменным, параллельные линии по-прежнему являются параллельными линиями, а угол между пересекающимися линиями остается неизменным). Пример аффинного преобразования показан на рисунке 2.13. Процесс его расчета выглядит следующим образом:

1. Вычисление повернутого прямоугольника;
2. Вычислить матрицу преобразования на основе центра и угла повернутого прямоугольника. Например, текстовая область, которую необходимо получить в этом методе, представляет собой горизонтальный прямоугольник, поэтому необходимо оценить угол, ширину и высоту, чтобы убедиться, что повернутый прямоугольник является горизонтальным прямоугольником;
3. Центр повернутого прямоугольника используется в качестве опорной точки для поворота всего изображения;
4. Поскольку координаты центральной точки прямоугольника после поворота не меняются, исходя из координат центральной точки, прямоугольная текстовая область обрезается по ширине и высоте.

Рисунок 2.13 – пример поворота прямоугольника

2.6.1.2 CRNN

Сетевая архитектура CRNN [19] показана на рисунке 1.7. Вход для этой

модели представляет собой изображение размером (32*128*1). В этой модели есть семь сверточных слоев, шесть из которых имеют размер ядра (3×3), а последний сверточный слой имеет размер ядра (2×2). А количество фильтров увеличилось с 64 слоев до 512. Также есть два слоя максимальных объединения размера (2×2), а затем два слоя максимальных объединения размера (2×1) для извлечения объектов с большей шириной для прогнозирования длинного текста. Затем выходные данные сверточного слоя сжимаются лямбда-функцией, чтобы сделать его совместимым со слоем LSTM [28].

Рекуррентные слои используют глубокую двунаправленную сеть LSTM для продолжения извлечения функций текстовой последовательности на основе функций свертки. Так называемая глубокая сеть RNN [28] относится к сети RNN с более чем двумя уровнями. Здесь используется стековая глубокая двунаправленная сеть RNN (рисунок 2.14). Так как выходная карта функций CNN имеет размер (1, 32, 512), максимальная продолжительность RNN составляет 32 временных входа, а каждый вектор-столбец входных данных равно 512.

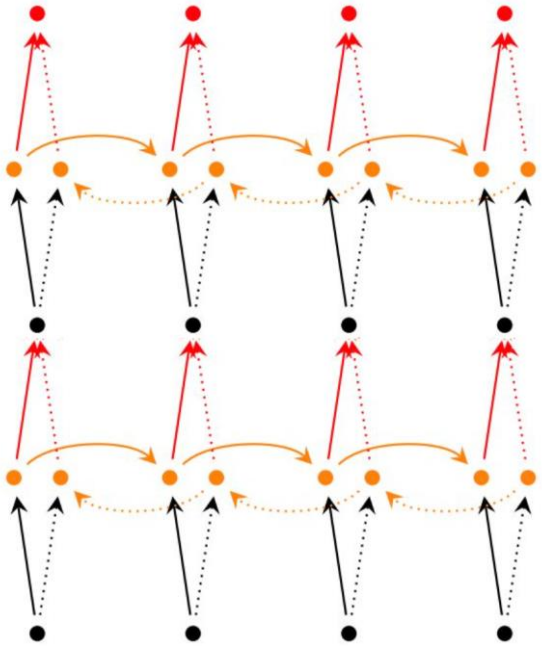

Рисунок 2.13 – глубокая двунаправленная сеть RNN в форме стека

Поскольку выход сети - прогнозируемая вероятность появления символа, поэтому функция активации softmax используется для сопоставления вывода

RNN с диапазоном [0, 1], что указывает на вероятность того, что символ правильный, рассчитывается по уравнению (2.14). Выходные данные модели (batch_size, 31, 63), где 63-общее количество выходных классов, содержащих прописные и строчные буквы и цифры 0-9.

$$\sigma(t) = \frac{1}{1 + e^{-t}}, \tag{2.14}$$

2.6.1.3 Коннекционистская временная классификация (CTC)

Для рекуррентных слоев, если используется общая потеря кросс-энтропии Softmax, каждый вывод столбца должен соответствовать элементу символа. Затем во время обучения каждое изображение-образец должно пометить положение каждого символа в изображении, а затем выровнять его с каждым столбцом карты объектов через принимающее поле CNN, чтобы получить метку, соответствующую выходным данным столбца для обучения. В реальной ситуации это очень сложно, а рабочая нагрузка очень велика. Поскольку количество символов в каждом образце отличается, стиль шрифта и размер шрифта различны, выходные данные каждого столбца могут не соответствовать каждому символу один к одному. Поэтому CTC [20] предлагает метод расчета потерь, не требующий выравнивания, который используется для обучения сети и широко используется при распознавании текста. Расчет по формуле (2.15).

$$L(S) = -\ln \prod_{(x,z)\in S} p(z|x) = -\sum_{(x,z)\in S} \ln(z|x), \tag{2.15}$$

где S - обучающая выборка.

3 Результаты модели

Для обучения модели используется GPU предоставленный Google colab pro — NVIDIA Tesla P100 PCIe 12GB и место для хранения, предоставляемое Google Cloud Drive.

3.1 Результаты детектирования текста для метода

Во входных изображениях используется исходный набор данных ICDAR 2015. Для слоя извлечения признаков ResNet50 используется для уменьшения разрешения изображения в 5 раз и извлечения признаков изображения в четыре раза. Карта признаков выводится после повышения частоты дискретизации и объединения признаков. Выходные данные сети - результат прогнозирования, составляющий четверть размера входного изображения. Поэтому перед входом в сеть размер исходного изображения необходимо отрегулировать до кратного 32. На рисунке 3.1 показаны результаты детектирования моделью текстовых областей после 600 эпохи.

Рисунок 3.1 – примеры результатов детектирования текста

Видно, что модель лучше детектирует четкий, крупный, высококонтрастный текст и какой-то мелкий, нечеткий текст не детектируется.

3.2 Результат вырезания области текста

Согласно описанию в разделе 2.6.1.1, если текст, детектированный на

изображении, должен быть подвергнут последующим этапам распознавания, Текстовые области в разных направлениях необходимо обрезать в соответствии с предсказанием прямоугольника. Примеры обрезанных текстовых изображений показаны на рис. 3.2.

Рисунок 3.2 – примеры обрезанных текстовых изображений

Видно, что обрезанные текстовые изображения имеют разный размер, а также есть изображения с размытым текстом, изображения со специальными шрифтами, изображения с наклонным текстом и изображения с низкой контрастностью. Все эти изображения будут изменены до одинакового размера и отправлены в CRNN для распознавания.

3.3 Результаты распознавания корейского и английского текста

В начале работы для сравнения вышеуказанных методов был выбран набор данных текстовых изображений сцены с корейским языком. Чтобы реализовать распознавание корейского текста, генератор данных для распознавания текста [29] использовался для создания набора данных для обучения сети распознавания текста. Генератор использовался для создания более 70 000 изображений, содержащих корейский, английский языки и числа, 70% из которых использовались для обучения, 20% для проверки и 10% для

тестирования. Название каждого изображения – этикетка. Набор данных показан на рисунке 3.3.

Рисунок 3.3 – примеры изображений с использованием генератора

В качестве сети, используемой для обучения, выбрана сеть CRNN, описанная в разделе 2.6.1.2. Размер выходного слоя заменяется на (31, 11234), где 11234 — общее количество выходных классов, содержащих цифры, прописные и строчные буквы английских и корейских букв. Оптимизатор Adam и функция потерь CTC используются для обучения сети. Модель достигла точности, показанной в таблице 3.1, после 240 эпохи. Где расчет точности — это показатель точности распознавания символов, то есть доля правильно распознанных символов в общем количестве символов.

Таблица 3.1 – Результат распознавания корейского текста

Модель	**Обучение**	**Проверка**	**Тестирование**
CRNN (основан на корейском)	0,989	0,919	0.907

Примеры результатов распознавания корейского текста показан на рисунке 3.4. Из результатов видно, что модель имеет высокую точность распознавания корейских и английских текстов. Однако между изображением, сгенерированным генератором данных для распознавания текста [29], и текстовым изображением, вырезанным из реального изображения сцены, имеется разрыв. А поскольку выходной слой содержит более 11 000 корейских символов, модель будет ошибочно распознавать размытые текстовые изображения на английском языке как корейские, что приведет к значительному

снижению точности распознавания текстовых изображений, вырезанных из изображений в наборе данных ICDAR 2015. Однако, поскольку модель EAST не обучена на реальных изображениях сцены с корейским текстом, ее эффект детектирования корейского текста очень плохой, что также приводит к очень низкой точности распознавания корейского текста.

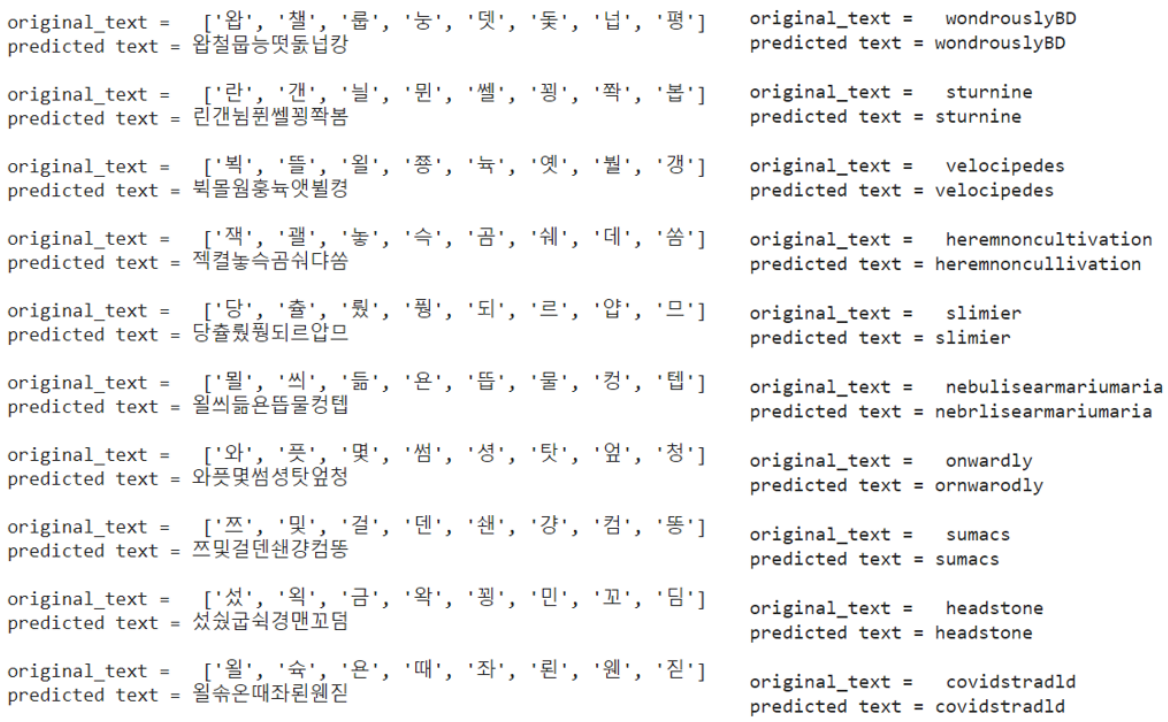

Рисунок 3.4 – примеры результатов

3.4 Результаты распознавания текста для модели EAST

Для обучения и проверки сети ветки распознавания текста используется синтетический набор данных MJSynth. Изображения в этой модели предварительно обрабатываются перед подачей в модель, их размер изменяется до того же размера и обрабатывается как изображения в градациях серого. Цветные изображения разных размеров, обрезанные в предыдущем разделе, также будут изменены до того же размера и обработаны как изображения в градациях серого, и использовать их в качестве испытуемых для ветки распознавания текста. На рисунках 3.5 и 3.6 показаны хорошие и худшие результаты, полученные с помощью используемого метода, соответственно.

