

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ДЕТАЛИЗИРОВАННАЯ СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЯ ЧЕЛОВЕКА И ГЕНЕРАЦИЯ НОВОГО ИЗОБРАЖЕНИЯ

Автор: Пенская Таисия Андреевна _____

Направление подготовки: 01.03.02 Прикладная
математика и информатика

Квалификация: Бакалавр

Руководитель ВКР: Фильченков А.А., к.ф.-м.наук _____

Санкт-Петербург, 2021 г.

Обучающийся Пенская Таисия Андреевна
Группа М3437 Факультет ИТиП

Направленность (профиль), специализация
Математические модели и алгоритмы в разработке программного обеспечения

Консультанты:

а) Ефимова В.А., без степени, без звания

ВКР принята «_____» _____ 20__ г.

Оригинальность ВКР _____%

ВКР выполнена с оценкой _____

Дата защиты «_____» _____ 20__ г.

Секретарь ГЭК Павлова О.Н.

Листов хранения _____

Демонстрационных материалов/Чертежей хранения _____

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»

УТВЕРЖДАЮ

Руководитель ОП
проф., д.т.н. Парфенов В.Г. _____
« ____ » _____ 20__ г.

ЗАДАНИЕ
НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ

Обучающийся Пенская Таисия Андреевна

Группа М3437 **Факультет** ИТиП

Квалификация: Бакалавр

Направление подготовки: 01.03.02 Прикладная математика и информатика

Направленность (профиль) образовательной программы: Математические модели и алгоритмы в разработке программного обеспечения

Тема ВКР: Детализированная сегментация изображения человека и генерация нового изображения

Руководитель Фильченков А.А., к.ф.-м.наук, доцент ФИТиП, Университет ИТМО

2 Срок сдачи студентом законченной работы до: «31» мая 2021 г.

3 Техническое задание и исходные данные к работе

Требуется сгенерировать новое изображение человека на основе детализированной сегментации существующего изображения человека, соответствующее данному текстовому описанию. Изначально сервису подается на вход которому подается JSON определенного вида на английском языке, по которому сервис должен возвращать RGBA изображение человека, соответствующее описанию.

4 Содержание выпускной квалификационной работы (перечень подлежащих разработке вопросов)

- а) Найти сегментацию лица и одежды с помощью существующих моделей.
- б) Вырезать человека из остального изображения с помощью alpha mating.
- в) Реализовать генерацию произвольного человека с помощью StyleGAN2.
- г) Сгенерировать новую сегментацию по отличительным признакам.
- д) По сегментации сгенерировать изображение человека.
- е) Сравнить результаты с результатами существующих методов по метрикам FID, IS, SOA.

5 Перечень графического материала (с указанием обязательного материала)

Графические материалы и чертежи работой не предусмотрены

6 Исходные материалы и пособия

- а) Маттинг изображений: <https://arxiv.org/pdf/2011.11961v1.pdf>.
- б) Pix2Pix: <https://arxiv.org/pdf/1611.07004.pdf>.
- в) Pix2PixHD: <https://arxiv.org/pdf/1711.11585.pdf>
- г) CPGAN: <https://arxiv.org/pdf/1912.08562.pdf>.
- д) Метрика для оценки качества: <https://arxiv.org/pdf/1910.13321.pdf>.

7 Дата выдачи задания «15» февраля 2020 г.

Руководитель ВКР _____

Задание принял к исполнению _____ «15» февраля 2020 г.

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»

АННОТАЦИЯ
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Обучающийся: Пенская Таисия Андреевна

Наименование темы ВКР: Детализированная сегментация изображения человека и генерация нового изображения

Наименование организации, в которой выполнена ВКР: Университет ИТМО

ХАРАКТЕРИСТИКА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

1 Цель исследования: Разработка и реализация алгоритма генерации изображения человека на основе детализированной сегментации существующего изображения человека, соответствующее данному текстовому описанию.

2 Задачи, решаемые в ВКР:

- а) Разработка алгоритма для детализированной сегментации изображения человека.
- б) Разработка алгоритма генерации новой позы человека по текстовому описанию.
- в) Разработка алгоритма генерации изображения по текстовому описанию и сегментации, основанной на новой позе человека.
- г) Генерация набора данных для оценки качества работы алгоритма.

3 Число источников, использованных при составлении обзора: 0

4 Полное число источников, использованных в работе: 7

5 В том числе источников по годам:

Отечественных			Иностранных		
Последние 5 лет	От 5 до 10 лет	Более 10 лет	Последние 5 лет	От 5 до 10 лет	Более 10 лет
0	0	0	7	0	0

6 Использование информационных ресурсов Internet: да, число ресурсов: 7

7 Использование современных пакетов компьютерных программ и технологий:

Пакеты компьютерных программ и технологий	Раздел работы
latex	1.1
python3 скрипты	

8 Краткая характеристика полученных результатов

Разработан и реализован алгоритм для детализированной сегментации изображения человека. Сгенерирован датасет высокого качества с единственным человеком на изображении.

9 Гранты, полученные при выполнении работы

Нет

10 Наличие публикаций и выступлений на конференциях по теме выпускной работы
Нет

Обучающийся Пенская Т.А. _____

Руководитель ВКР Фильченков А.А. _____

« ____ » _____ 20 ____ г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1. Обзор современных результатов и постановка задачи.....	6
1.1. Алгоритмы для сегментации изображений	6
1.1.1. Алгоритмы для сегментации человека полностью	6
1.1.2. Алгоритмы для детализированной сегментации человека .	7
1.2. Алгоритмы для трансляции изображения.....	7
1.3. Постановка задачи	7
2. Описание разработанных алгоритмов.....	8
2.1. Подробное описание предполагаемого решения.....	8
2.2. Алгоритм для детализированной сегментации изображения человека.....	8
2.3. Алгоритм генерации новой сегментации по текстовому описанию	8
3. Полученные результаты алгоритмов	9
ЗАКЛЮЧЕНИЕ	10
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	11

ВВЕДЕНИЕ

Компьютерное зрение — это научное направление в области искусственного интеллекта и связанные с ним технологии получения изображений объектов реального мира, их обработки и использования полученных данных для решения разного рода прикладных задач без участия (полного или частичного) человека.

Одной из популярных задач компьютерного зрения является задача автоматического создания изображений на основе текста. Качественное решение этой задачи может быть использовано в любой литературе, ведь читателю проще воспринимать текст, который сопровождается иллюстрацией.

В последнее время (2017-2020) появилось много видов порождающих состязательных сетей (англ. Generative Adversarial Networks, GANs), решающих ее. Однако несмотря на рост популярности этой задачи, текущие модели генерируют нереалистичные изображения низкого качества. Данную задачу можно разбить на несколько подзадач, в числе которых будет генерация нового изображения человека по существующему. Это позволит создавать изображения несуществующих людей, не защищенные авторским правом.

Целью настоящей работы является разработка и реализация алгоритма генерации изображения человека на основе детализированной сегментации существующего изображения человека, соответствующее данному текстовому описанию.

Предлагаемый алгоритм принимает на вход фотографию и текстовое описание, в виде JSON, на английском языке. После этого, первым важным этапом нашей