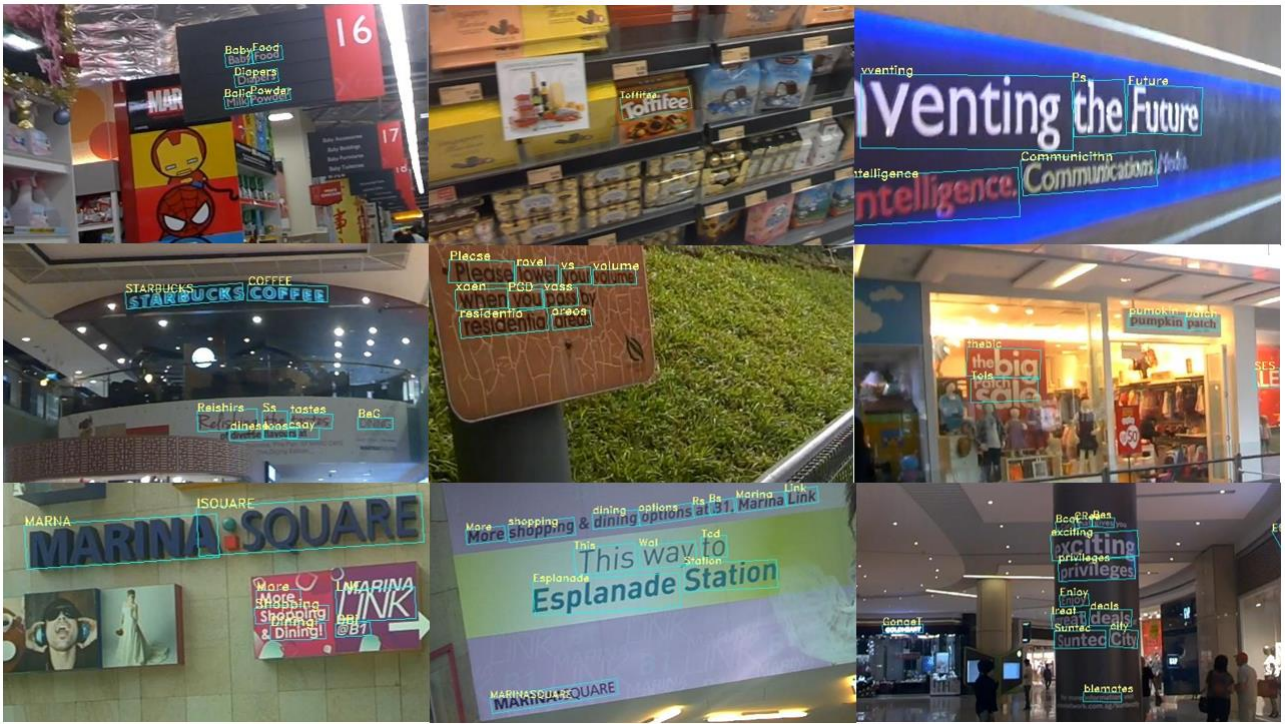

Рисунок 3.5 – примеры хороших результатов

Рисунок 3.6 – примеры худших результатов

Как видно из рисунка, на эффект распознавания текста ветвью распознавания текста в большей степени влияет ветвь детектирования текста. Ветка распознавания текста лучше распознает четкий, крупный, высококонтрастный текст. И эффект распознавания хуже для наклонного текста, текста со специальными шрифтами, размытого текста, мелкого и изогнутого текста.

3.5 Результат точности метода распознавания текста

Метод, используемый для вычисления точности ветви детектирования текста, описан в разделе 2.5. Для точности ветки распознавания, ее точность обучения и проверки основаны на наборе данных MJSynth, точность теста на основе набора данных ICDAR 2015. Расчет точности — это показатель точности распознавания символов, то есть доля правильно распознанных символов в общем количестве символов.

Точность метода и сравнение с EAST (PVANet) и EAST (VGG16), использованными первоначальным автором EAST, показаны в таблице 3.1.

Таблица 3.1 – Сравнение методов

Ветки	**Модели**	**Обучение**	**Проверка**	**Тестирование**		
		Точность	Точность	Точность	Полнота	F-мера
Детектирование	EAST (PVANet) [17]	\	\	**0,806**	0,713	**0,782**
	EAST (VGG16) [17]	\	\	0,805	0,728	0,764
	EAST (ResNet50)	\	\	0,798	**0.729**	0.762
Распознавание (CRNN)		0,933	0,918	0,729	\	\

По полученным результатам можно сделать вывод, что метод соответствует требованиям по детектированию и распознаванию надписей на изображениях реальных сцен.

4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

4.1 Предпроектный анализ

Целью выпускной квалификационной работы является разработка нейросетевого алгоритма детектирования и распознавания надписей на изображениях реальных сцен.

Реализация алгоритма подразумевает собой написание сверточной нейронной сети EAST и CRNN. Для обучения сети используется находящаяся в открытом доступе база изображения ICDAR_2015. Результатом работы алгоритма для каждого изображения является текстовая область на изображении, на которой отмечена текстовая информация.

Разработанное программное обеспечение может быть использовано в исследовательских учреждениях, предприятиях или независимыми исследователями в качестве детектирования и распознавания надписей на изображениях реальных сцен.

Для эффективного использования научного потенциала научно-исследовательского проекта необходимо не только провести исследовательскую работу, но и провести анализ исследования с точки зрения финансового менеджмента. Задачами данного раздела являются:

1. Определение перспективности технологии с помощью технологии QuaD;
2. Определение сильных и слабых сторон, а также возможностей и угроз с помощью SWOT-анализа;
3. Планирование научно-исследовательских работ;
4. Бюджет научно-технической разработки;
5. Расчет прибыли, НДС и Цена разработки НИР.

Технология QuaD представляет собой гибкий инструмент измерения характеристик, описывающих качество новой разработки и её перспективность

на рынке и позволяющие принимать решение целесообразности вложения денежных средств в научно-исследовательский проект.

Ввиду отсутствия прямых конкурентных алгоритмов, применяющих как используемую структуру сети, так и используемую базу изображений, сравнение производится с решениями, использующими только указанную структуру сети. Были выбраны алгоритмы, описанные в следующих статьях:

1. «TextBoxes++: A Single-Shot Oriented Scene Text Detector», основанная на оригинальной сети “TextBoxes”, было добавлено детектирование наклоненного текста [30];
2. «FOTS: Fast Oriented Text Spotting with a Unified Network», в котором используется для сквозного детектирования и распознавания текста сцены [31];
3. «Detecting Text in Natural Image with Connectionist Text Proposal Network», в котором используется детектирования горизонтального текста [32].

Были выбраны следующие критерии конкурентоспособности:

- Доступность программного кода. Наличие в открытом доступе описанного в статье алгоритма в виде программного кода или структуры;
- Точность обучения алгоритма. Значения конечных показателей точности обученного сети на выборках данных;
- Универсальность относительно входных данных. Возможность подачи на вход обученной сети различных входных данных, например, изображения реальных сцен, отсканированные страницы документов или рукописных текстов;
- Детектирование и распознавание. Может детектировать и распознавать текст;
- Потребность в вычислительных ресурсах. Уровень требований к вычислительной машине, на которой производится обучение сети.

Был оценен вес каждого критерия, после чего конкурентные решения

были оценены по выбранным факторам по 10 бальной шкале путем осуществления экспертной оценки. Затем, на основании полученных значений, были подсчитаны итоговые оценки конкурентоспособности. В таблице 4.1 представлена оценочная карта технологии QuaD.

Таблица 4.1 – Оценочная карта технологии QuaD

Критерии оценки	Вес критерия	Баллы				Конкурентоспособность			
		$Б_ф$	$Б_{к1}$	$Б_{к2}$	$Б_{к3}$	$К_ф$	$К_{к1}$	$К_{к2}$	$К_{к3}$
Доступность	0,2	9	6	3	6	1,8	1,2	0,6	1,2
Точность	0,2	8	8	9	7	1,6	1,6	1,8	1,4
Универсальность	0,2	6	5	5	4	1,2	1,0	1,0	0,8
Детектирование и распознавание	0,25	9	2	9	1	2,25	0,5	2,25	0,25
Потребность	0,15	8	6	6	8	1,2	0,9	0,9	1,2
Итого	1	**40**	27	32	26	**8,05**	5,2	6,55	4,85

Значение качества по технологии QuaD составило 0,8525. Можно сказать, что подобная разработка может считаться перспективной, так как у рассматриваемой разработки высокие показатели по всем наиболее важным критериям и может выполнять сквозное детектирование и распознавание текста сцены.

SWOT – Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы) – это комплексный анализ научно-исследовательского проекта. Такой анализ применяют для исследования внешней и внутренней среды проекта. Первый этап заключается в описании сильных и слабых сторон проекта, в выявлении возможностей и угроз для реализации проекта, которые проявились или могут появиться в его внешней среде. Составленная матрица SWOT представлена в таблице 4.2.

В рамках SWOT анализа были рассмотрены по четыре слабых и сильных сторон проекта, а также возможности и угрозы. Указаны стратегии, которые будут ликвидировать слабые стороны разработки и бороться с угрозами. И выявляются соответствия сильных и слабых сторон проекта условиям внешней среды. Это необходимо для оценки необходимости каких-либо изменений проекта.

Таблица 4.2 – SWOT-анализ

		Внутренние факторы	
		Сильные стороны: С1. Доступность для широкой аудитории. С2. Высокая точность. С3. Воспроизводимость. С4. Большое количество библиотек с открытым исходным кодом.	Слабые стороны: Сл1. Потребность в вычислительных ресурсах при переобучении. Сл2. Сложность переобучения для конечного пользователя. Сл3. Высокая трудоемкость методов исследования.
Внешние факторы	Возможности: В1. Расширение для обработки пакетов изображений. В2. Переобучение нейронной сети на новых базах данных изображений.	В1С1С3. Реализация автоматического сохранения файла с данными о найденных текстовых областях при обработке пакета изображений. В2С1С2С3. Расширение функциональности в зависимости от нужд потребителя. В2С1С2С4. Повышение точности путём использования новых баз данных изображений.	В1Сл1. Модульное программирование частей проекта для повторного использования. В2Сл2. Переобучение нейронной сети для распознавания текста в найденных текстовых областях. В1Сл1Сл3. Реализация возможности обновления вместо повторного скачивания программы с переобученной сетью.
	Угрозы: У1. Невостребованность системы. У2. Конкуренция с аналогами.	У1С1С3. Публикация статей с описанием программы для привлечения внимания. У2С2С4. Отслеживание конкурентных решений и улучшение возможностей на их основании	У1У2Сл1Сл2Сл3. Отслеживание и применение новых технологий. У2Сл1Сл2. Оптимизация переобучения для сокращения требуемых ресурсов.

Для определения комбинаций взаимосвязей областей матрицы SWOT-таблицы была составлена интерактивная матрица проекта, представленная в таблице 4.3.

Таблица 4.3 – Интерактивная матрица проекта

		Сильные стороны проекта				Слабые стороны проекта		
		С1	С2	С3	С4	Сл1	Сл2	Сл3
Возможности	В1	+		+		+		+
	В2	+	+	+	+		+	
Угрозы	У1	+		+		+	+	+
	У2		+		+	+	+	+

В результате проведения SWOT-анализа были предложены варианты улучшения разработки на основании сильных сторон и возможностей, нейтрализации слабых сторон с помощью возможностей и нейтрализации угроз с помощью сильных сторон, а также способы избегания возможных потерь от угроз при наложении на слабые стороны. Общим направлением среди всех вариантов можно считать улучшение технологий на основании данных, полученных при наблюдении за конкурентами, или на основании возможных требований пользователей.

4.2 Планирование управления научно-техническим проектом

Для организации и систематизации работы был сформирован ряд работ и назначены исполнители для каждого этапа работы. Для реализации проекта необходимо 2 исполнителя – научный руководитель (НР), инженер (И). Для определения трудоемкости работ использовались следующие показатели:

- Ожидаемое значение трудоемкости ($t_{ож}$), выполнено по формуле (4.1):

$$t_{ож} = \frac{3 \times t_{min} + 2 \times t_{max}}{5}, \quad (4.1)$$

где t_{min} – минимальная продолжительность работы, дни.,

t_{max}– максимальная продолжительность работы, дни.

- Продолжительность каждой работы в рабочих днях (Tр),
- Продолжительность каждой работы в календарных днях (Tк),

• Коэффициент календарности (kкал).

Согласно производственному календарю в 2021 году 365 календарных дней, из них для 6-дневной рабочей недели 66, а для 5-дневной рабочей недели 118 выходных или праздничных дней. Расчеты по трудоемкости выполнения работ представлены в таблице 4.4.

Таблица 4.4 – Результаты трудозатрат

№	Вид	Исполн-ители	Трудоемкость работ, дни			Длительность работ, дни	
			t_{min}	t_{max}	$t_{ож}$	$T_р$	$T_к$
1	Выбор направления исследования	И	1	2	1,4	1	1
2	Составление и утверждение темы	НР	1	2	1,4	1	1
		И	1	2	1,4	1	1
3	Составление календарного плана графика	НР	1	2	1,4	1	1
		И	1	2	1,4	1	1
4	Изучение литературы	И	8	10	8,8	9	13
5	Выбор алгоритмов и подготовка данных	И	8	10	8,8	9	13
6	Проектирование программного обеспечения	И	17	25	20,2	20	30
7	Разработка программного обеспечения	И	22	35	27,2	27	40
8	Тестирование программного обеспечения и анализ полученных результатов	НР	3	4	3,4	2	2
		И	9	10	9,4	5	7
9	Согласование выполненной работы с научным руководителем	НР	2	3	2,4	1	1
		И	2	3	2,4	1	1
10	Подведение итогов, оформление работы	И	8	10	8,8	9	13
Итого		НР				5	5
		И				83	120

Диаграмма Ганта – это тип столбчатых диаграмм (гистограмм), который используется для иллюстрации календарного плана проекта, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ. Диаграмма Ганта, построенная по рассчитанным значениям, представлена на

рисунке 4.1.

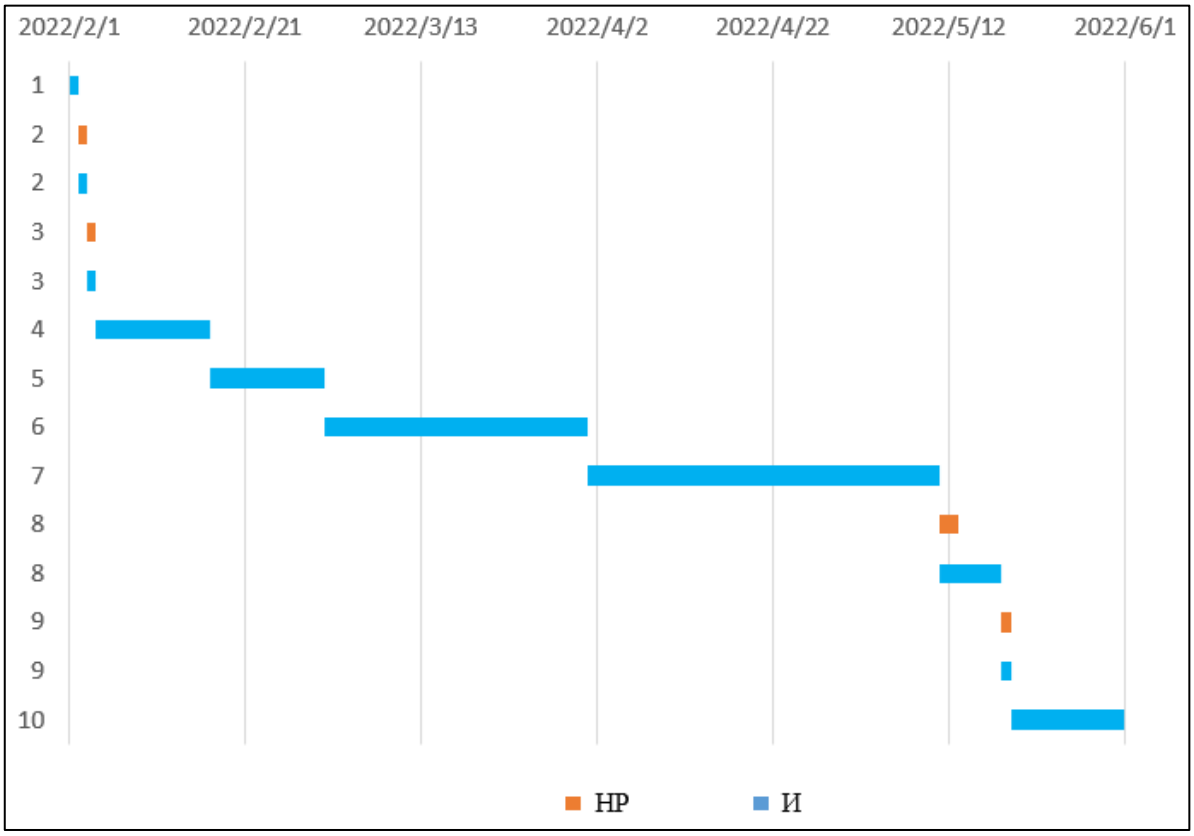

Рисунок 4.1 – Диаграмма Ганта

При планировании бюджета научного исследования должно быть обеспечено полное и достоверное отражение всех видов планируемых расходов, необходимых для его выполнения.

В статью расходов «Специальное оборудование» включает сумму, необходимую для обеспечения амортизации используемого оборудования, а также сумму, необходимую для приобретения средств разработки и среды. Расчет амортизации персонального компьютера, используемого при написании работы: первоначальная стоимость персонального ноутбука инженера составляет 50 000 рублей, при нормальных обстоятельствах годовая норма износа ноутбуков составляет 18%; планируется использовать персональный компьютер для написания работы в течение 4 месяцев. Стоимость приобретения среды разработки Google Colab Pro составляет 2500 рублей. Тогда амортизационная стоимость ноутбука за четыре месяца составляет:

$$A_M = \frac{50000 \times 0{,}18}{12} \times 4 = 3000 \text{ руб.}, \quad (4.2)$$

Итоговая сумма затрат на амортизацию составила 3000 рублей.

В статью затрат «Основная заработная плата исполнителей темы» входят основная заработная плата с учетом премий и доплат для исполнителей работы: инженера (студента) и научного руководителя.

Для должности доцента и степень кандидата наук месячный оклад составляет 35 111,5 рубля, для инженера был взят оклад ассистента без ученой степени – 22 695 рублей. Основная заработная плата от предприятия рассчитывается по следующей формуле:

$$З_{ОСН} = Т_Р \times З_Д, \quad (4.3)$$

где $З_д$ – среднедневная заработная плата работника, руб; $Т_р$ – продолжительность работ, выполняемых научно-техническим работникам, работающим.

Среднедневная заработная плата рассчитывается по формуле:

$$З_Д = \frac{З_М \times М}{F_Д}, \quad (4.4)$$

где $З_м$ – месячный должностной оклад работника, руб.; М – количество месяцев работы без отпуска в течение года:

для 5-дневной рабочей недели М = 11,2 месяца; для 6-дневной рабочей недели М = 10,4 месяца.

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

В таблице 4.5 показаны количества календарных, нерабочих и праздничных дней, дней, пришедшихся на потерю рабочего времени и действительный годовой фонд рабочего времени.

Таблица 4.5 – Баланс рабочего времени

Показатели рабочего времени	**Дни**	
	Научный руководитель	Инженер
Календарные дни	365	365
Нерабочие дни	66	118
Потери рабочего времени	56	24

Действительный годовой фонд рабочего времени	243	223

Месячный должностной оклад работника рассчитывается по формуле:

$$З_М = З_{ОК} \times k_р, \quad (4.5)$$

где $З_{ок}$ – размер оклада руководителя (указываете должность руководителя); $k_р$ - районный коэффициент, равный 1,3 (для Томска).

Таблица 4.6 - Расчёт основной заработной платы

Исполнители	$З_{ок}$, руб.	$k_р$	$З_м$, руб.	$З_{дн}$, руб.	$Т_р$, раб. дн.	$З_{осн}$, руб.
Инженер	22695,00	1,3	29503,50	1481,79	83	122988,58
Научный руководитель	35111,50	1,3	45644,95	1953,53	**5**	9767,64
Итого:						132756,22

Дополнительная заработная плата включает оплату за непроработанное время (очередной и учебный отпуск, выполнение государственных обязанностей, выплата вознаграждений за выслугу лег и т.п.) и рассчитывается исходя из 10-15% от основной заработной платы, работников, непосредственно участвующих в выполнение темы:

$$З_{ДОП} = k_{ДОП} \times З_{ОСН}, \quad (4.6)$$

где $З_{доп}$ – дополнительная заработная плата, руб.;

$k_{доп}$ – коэффициент дополнительной зарплаты ($k_{доп}$ =0,12);

$З_{осн}$ – основная заработная плата, руб.

Для руководителя:

$$З_{ДОП} = 0{,}12 \times 9767{,}64 = \ 1172{,}12 \text{ руб.}, \quad (4.7)$$

Для инженера:

$$З_{ДОП} = 0{,}12 \times 122988{,}58 = \ 14758{,}63 \text{ руб.}, \quad (4.8)$$

Итого 15930,75 рублей.

Отчисления во внебюджетные фонды рассчитываются как:

$$З_{внеб} = \ k_{внеб} \times (З_{ОСН} + З_{ДОП}), \quad (4.9)$$

где $k_{внеб}$ – коэффициент отчислений во внебюджетные фонды, для учреждений, осуществляющих образовательную и научную деятельность в 2014 году водится

пониженная ставка – 27%.;

$З_{осн}$ – основная заработная плата;

$З_{доп}$ – дополнительная заработная плата.

Для руководителя:

$$З_{внеб} = 0{,}27 \times (9767{,}64 + 1172{,}12) = 2953{,}74 \text{ руб.}, \quad (4.10)$$

Для инженера:

$$З_{внеб} = 0{,}27 \times (122988{,}58 + 14758{,}63) = 37191{,}75 \text{ руб.}, \quad (4.11)$$

Итого 40145,49 рублей.

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов, оплата услуг связи, электроэнергии и т.д. Их величина определяется согласно следующей формуле:

$$З_{накл} = k_{накл} \times (\text{сумма статей расходов}), \quad (4.12)$$

где $k_{накл}$ – коэффициент накладных расходов. Накладные расходы составляют 16% от суммы основной и дополнительной заработной платы, работников, непосредственно участвующих в выполнение темы.

$$З_{накл} = 0{,}16 \times 194484{,}14 = 31093{,}19, \quad (4.13)$$

Рассчитанные величины затрат научно-исследовательской работы являются основой для формирования бюджета затрат проекта. Результаты составления итогового бюджета разработки представлены в таблице 4.7.

Таблица 4.7 – Бюджет затрат НТИ

Наименование	**Сумма, руб.**
Затраты на Google Colab Pro	2500
Затраты на амортизацию	3000
Осн зп	132756,22
Доп зп	15930,75
Внеб зп	40145,49
Накл рас	31093,19
Итого	225425,65

В результате было получено, что бюджет на разработку НТИ составит 225425,65 руб

4.3 Выводы по разделу

Исследование проводилось при помощи ЭВМ. Данный исследовательский проект является перспективным согласно расчету с использованием технологии QuaD. Во время проведения SWOT анализа были выделены сильные и слабые стороны проекта, а также возможности и угрозы его реализации. Для определения комбинаций взаимосвязей областей матрицы SWOT-таблицы была составлена интерактивная матрица проекта.

В рамках процессов инициации определены внутренние и внешние заинтересованные стороны проекта с их ожиданиями от проекта, цели и результат проекта и были определены структуры и трудоемкости работ.

План проекта представлен на диаграмме Ганта, из которого видно какой исполнитель (студент или руководитель) какой вид работ осуществлял и в течении какого количества дней.

Был сформирован бюджет разработки, который включил в себя затраты на амортизацию, заработную плату, отчисления во внебюджетные фонды и накладные расходы. Общий бюджет проекта получился равным 225425,65 рублей.

5 СОЦИАЛЬНАЯ ОТВЕТСВЕННОСТЬ

Целью выпускной квалификационной работы является разработка нейросетевого алгоритма детектирования и распознавания надписей на изображениях реальных сцен.

Реализация алгоритма подразумевает собой написание сверточной нейронной сети EAST и CRNN. Для обучения сети используется находящаяся в открытом доступе база изображения ICDAR_2015. Результатом работы алгоритма для каждого изображения является текстовая область на изображении, на которой отмечена текстовая информация.

Разработанное программное обеспечение может быть использовано в исследовательских учреждениях, предприятиях или независимыми исследователями в качестве детектирования и распознавания надписей на изображениях реальных сцен.

Рабочей зоной при разработке алгоритма является рабочее место для выполнения работ сидя с использованием персональной электронно-вычислительной машины.

5.1 Правовые и организационные вопросы обеспечения безопасности

5.1.1 Специальные правовые нормы трудового законодательства

В данном разделе рассмотрены основные нормативные документы, выполнение которых необходимо для безопасного исполнения рабочих обязанностей диспетчером, который проводит большую часть времени за компьютером.

Правовые нормы регулирует специальный государственный орган. Основные требования описаны в трудовом кодексе Российской Федерации. Время продолжительности рабочего времени не может превышать 40 часов в неделю. В течение рабочего дня работнику должен быть предоставлен перерыв

для отдыха и питания продолжительностью не более двух часов и не менее 30 минут.

Первопричиной всех травм и заболеваний, связанных с процессом труда, является неблагоприятное воздействие на организм занятого трудом человека тех или иных факторов производственной среды и трудового процесса. Это воздействие, приводящее в различных обстоятельствах к различным результирующим последствиям, зависит от наличия в условиях труда того или иного фактора, его потенциально неблагоприятных для организма человека свойств, возможности его прямого или опосредованного действия на организм, характера реагирования организма в зависимости от интенсивности и длительности воздействия (экспозиции) данного фактора [34]. Согласно ТК: Федеральный закон от 30 декабря 2001г. N 197-ФЗ работник имеет право на [33]:

- рабочее место, соответствующее требованиям охраны труда;
- обязательное социальное страхование от несчастных случаев на производстве и профессиональных заболеваний в соответствии с федеральным законом;
- 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;
- внеочередной медицинский осмотр в соответствии с медицинскими рекомендациями с сохранением за ним места работы (должности) и среднего заработка во время прохождения указанного медицинского осмотра.

Согласно ТК: Федеральный закон от 30 декабря 2001г. N 197-ФЗ устанавливает обязательства работодателя, некоторые из них [33]:

- соблюдать трудовое законодательство и иные нормативные правовые

акты, содержащие нормы трудового права, локальные нормативные акты, условия коллективного договора, соглашений и трудовых договоров;

- предоставлять сотрудникам работу, обусловленную трудовым договором;
- обеспечивать безопасность и условия труда, соответствующие государственным нормативным требованиям охраны труда;
- обеспечивать работников оборудованием, инструментами, технической документацией и иными средствами, необходимыми для исполнения ими трудовых обязанностей; обеспечивать работникам равную оплату труда;

5.2 Организационные мероприятия при компоновке рабочей зоны

Исходя из общих принципов организации рабочего места, основными элементами рабочего места программиста являются: рабочий стол, рабочий стул (кресло), системный блок, монитор, клавиатура, мышь. При взаимном расположении элементов рабочего места необходимо учитывать:

- рабочую позу человека-оператора;
- пространство для размещения человека-оператора;
- возможность обзора элементов рабочего места;
- возможность обзора пространства за пределами рабочего места;
- возможность ведения записей, размещения документации и материалов, используемых человеком-оператором [35].

Предъявляют требования к организации рабочих мест пользователей персональных электронных средств обучения (ЭСО): должна обеспечиваться зрительная дистанция до экрана не менее 50 см. Линейные размеры (диагональ) экрана ЭСО должны соответствовать гигиеническим нормативам: минимальный размер монитора персонального компьютера и ноутбука - не менее 39,6 см [36]. Рабочее место при выполнении работ сидя регламентировано ГОСТ 12.2.032-78.

Конструкция рабочего места должна обеспечивать оптимальное положение работающего и обеспечивать трудовые операции в пределах зоны досягаемости моторного поля. Оптимальное положение может быть достигнуто регулированием высоты рабочей поверхности, сиденья и пространства для ног [37]. Кресло должно обеспечивать человеку-оператору соответствующую характеру и условиям труда физиологически рациональную рабочую позу. Кресло должно обеспечивать длительное поддержание основной рабочей позы в процессе трудовой деятельности. При невозможности покинуть рабочее место длительное время конструкция кресла должна обеспечивать условия для отдыха человека-оператора в кресле [38].

Также регламентированы требования к размещению органов управления: в данном контексте клавиатура и мышь. В соответствии с разделение рабочего пространства на зоны данные органы управления являются наиболее важными и располагаются в зоне, соответствующей горизонтальной плоскости под углом ±30° от сагиттальной плоскости. Размещение средств отображения информации касается расположения монитора, как главного средства отображения. Монитор является часто используемым средством, однако он не требует точного и быстрого считывания показаний, поэтому допускается располагать в вертикальной плоскости под углом ±30° от нормальной линии взгляда и в горизонтальной плоскости под углом ±30° от сагиттальной плоскости [37].

5.3 Производственная безопасность

Для обеспечения производственной безопасности необходимо проанализировать воздействия на человека вредных и опасных производственных факторов, которые могут возникать при разработке или эксплуатации проекта. Производственный фактор считается вредным, если воздействие этого фактора на работника может привести к его заболеванию. Производственный фактор считается опасным, если его воздействие на работника может привести к его травме. Все производственные факторы

классифицируются по группам элементов: физические, химические, биологические и психофизические. Для данной работы целесообразно рассмотреть физические и психофизические вредные и опасные факторы производства, характерные как для рабочей зоны программиста, как разработчика рассматриваемой в данной работе системы, так и для рабочей зоны пользователя готового продукта. В таблице 5.1 приведены возможные опасные и вредные факторы.

Таблица 5.1 – Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работы			Нормативные Документы
	Разработка	Изготовление	Эксплуатация	
1. Отсутствие или недостаток необходимого искусственного освещения	+	+	+	СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95 [39]
2. Отклонение показателей микроклимата в закрытом помещении	+	+	+	СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания [40]. ГОСТ 12.1.005-88 Система стандартов безопасности труда (ССБТ). Общие санитарно-гигиенические требования к воздуху рабочей зоны [41].
3. Перенапряжение зрительного анализатора	+	+	+	ГОСТ Р 50923-96. Дисплеи. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения [37].
4. Повышенный уровень шума на рабочем месте	+	+	+	ГОСТ 12.1.003-83 Система стандартов безопасности труда (ССБТ). Шум. Общие требования безопасности [42].
5. Повышенное значение напряжения	+	+	+	ГОСТ 12.1.030-81 Система стандартов безопасности труда (ССБТ). Электробезопасность. Защитное заземление. Зануление [43]. ГОСТ 12.1.038-82 Система стандартов безопасности труда (ССБТ). Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов [44].

5.3.1 Вредные производственные факторы

5.3.1.1 Недостаточная освещенность рабочей зоны

Освещение на рабочем месте в значительной степени влияет на восприятие получаемой информации. Даже при работе с экраном компьютера необходимо дополнительное освещение. В качестве основного освещение было использовано искусственное освещение, которое нормируется в соответствии с СП 52.13330.2016 [39]. Данный тип освещения может применяться как индивидуально, так и совместно с естественным освещением, тогда освещение называется совмещенным.

Работа с компьютером подразумевает постоянный зрительный контакт с дисплеем ПЭВМ и занимает от 80 % рабочего времени. Недостаточность освещения снижает производительность труда, увеличивает утомляемость и количество допускаемых ошибок, а также может привести к появлению профессиональных болезней зрения. Существуют общие требования и рекомендации к организации освещения на рабочем месте: Искусственное освещение в помещениях для эксплуатации ПЭВМ должно осуществляться системой общего равномерного освещения. Рабочие места следует размещать таким образом, чтобы естественный свет падал преимущественно слева, а дисплеи монитора были ориентированы боковой стороной к световым проемам.

5.3.1.1.1 Расчет искусственного освещения

Рассчитывается система общего равномерного освещения для помещения с размерами: длина A = 8,0 м, ширина B = 4,2 м, высота H = 2,5 м. Высота рабочей поверхности h_{pn} = 0,75 м. Характеристики зрительной работы в компьютерных залах соответствуют очень высокой точности для рабочего стола при относительной продолжительности зрительной работы при направлении зрения на рабочую поверхность менее 70%. Тогда, освещенность рабочих поверхностей при общем освещении должна быть E = 300 лк.

В комнате есть окно, Размер помещения составляет S = 33,6 м^2. Стены оклеены светлыми обоями, соответственно, коэффициент отражения стен R_c = 30%, потолок белый матовый натяжной, поэтому приравниваем его к побеленному с коэффициентом отражения $R_п$ = 50%. Рабочее пространство относится к помещениям с малым выделением пыли, поэтому коэффициент запаса K =1,5.

В рабочем помещении работник должен находиться в течении всего рабочего дня, а также производить работу очень высокой точности, которая, однако не имеет повышенных требований к передаче цветов. Также помещение не обладает высокими полками. Вследствие всех перечисленных условий наилучшим выбором для общего освещения будут люминесцентные лампы белой цветности. Для люминесцентных ламп коэффициент неравномерности освещения берётся Z = 1,1 (для люминесцентных ламп).

Для нормальных помещений с хорошим отражением потолка и стен выбираются открытые двухламповые светильники типа ОД, ОДОР, ШОД, ОДО или ООД. Так как высота помещения 2,5 м, то в соответствии с наименьшей допустимой высотой подвеса над полом необходимо выбрать тип светильника ШОД. Наивыгоднейшее расположение светильников данного типа берется с критерием λ = 1,1- 1,3. Световой поток одной лампы Ф = 2500 лм. Расстояние светильников от перекрытия h_c = 0,1 м, так чтобы высота светильника над полом не превышала наименьшую допустимой высотой подвеса.

Тогда расчётная высота светильника над рабочей поверхностью вычисляется по формуле 5.1:

$$h = H - h_c - h_{\text{рп}} = 2{,}5 - 0{,}1 - 0{,}75 = 1{,}65 \text{ м.}, \quad (5.1)$$

Рассчитали индекс помещения i по формуле 5.2:

$$i = \frac{S}{H \times (A + B)} = \frac{33{,}6}{2{,}5 \times (8 + 4{,}2)} = 1{,}10, \quad (5.2)$$

где S – площадь помещения; H – высота помещения; A – длина помещения; B – ширина помещения.

Получили в результате значение i = 1,10. Исходя из значения индекса помещения определили коэффициент использования светового потока η = 46 %. Требуемое количество люминесцентных ламп по формуле 5.3:

$$N = \frac{E \times S \times K \times Z \times 100\%}{\Phi \times \eta}, \quad (5.3)$$

где Ф – световой поток; E – освещенность рабочей поверхности от систем общего освещения по СП 52.13330.2016; Z – коэффициент неравномерности освещения; N – количество рядов светильников

$$N = \frac{300 \times 33{,}6 \times 1{,}5 \times 1{,}1 \times 100\%}{2500 \times 46\%} = 14{,}5, \quad (5.4)$$

Получили количество светильников, равное 14,5. Округлив до ближайшего целого, получим 15 ламп. Такое количество огней можно получить, применив 3 ряда огней по 5 огней в ряд. На основании полученного значения рассчитали фактическое значение освещения в помещении по формуле 5.5:

$$E_{\text{факт}} = \frac{\Phi \times N \times \eta}{S \times K \times Z \times 100\%}, \quad (5.5)$$

где Ф – световой поток; E – освещенность рабочей поверхности от систем общего освещения; Z – коэффициент неравномерности освещения; N – количество рядов светильников.

$$E_{\text{факт}} = \frac{2500 \times 15 \times 46\%}{33{,}6 \times 1{,}5 \times 1{,}1 \times 100\%} = 311, \quad (5.6)$$

В результате получили значение $E_{\text{факт}}$ = 311 лк. На основании проделанных расчетов можно сделать вывод, исходя из данных таблицы 4.1 СП 52.13330.2016, что помещение удовлетворяет нормам освещения.

Мощность белой люминесцентной лампы составляет 28 Вт, поэтому общая мощность, необходимая для освещения, рассчитывается по формуле 5.7:

$$M_{\text{о}} = M_{\text{л}} \times N = 28 \times 15 = 420 \text{ Вт}, \quad (5.7)$$

где $M_{\text{л}}$ — мощность одной лампы, а N — количество ламп.

5.3.1.2 Повышенный уровень электромагнитных излучений

При работе с компьютерной техникой неизбежно возникновение

электромагнитных полей. Электромагнитные поля воздействуют на нервную систему человека, чувствительную к электрическим сигналам. Данное воздействие может вызывать повышение температуры тела, что вредно для тканей со слаборазвитой сосудистой системой или недостаточным кровообращением. Также при действии поля происходит поляризация макромолекул тканей и ориентация их параллельно электрическим силовым линиям, что может привести к изменению их свойств: нарушению функций сердечно-сосудистой системы и обмена веществ, уменьшению количество эритроцитов в крови. Степень воздействия электромагнитных полей на организм человека зависит от расстояния до источника, диапазона частот излучения, интенсивности воздействия, продолжительности, характера и режима облучения, размера облучаемой поверхности и особенностей организма. В таблице 5,2 указаны допустимые уровни электромагнитного поля, создаваемым персональным компьютером.

Таблица 5.2 –Допустимые уровни ЭМП, создаваемых ПК

Наименование параметров		ВДУ ЭМП
Напряженность электрического поля	в диапазоне частот 5 Гц - 2 кГц	25 В/м
	в диапазоне частот 2 кГц - 400 кГц	2,5 В/м
Плотность магнитного потока	в диапазоне частот 5 Гц - 2 кГц	250 нТл
	в диапазоне частот 2 кГц - 400 кГц	25 нТл
Электростатический потенциал экрана видеомонитора		500 В

Сила влияния электромагнитного поля уменьшается по мере удаления от его источника, поэтому одним из вариантов снижения влияния излучения на работника является расположение источников на некотором расстоянии от рабочего места и места отдыха. Другим вариантом защиты является экранирование источника излучения и/или рабочего места.

5.3.1.3 Отклонение показателей микроклимата

Микроклимат определяется действующими на организм человека показателями температуры, влажности и скорости движения воздуха. Длительное воздействие неблагоприятных условий на человека ухудшает его самочувствие, снижает производительность труда и приводит к заболеваниям. Оптимальными для микроклимата жилых и общественных помещений в теплое время года считаются: температура воздуха 22-25°С, относительная влажность 40-60 %, скорость движения воздуха не более 0,25 м/с; в холодное время года эти показатели составляют соответственно 20-22° С, 30-45 % и 0,1-0,15 м/с [8].

К параметрам, характеризующим микроклимат в производственных помещениях, относятся:

- Температура воздуха (t, °С);
- Температура поверхностей (t, °С);
- Относительная влажность воздуха (φ, %);
- Скорость движения воздуха (υ, м/с);
- Интенсивность теплового облучения (I, Вт/м2).

Самочувствие работника напрямую зависит от значений данных параметров. Высокая или низкая температура в комплексе с повышенной влажностью воздуха могут привести соответственно к перегреву или переохлаждению организма. При отсутствии движения воздуха в рабочем помещении перегрев наступит намного быстрее, аналогично, высокая скорость движения воздуха ускорит переохлаждение.

В таблице 5.3 представлены оптимальные и допустимые величины показателей микроклимата на рабочих местах производственных помещений. Величины соответствуют Iа категории интенсивности энергозатрат, так как вся работа производится сидя и сопровождающиеся незначительным физическим напряжением [40].

Таблица 5.3 - Оптимальные и допустимые величины показателей микроклимата на рабочих местах производственных помещений

Величины показателе	Период года	Температура воз-	Температура поверхностей, °	Относительная влажность	Скорость движения

й		духа, °С	С	воздуха, %	воздуха, м/с
Оптимальные	Холодный	22 – 24	21 – 25	40 – 60	0,1
	Теплый	23 – 25	22 – 26	40 – 60	0,1
Допустимые	Холодный	20 – 25	19 – 26	15 – 75	0,1
	Теплый	21 – 28	20 – 29	15 – 75	0,1 – 0,2

Для создания благоприятных условий труда и повышения производительности, необходимо поддерживать оптимальные параметры микроклимата производственных помещений. Для этого должны быть предусмотрены следующие средства: центральное отопление, вентиляция (искусственная и естественная), искусственное кондиционирование. Методы обеспечения нормальных микроклиматических условий:

1. Отопление – совокупность конструктивных элементов со связями между ними, предназначенных для получения, переноса и передачи необходимого количества теплоты в обогреваемых помещениях.

2. Защита от теплового излучения производится с помощью теплоизоляции, экранирования, оптимальных размещений оборудования и рабочих мест.

3. Герметизация помещений;

4. Кондиционирование – искусственная автоматическая обработка воздуха с целью поддержания в помещениях заранее заданных метеорологических условий, независимо от изменения наружных условий и режимов внутри помещения [41].

5.3.1.4 Повышенный уровень шума на рабочем месте

Согласно ГОСТ 12.1.003-2014 шум на рабочем месте оказывает раздражающее влияние на работника, повышает его утомляемость, а при выполнении задач, требующих внимания и сосредоточенности, способен привести к росту ошибок и увеличению продолжительности выполнения задания. Длительное воздействие шума влечет тугоухость работника вплоть до его

полной глухоты [42].

Вид трудовой деятельности и рабочее место в соответствии с СН 2.2.4/2.1.8.562-96 определяют предельно допустимые уровни звукового давления, уровни звука и эквивалентные уровни звука [42]. Разработка алгоритма соответствует виду трудовой деятельности, входящему в перечень «творческая деятельность, руководящая работа с повышенными требованиями, научная деятельность, конструирование и проектирование, программирование, преподавание и обучение, врачебная деятельность»; а рабочее место соответствует рабочему месту, входящему в перечень «рабочие места в помещениях дирекции, проектно-конструкторских бюро, расчетчиков, программистов вычислительных машин, в лабораториях для теоретических работ и обработки данных, приема больных в здравпунктах». Для данного вида трудовой деятельности и рабочего места в таблице 5.4 представлены предельно допустимые значения шума.

Таблица 5.4 - Предельно допустимые уровни звукового давления, уровни звука и эквивалентные уровни звука

Уровни звукового давления, дБ, в октавных полосах со среднегеометрическими частотами, Гц									Уровни звука и эквивалентные уровни звука (дБА)
31,5	63	125	250	500	1000	2000	4000	8000	50
86	71	61	54	49	45	42	40	38	

Наиболее типичным источником шума при данном виде деятельности будет являться компьютерная техника, в частности работающая система охлаждения ПЭВМ. Для снижения негативного влияния шума на рабочем месте могут быть применены следующие средства:

1. Использование охлаждающих систем с меньшим уровнем шума;
2. Использование звукопоглощающих материалов для системного блока компьютера для смягчения вибрации.

При отсутствии возможности избегания шума во время работы необходимо организовать места отдыха на некотором расстоянии от источников

шума.

5.3.1.5 Умственное перенапряжение

Основными обусловливающими развитие производственно-профессионального стресса (стрессового состояния) трудовыми нагрузками, являются [45]:

- при умственной нагрузке - длительный и ненормированный рабочий день с работой в сменном режиме, служебные командировки, работа в состоянии дефицита времени, длительность сосредоточенного внимания, плотность сигналов и сообщений в единицу времени, высокая степень сложности задания, выраженная ответственность, наличие риска для жизни;
- при зрительной нагрузке - высокая точность выполняемой работы, необходимость высокой координации сенсорных и моторных элементов зрительной системы, т. е. координации зрения с системой органов движения, время работы с оптическими приборами и время работы непосредственно с экраном видео дисплейных терминалов (ВДТ) и персональных электронно-вычислительных машин (ПЭВМ).
- при физической нагрузке - динамические и статические мышечные нагрузки, связанные с подъемом, перемещением и удержанием различного по массе груза, значительные усилия, прикладываемые к органам управления и ручным инструментам, многократно повторяющиеся движения рук различной амплитуды, выполнение глубоких наклонов корпуса, длительное поддержание физиологически нерациональных рабочих поз.

В соответствии с указанными в руководстве критериями определения степени сложности решаемых задач разработка алгоритма относится к сложным задачам. В течение рабочего дня работнику должен быть предоставлен перерыв для отдыха, в том числе для снижения умственного напряжения при выполнении

задач.

5.3.1.6 Повышенное значение напряжения в электрической цепи

В процессе работы необходимо управлять электрооборудованием, в результате чего создается вероятность прохождения электрического тока через тело человека. Несоблюдение правили ГОСТ 12.1.030-81 [43] и ГОСТ 12.1.038-82 ССБТ [44]. может привести к опасным последствиям. Опасное и вредное воздействия на людей электрического тока проявляются в виде электротравм (судороги, остановка сердца, остановка дыхания, ожоги и др.) и заболеваний.

Электрический ток оказывает на человека термическое, электролитическое, биологическое и механическое воздействие. Действие электрического тока на человека приводит к травмам или гибели людей. Для переменного тока частотой 50 Гц допустимое значение (не более) напряжения прикосновения составляет 2 В, а силы тока – 0,3 мА, для тока частотой 400 Гц, соответственно – 3 В и 0,4 мА, для постоянного тока – 8 В и 1 мА. Мерами защиты от воздействия электрического тока являются оградительные устройства, устройства автоматического контроля и сигнализации, изолирующие устройства и покрытия, устройства защитного заземления, устройства автоматического отключения, предохранительные устройства.

Токи статического электричества, наведенные в процессе работы компьютера на корпусах монитора, системного блока и клавиатуры, могут приводить к разрядам при прикосновении к этим элементам. Такие разряды опасности для человека не представляют, но могут привести к выходу из строя компьютера. Для снижения величин токов статического электричества используются нейтрализаторы, местное и общее увлажнение воздуха, использование покрытия полов с антистатической пропиткой.

5.4 Экологическая безопасность

5.4.1 Анализ влияния объекта исследования на окружающую среду

Так как основным объектом исследования данной работы являются разрабатываемый программный продукт, серьезной проблемой является воздействие на литосферу при утилизации компьютера и периферийных устройств (принтеры, МФУ, веб-камеры, наушники, колонки, телефоны).

Согласно ГОСТ 17.4.3.04-85 Охрана природы (ССОП) [47] отходах производства и потребления выделяют пять классов опасности отходов в зависимости от степени негативного воздействия на окружающую среду. Согласно Федеральному классификационному каталогу к IV классу малоопасных отходов относятся следующая техника:

- системный блок компьютера, утративший потребительские свойства;
- мониторы компьютерные плазменные/жидкокристаллические/ электроннолучевые, утратившие потребительские свойства;
- клавиатура, манипулятор "мышь" с соединительными проводами, утратившие потребительские свойства.

В то время как к I классу чрезвычайно опасных отходов относятся лампы ртутные, ртутно-кварцевые, люминесцентные, утратившие потребительские свойства. Каждая такая лампа, кроме стекла и алюминия, содержит около 60 мг ртути, поэтому представляет собой опасный источник токсичных веществ. Более точная оценка наличия вредных веществ и класса опасности производится либо специалистом на предприятии, либо сторонней специализированной организацией [47].

5.4.2 Обоснование мероприятий по защите окружающей среды

Для защиты окружающей среды от вредного воздействия выбросов необходимо осуществить полный переход к энергосберегающим и малоотходным (в идеале безотходным) производствам. Также на данный момент во многих странах используются альтернативные источники энергии (солнечные батареи, энергия ветра).

В аудитории не ведется никакого производства. К отходам,

производимым в помещении, можно отнести, в первую очередь, это бумажные отходы – макулатура, пластиковые отходы, неисправные детали персональных компьютеров и других видов ЭВМ. Бумажные отходы рекомендуется накапливать и передавать их в пункты приема макулатуры для дальнейшей переработки. Пластиковые бутылки складывать в специально предназначенные контейнеры. Неисправные комплектующие персональных компьютеров и других ЭВМ, а также люминесцентные лампы возможно сдавать организациям, имеющим специальную лицензию, которые занимаются переработкой отходов. Дополнительным методом снижения отходов является увеличение доли электронного документооборота.

5.5 Безопасность в чрезвычайных ситуациях

5.5.1 Анализ вероятных ЧС, которые может инициировать объект разработки

Среди чрезвычайных ситуаций, возможных для данного рабочего места можно выделить: пожар. Пожар является наиболее типичной чрезвычайной ситуацией ввиду наличия в помещении компьютерной техники, неполадки с которой могут привести его появлению. Он может возникнуть вследствие причин электрического и неэлектрического характеров. К причинам электрического характера можно отнести короткое замыкание, искрение, статическое электричество. К причинам неэлектрического характера относится неосторожное обращение с огнём, оставление без присмотра нагревательных приборов. Пожар может привести к травме, отравлению или гибели человека, а также к материальному ущербу.

5.5.2 Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС

Наиболее вероятной чрезвычайной ситуацией, которая может возникнуть в офисе во время разработки проекта, является пожар. Его могут вызвать

следующие причины:

- несоблюдение мер пожаробезопасности;
- обрыв проводов;
- замыкание электропроводки оборудования.

Исходя из возможных причин возникновения пожара для его предотвращения могут быть применены следующие меры [46]:

- регулярная проверка исправности работы электрических приборов, целостности соединительных проводов и вилок;
- соблюдение работниками техники безопасности при работе с электрическими приборами;
- заземления частей электрических приборов для снижения статического заряда. В случае возникновения пожара необходимо:
- незамедлительно сообщить об это в пожарную охрану;
- принять меры по эвакуации людей;
- отключить электроэнергию, приступить к тушению пожара первичными средствами пожаротушения.

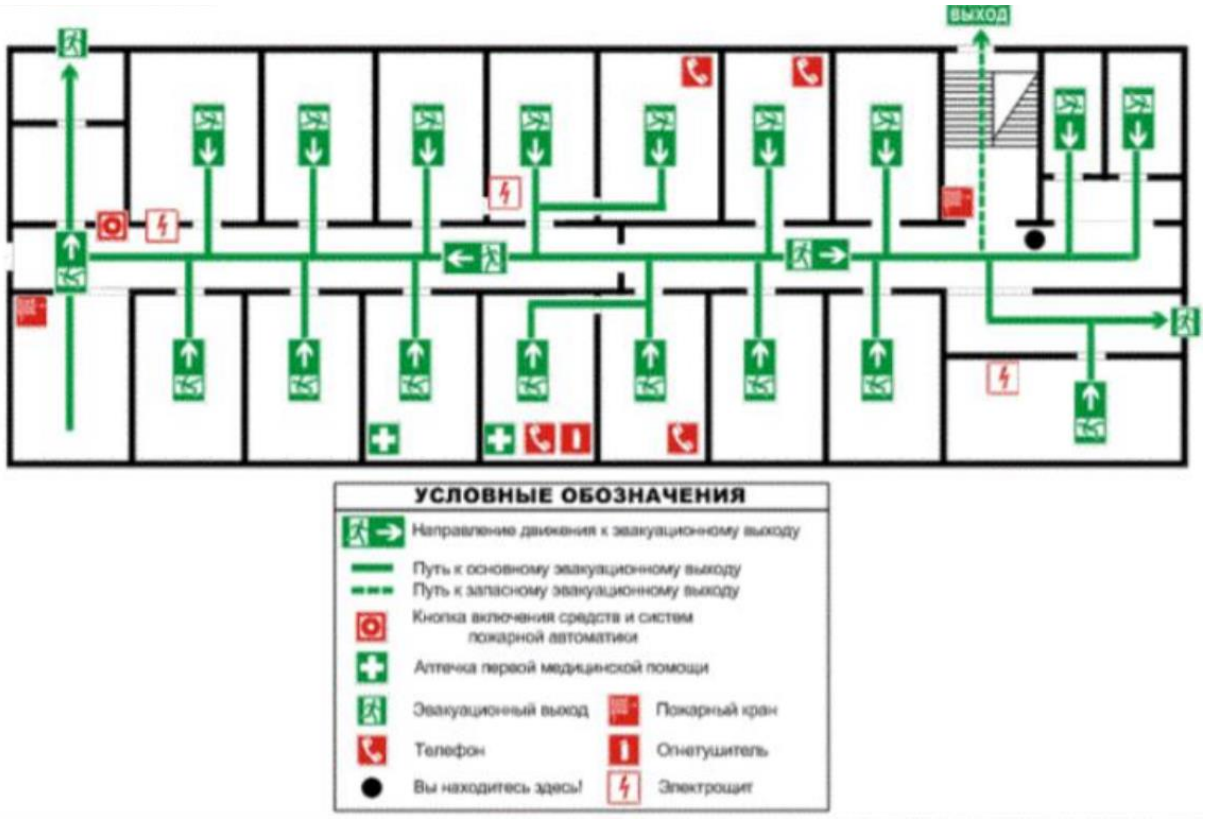

Рисунок 5.1 – План эвакуации

Пожарная безопасность в офисном помещении обеспечивается системой пожарной сигнализации. Имеется план эвакуации (рисунок 5.1), закреплен ответственный за пожарную безопасность.

5.6 Выводы по разделу

В ходе выполнения работы по разделу «Социальная ответственность» были рассмотрены правовые и организационные вопросы обеспечения безопасности, суть которых заключалось в анализе основных нормативных документов, регулирующих и регламентирующих производственную деятельность программиста-разработчика. Также были рассмотрены вредные факторы, такие как отклонение показателей микроклимата, производственный шум, отсутствие или недостаток естественного света, повышенный уровень электромагнитных излучений, риск поражения электрическим током. Были рассмотрены такие важные вопросы, как экологическая безопасность и безопасность в чрезвычайных ситуациях при использовании персонального компьютера. Следование правилам, описанным в данном разделе диссертации, помогут помочь избежать данных ситуаций, а также обеспечить здоровье персонала и сохранность окружающей среды.

ЗАКЛЮЧЕНИЕ

В этой работе анализируются и изучаются различные отличные статьи в смежных областях, а также рассматриваются и сравниваются существующие методы детектирования объектов и детектирования текста. EAST+CRNN был выбран в качестве сети для детектирования и распознавания текста.

Набор данных текстовых изображений реальных сцен ICDAR 2015 был выбран в качестве базы данных изображений для обучения и тестирования сети EAST. Искусственно синтезированный набор текстовых данных MJSynth был выбран в качестве базы данных изображений для обучения и проверки сети CRNN. И используйте набор данных ICDAR 2015 для тестирования общей архитектуры.

Разработан метод детектирования и распознавания надписей на изображениях реальных сцен, для чего выбираются параметры обучения нейронной сети, такие как размер входного изображения, замена базовой сети, выбор функции потерь.

Ветка детектирования текста метода достигла точности 79,8% в наборе данных ICDAR 2015, ветка распознавания достигла точности обучения 93,3% и точности проверки 91,8% в наборе данных MJSynth, в то время как в ICDAR 2015 г. Точность теста в наборе данных составляет 72,9%.

Этот метод в основном удовлетворяет требованиям идентификации и детектирования надписей на изображениях реальных сцен. А если добавить набор данных, похожий на обрезанные текстовые изображения, считается, что точность распознавания текста еще больше повысится.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. Ssd: Single shot multibox detector //European conference on computer vision. – Springer, Cham, 2016. – С. 21-37.
2. Ren, S., He, K., Girshick, R., & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks //Advances in neural information processing systems. – 2015. – Т. 28.
3. Matas, J., Chum, O., Urban, M., & Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions //Image and vision computing. – 2004. – Т. 22. – №. 10. – С. 761-767.
4. Dalal N., Triggs B. Histograms of oriented gradients for human detection //2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). – Ieee, 2005. – Т. 1. – С. 886-893.
5. Papageorgiou C. P., Oren M., Poggio T. A general framework for object detection //Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271). – IEEE, 1998. – С. 555-562.
6. Neumann L., Matas J. Real-time scene text localization and recognition //2012 IEEE Conference on Computer Vision and Pattern Recognition. – IEEE, 2012. – С. 3538-3545.
7. He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask r-cnn //Proceedings of the IEEE international conference on computer vision. – 2017. – С. 2961-2969.
8. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – С. 3431-3440.
9. Deng, D., Liu, H., Li, X., & Cai, D. Pixellink: Detecting scene text via instance segmentation //Proceedings of the AAAI Conference on Artificial Intelligence. – 2018. – Т. 32. – №. 1.

10. Lyu, P., Liao, M., Yao, C., Wu, W., & Bai, X. Mask textspotter: An end-to-end trainable neural network for spotting text with arbitrary shapes //Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – С. 67-83.
11. Zhang, Z., Zhang, C., Shen, W., Yao, C., Liu, W., & Bai, X. Multi-oriented text detection with fully convolutional networks //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 4159-4167.
12. Shi B., Bai X., Belongie S. Detecting oriented text in natural images by linking segments //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 2550-2558.
13. Dai, J., Li, Y., He, K., & Sun, J. R-fcn: Object detection via region-based fully convolutional networks //Advances in neural information processing systems. – 2016. – Т. 29.
14. Jiang, Y., Zhu, X., Wang, X., Yang, S., Li, W., Wang, H. R2CNN: Rotational region CNN for orientation robust scene text detection //arXiv preprint arXiv:1706.09579. – 2017.
15. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.
16. Neubeck A., Van Gool L. Efficient non-maximum suppression //18th International Conference on Pattern Recognition (ICPR'06). – IEEE, 2006. – Т. 3. – С. 850-855.
17. Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. East: an efficient and accurate scene text detector //Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. – 2017. – С. 5551-5560.
18. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation //International Conference on Medical image computing and computer-assisted intervention. – Springer, Cham, 2015. – С. 234-241.
19. Shi B., Bai X., Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition //IEEE transactions on pattern analysis and machine intelligence. – 2016. – Т. 39. – №. 11. – С. 2298-2304.

20. Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks //Proceedings of the 23rd international conference on Machine learning. – 2006. – С. 369-376.

21. Hochreiter S., Schmidhuber J. Long short-term memory //Neural computation. – 1997. – Т. 9. – №. 8. – С. 1735-1780.

22. Kim K. H. et al. Pvanet: Deep but lightweight neural networks for real-time object detection //arXiv preprint arXiv:1608.08021. – 2016.

23. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 770-778.

24. KAIST Scene Text Database. [Электронный ресурс]. – URL: http://www.iapr-tc11.org/mediawiki/index.php/KAIST_Scene_Text_Database (дата обращения 01.02.2022).

25. Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., ... & Valveny, E. ICDAR 2015 competition on robust reading //2015 13th international conference on document analysis and recognition (ICDAR). – IEEE, 2015. – С. 1156-1160.

26. Jaderberg, M., Simonyan, K., Vedaldi, A., & Zisserman, A. Synthetic data and artificial neural networks for natural scene text recognition //arXiv preprint arXiv:1406.2227. – 2014.

27. Yu J., Jiang Y., Wang Z., Cao Z., & Huang T. Unitbox: An advanced object detection network //Proceedings of the 24th ACM international conference on Multimedia. – 2016. – С. 516-520.

28. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network //Physica D: Nonlinear Phenomena. – 2020. – Т. 404. – С. 132306.

29. TextRecognitionDataGenerator’s documentation [Электронный источник]. URL: https://textrecognitiondatagenerator.readthedocs.io/en/latest/index.html. Дата обращения: 05.05.2022.

30. Liao M., Shi B., Bai X. Textboxes++: A single-shot oriented scene text detector //IEEE transactions on image processing. – 2018. – Т. 27. – №. 8. – С. 3676-3690.
31. Liu, X., Liang, D., Yan, S., Chen, D., Qiao, Y., & Yan, J. Fots: Fast oriented text spotting with a unified network //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 5676-5685.
32. Tian, Z., Huang, W., He, T., He, P., & Qiao, Y. Detecting text in natural image with connectionist text proposal network //European conference on computer vision. – Springer, Cham, 2016. – С. 56-72.
33. Трудовой кодекс: Федеральный закон от 30 декабря 2001г. N 197-ФЗ.
34. ГОСТ 12.0.003-2015 Опасные и вредные производственные факторы. Классификация. Перечень опасных и вредных факторов.
35. ГОСТ 22269-76 «Рабочее место оператора. Взаимное расположение элементов рабочего места».
36. ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
37. ГОСТ Р 50923-96. Дисплеи. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения.
38. ГОСТ 21889-76 «Система "Человек-машина". Кресло человека-оператора. Общие эргономические требования».
39. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95.
40. СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания.
41. ГОСТ 12.1.005-88 Система стандартов безопасности труда (ССБТ). Общие санитарно-гигиенические требования к воздуху рабочей зоны.
42. ГОСТ 12.1.003-83 Система стандартов безопасности труда (ССБТ). Шум. Общие требования безопасности.

43. ГОСТ 12.1.030-81 Система стандартов безопасности труда (ССБТ). Электробезопасность. Защитное заземление. Зануление.
44. ГОСТ 12.1.038-82 Система стандартов безопасности труда (ССБТ). Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов.
45. МР 2.2.9.2311 – 07 «Профилактика стрессового состояния работников при различных видах профессиональной деятельности».
46. ГОСТ 12.1.004-91 Система стандартов безопасности труда (ССБТ). Пожарная безопасность. Общие требования.
47. ГОСТ 17.4.3.04-85 Охрана природы (ССОП). Почвы. Общие требования к контролю и охране от загрязнения.

Приложение А

Раздел 1

Text detection and recognition methods

Обучающийся:

Группа	ФИО	Подпись	Дата
8ВМ03	Гао Жэньцзе		

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Иванова Юлия Александровна	к.т.н.		

Консультант – лингвист отделения иностранных языков, школы ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Ануфриева Татьяна Николаевна			

1 Text Detection and Recognition Methods

1.1 Text detection methods

Text detection can be considered as a special form of the task of detecting computer vision objects. The input to this problem is an image containing text, and the output is predictive information in the form of bounding boxes. General object detection tasks result in the position and area of objects such as animals, furniture, and cars in an image, while text detection mainly focuses on the precise positioning of text in an image. Compared with conventional object detection, text in a real scene has the characteristics of multi-directionality, irregular shape, extreme aspect ratio, different fonts, colors, and backgrounds. Thus, algorithms that are successful in general object detection often cannot be directly transferred to text detection.

Text detection has advanced significantly from detecting horizontal text at the beginning, to detecting multidirectional, curved and irregularly shaped texts. Text detection methods can be divided into two categories: traditional text detection methods and deep learning-based text detection methods. And the text detection methods based on deep learning are mainly divided into segmentation-based text detection and regression-based text detection.

1.1.1 Traditional text detection methods

Traditional text detection methods are mainly divided into the following steps:

- Image pre-processing: grayscale image, binarization, smoothing, tilt detection and correction, etc.;
- Image segmentation;
- Extraction of features from the image and their analysis;
- Training models for text recognition;
- Post-processing of recognition results.

These conventional detection methods have poor adaptability and anti-interference capability for various texts due to their limited image processing

capabilities. For example, the method of maximally stable extreme regions (MSER) [3], histogram of oriented gradient (HOG) [4], and Haar-like features [5].

The MSER method uses extreme areas for text area recognition. Extreme areas in the image are determined by two conditions regarding the set of pixels inside the area [3]:

The MSER method uses extreme areas for text area recognition. Extreme areas in the image are determined by two conditions regarding the set of pixels inside the area [3]:

- the area remains constant when transforming coordinates (rotation, stretching);
- the area remains constant as the brightness of the image changes. These methods require a small amount of memory for implementation, can work in real time, but poorly recognize text areas on blurry or low-contrast images, and often give false positive results [6].

The HOG method [4] divides the image into cells, then a histogram of oriented gradient is calculated for each pixel inside the cell. The union of the obtained values for each cell is called HOG [4] – descriptor. The resulting gradient image highlights the contours in the image and excludes irrelevant information from consideration [4]. Czarnek [5] in his work considers individual figures, and not a full text. Such data is easy to visualize and interpret, unlike images of real scenes, in which other objects are present in addition to text. Therefore, the directional gradient histogram method is more suitable for image preprocessing, but not for independent use.

Haar features can be understood as convolution patterns (e.g. Prewitt, Sobel operators are of course not quite the same). Haar features mainly consist of four categories in feature templates: edge features, line features, center features and diagonal features (Figure 1.1). There are two types of rectangles in the feature template, white and black, and the template feature value is defined as the sum of the pixels of the white rectangle minus the sum of the pixels of the black rectangle. The value of the Haar sign reflects the change in grayscale in the image.

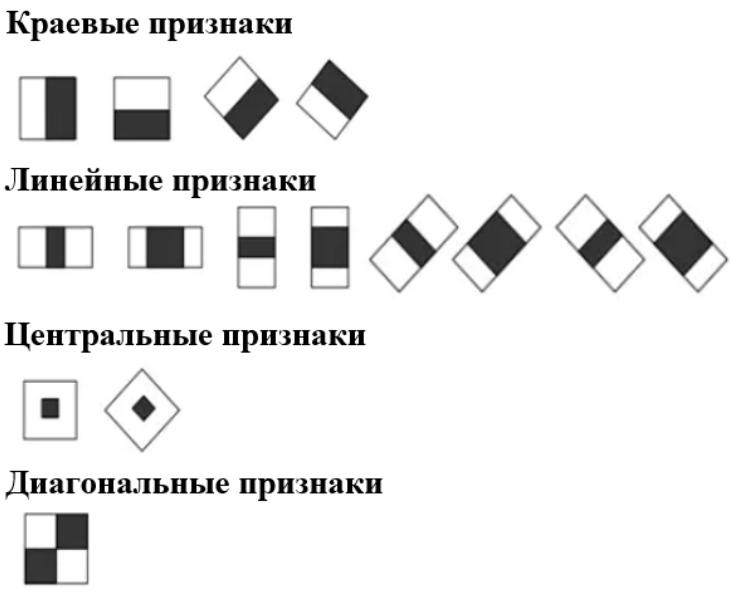

Figure 1.1 – Haar features

This method has major disadvantages for text detection as the font, position and text parameters vary from image to image.

1.1.2 Deep learning methods for text detection

The deep learning text detection method is a special learning-based text detection method. The traditional learning-based text detection methods mainly use two steps: "artificial feature extraction" and "classifier prediction", and the detection results are limited by the ability to represent artificial features. And deep neural networks have the ability to automatically learn image features.

Currently, a large number of text detection methods based on deep learning have appeared in real scenes, and they have achieved better results than traditional text detection methods. Generally speaking, real scene text detection methods based on deep learning mainly use two deep learning image processing strategies: regression-based text detection and segmentation-based text detection.

1.1.2.1 Text detection based on regression

This class of methods uses the network for detecting common objects as a basic model, and based on it, the algorithm is improved in combination with the practical application of text detection. For example, Faster R-CNN (Faster region-based

convolutional network) method [2], SSD (Single shot multi-box detector) method [1], R-FCN (Region-based fully convolutional networks) method [13].

Faster R-CNN [2] consists of four submodules: a convolutional layer, a Region proposal network (RPN), a RoI (regions of interest) pooling layer (RoI Pooling layer) and a classification and regression layer. Convolutional layers are used to extract image features, RPN is used to generate multiple candidate blocks associated with text objects, RoI pooling layer converts candidate blocks of different sizes to a fixed size, classification and regression layer make predictions for the candidate area and get the exact location of the area -candidate in the image. Since the original Faster R-CNN can only detect horizontal text, Jiang et al. therefore proposed the R2CNN (Rotational region CNN) algorithm [14] to detect text rotated at any angle in real scenes. Based on the original Faster R-CNN, this algorithm uses an RPN network to generate axisymmetric candidate blocks in different directions for a text area. After that, the features of the candidate areas in each direction are combined with different pool sizes to combine objects，and determine an axisymmetric candidate block and a slanted candidate block. The algorithm achieves a prediction result with an F-measure value of 82.54.

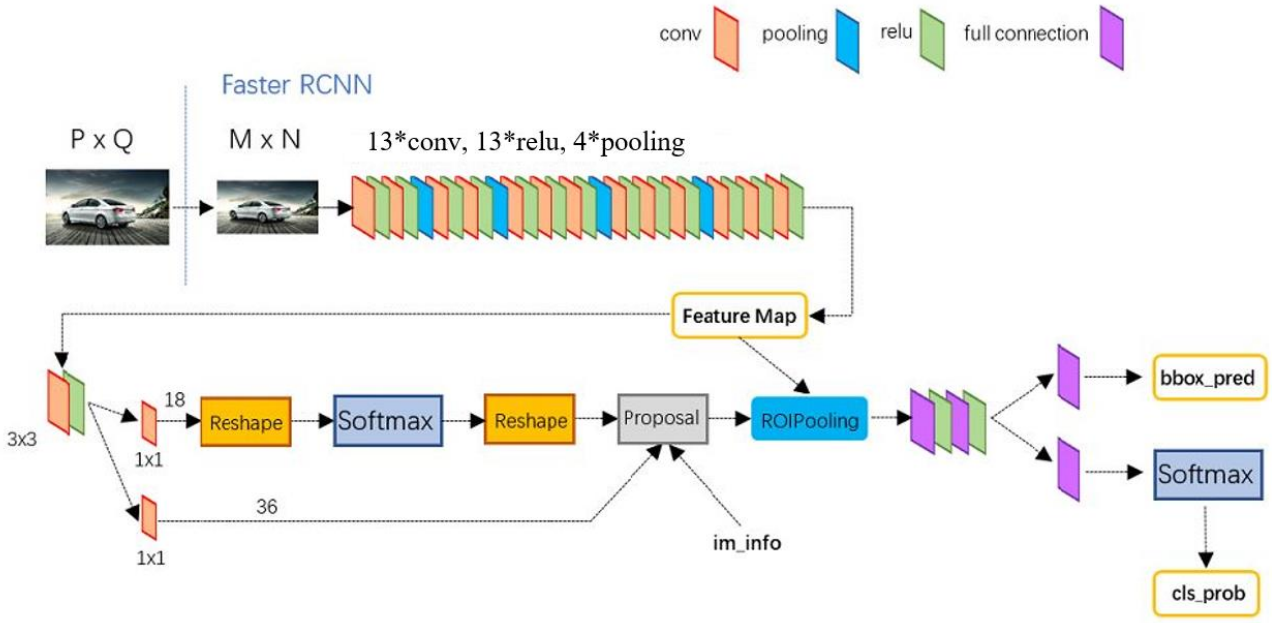

Figure 1.2 - An example of the Faster R-CNN structure [2]

SSD [1] injects an image into a modified VGG16 [15] to obtain feature maps

of different sizes, and then extracts feature maps of six convolutional layers Conv4_3, Conv7, Conv8_2, Conv9_2, Conv10_2, Conv11_2, and build reference fields of different scales at each point of these maps features, respectively. Finally, Non-Maximum Suppression (HMS) [16] is used to process all reference fields, suppress non-optimal reference fields, and display the final detection result.

The SegLink algorithm [12] proposed by Shi et al. injects an image into the SSD network and extracts the connection information between different text fields. The output of the model is not for the entire text string or word, but for a part of the text string or word, called a "segment", a segment can be 1 or more characters or 1 word. When extracting connection information from a text field, the algorithm takes a combination of different segments as the end result, avoiding the post-processing process of connecting segments to construct text strings. It is worth noting that the text field displayed by SegLink contains information about the angles, and only one field is displayed for each point in the feature map, which greatly reduces the computational complexity.

1.1.2.2 Text detection based on segmentation

This type of method uses semantic segmentation as the main technical means, processes images of real scenes through a deep learning semantic segmentation network, and obtains pixel-level label prediction, this pixel-level output is the basis for constructing text strings. Segmentation networks that are often used for text detection include Mask R-CNN [7], fully convolutional network (FCN) [8], PixelLink [9], EAST [17], etc., and the classical U- net [18] for medical image segmentation.

The R-CNN mask (Figure 1.2) is extended from Faster R-CNN [2] and Fast RCNN. In addition to the two detection branches (classification, bounding box regression) of the original network, a mask branch with a pixel-level prediction function for semantic segmentation is added at the end of the network. The mask branch takes the average binary cross-entropy loss and forms the network loss function along with the classification loss and the regression loss of the bounding box. Liu et al. at the

ECCV (European Conference on Computer Vision) proposed the Mask TextSpotter network [10] based on the Mask R-CNN. Its main innovation is to restructure the output of the Mask branch to include global text instance segmentation and character segmentation functions. Mask TextSpotter uses character-level segmentation and recognition so it can handle any irregularly shaped text (such as curved text), but its limitation is that it requires a character-level annotation to complete model training.

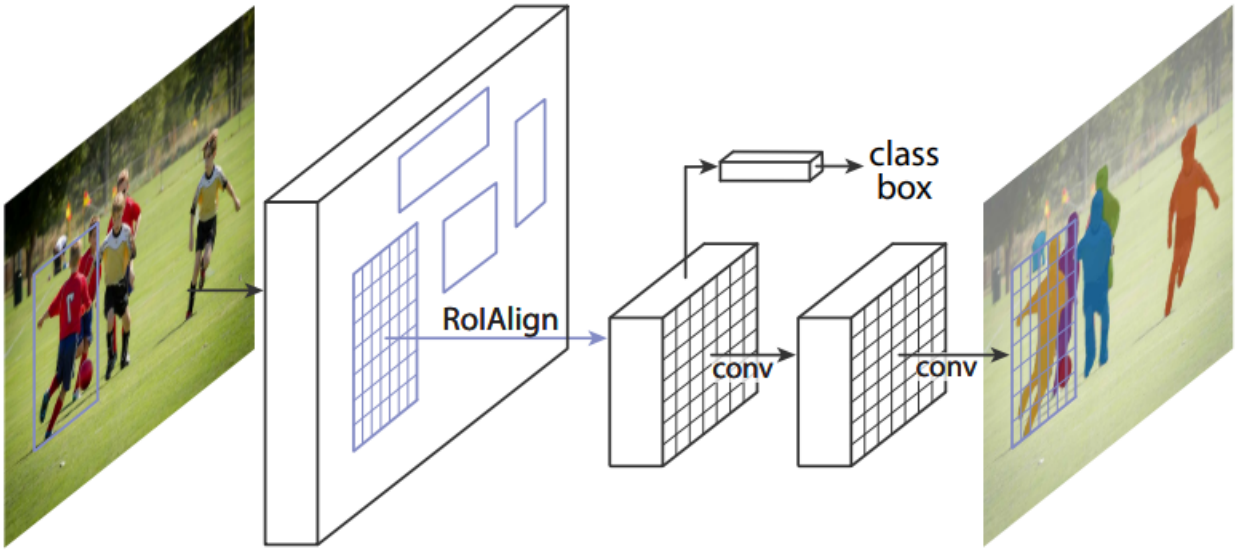

Figure 1.3 - An example of the structure of the Mask R-CNN [7]

The fully convolutional network FCN [8] is an end-to-end semantic segmentation method that differs from the area classification module and the bounding box regression module in algorithms such as Mask R-CNN. In FCN, network inference is a prediction of the entire image at the pixel level. Article [11] uses FCN to process an image, obtain a Salient map of a text area, and analyze its associated components to obtain a text block. On this basis, a region of candidate characters in a text block is extracted by the MSER method, and each text line is generated by concatenating the bounding boxes of the candidate characters. When using a fully convolutional network, false positives or missing characters may appear with very low contrast, curvature, strong light reflection, dense arrangement of lines of text or large spaces between characters. Another disadvantage of this type of network is speed.

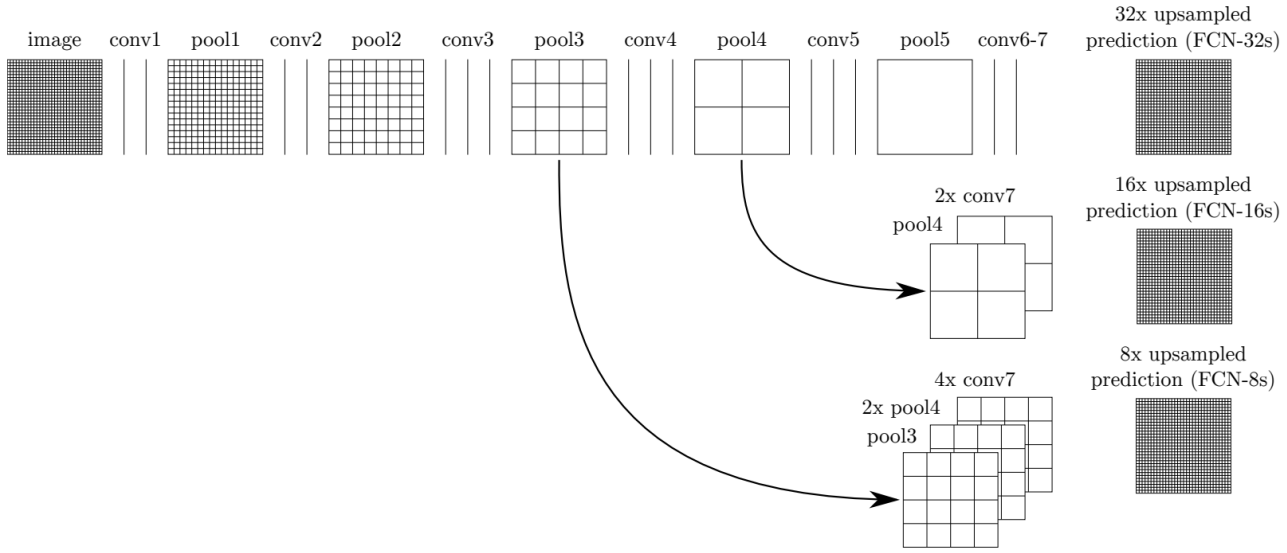

Figure 1.4 – An example of the FCN structure [8]

The PixelLink method [9] predicts pixels and link relationships associated with text through a deep learning network, applies an instance segmentation method to segment the region of a text string, and then directly finds the bounding box of the corresponding text string. The whole process consists of two parts: predicting and concatenating "positive pixels" according to the result of the "link positive" prediction, then obtaining a segmentation map of text instances, and directly extracting the bounding boxes of text strings from the segmentation map. Compared with image segmentation, text location in text detection is more accurate, and the segmentation method alone cannot accurately locate text at a small distance. Therefore, PixelLink adopts the idea of "link" in SegLink [12]. In prediction, it doesn’t only predict whether the pixels are text, but also predicts whether these pixels can be connected to form a good text field and output a more accurate text field.

U-net [18] is a widely used image segmentation model that is efficient, understandable, easy to build, and can be trained on small datasets. U-net is enhanced based on FCN. The original purpose of the network was to solve the problem of segmentation of medical images. Its network structure is shown in the figure 1.5.

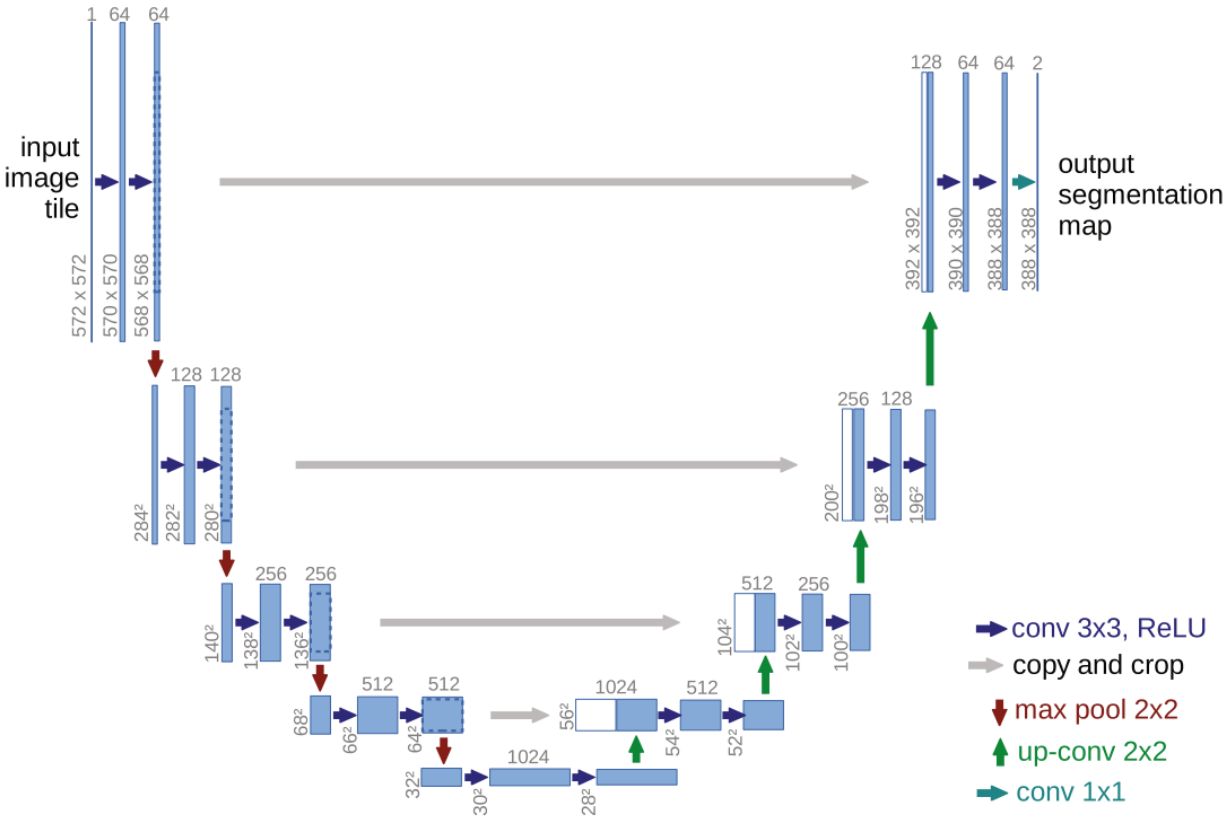

Figure 1.5 - An example of a U-net structure [8]

The U-net is relatively simple: the first half (left side of the figure) is used for feature extraction and the second half (right side of the figure) is used for upsampling. Its structure is also called the encoder-decoder structure. Since the general structure of this network is similar to a capital U, it is called U-net. U-net is very different from other common segmentation networks: U-net uses a completely different feature fusion method: splicing, which uses splicing features together in a channel dimension to form more meaningful features.

The semantics of medical images is relatively simple, the structure is fixed, and the semantic information is relatively simple, so there is no need to filter out useless information. All features of medical images are important, so low-level features and high-level semantic features are important, so U-shaped structures can be better used. There are many similarities between text label image segmentation and medical image segmentation.

EAST is a pixel-based scene text detection algorithm that can predict two different bounding boxes: one is RBOX (Rotated Box), which directly predicts the

distance from the center point to four sides (This is similar to the idea of FCOS) and the rotation angle of the text box; another is QUAD which predicts four vertices of an irregular quadrilateral. Thus, EAST can detect rotated or skewed text areas, but not curved text areas. In addition, the article [17] also proposes a new NMS: Locality Aware NMS (LANMS), which is a field-based string merge detection strategy that reduces NMS complexity from $O(n^2)$ to optimal $O(n)$, detection post-processing speed is greatly improved. The results obtained by EAST are useful for future research on text recognition. Its network structure is shown in Figure 1.6.

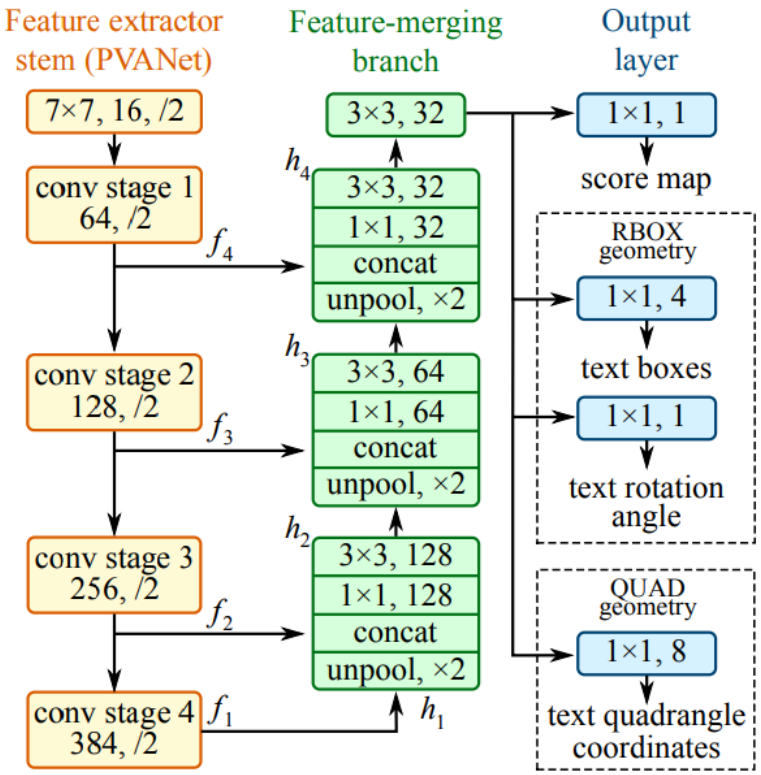

Figure 1.6 - An example of a U-net structure [17]

The EAST network structure can be divided into three parts:

Feature extractor: the figure uses PVANet [22], in fact, any model trained on large datasets can be used, for example, ResNet [23], VGG [15], etc.

Feature merging: here is a U-shaped structure, f1, f2, f3, f4 describe information about each scale, h2, h3, h4 merge information about each scale, and to ensure less calculation, the number of channels set to 128, 64, 32. Personally I believe that this is the result of the author's experiment.

Output layer: dashed border is optional here. A scorecard and RBOX geometry

form an output layer, or a scorecard and QUAD form an output layer. That is, the network can output quadrilaterals or rectangles with corners.

1.2 Text recognition method

Text recognition methods are also divided into traditional recognition methods and recognition methods based on deep learning. Among them, the conventional recognition method which is similar to a detection method that processes and recognizes an image through a sliding window. Since traditional methods can only recognize text in simple scenes, the research focuses on recognition methods based on deep learning.

CRNN+CTC [19–20] is a classic neural network architecture that combines feature extraction, sequence modeling, and transcription into a single structure [19]. The architecture used can naturally handle sequences of any length and does not require character segmentation or horizontal size normalization. It is not limited to any predefined lexicon and generates an efficient but much smaller model. Its network structure is shown in Figure 1.7.

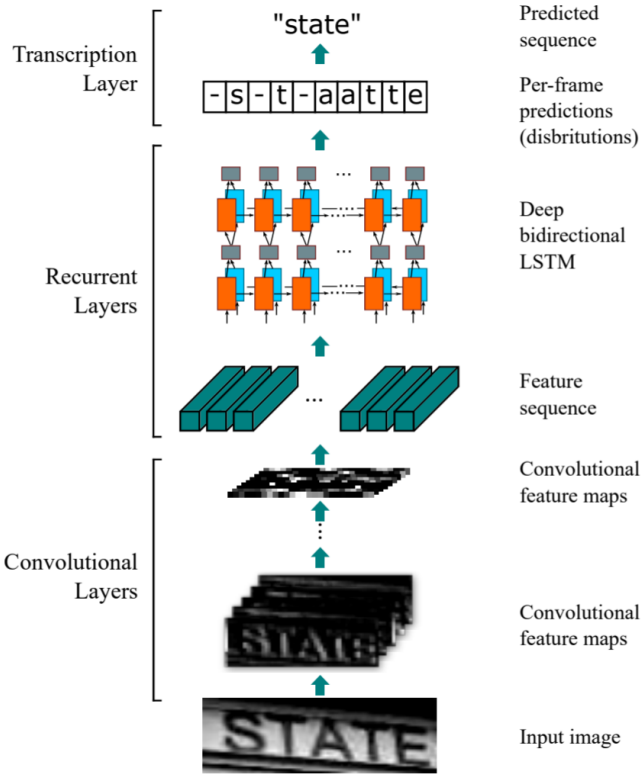

Figure 1.7 - CRNN network structure [19]

The CRNN + CTC network architecture mainly uses convolutional layers and pooling layers to obtain image features, and also uses LSTM [21] as RNN to predict text information, and the predicted text information is processed by CTC [20] to obtain the final prediction result.

1.3 Comparison of text detection methods

Due to the complexity of the inscriptions on images of real scenes, traditional detection methods cannot provide sufficiently good detection results, so the considered methods based on deep learning are compared according to Table 1. The following comparison methods were considered:

- size of occupied computing resources;
- speed results;
- method evaluation results;
- executability of the method program.

Table 1.1 - Comparison of methods

Methods	Comparison Methods			
	The size of the occupied computing resources	Speed results	Method Evaluation Results	Method program executability
Faster R-CNN	Medium	Low	High	Medium
SSD	Medium	Medium	Medium	Medium
SegLink	High	Medium	High	Medium
R-CNN mask	High	Medium	Medium	Medium
FCN	Medium	Low	High	High
PixelLink	High	Low	Medium	Medium
U-net	Medium	Medium	High	High
East	Medium	Medium	High	Medium

As can be seen from the table, detection methods based on deep learning are generally slower, but usually achieve good results. And according to the size of the computing resources occupied by the method and the possibility of executing the

program, the FCN, U-net and EAST methods were initially chosen and then compared.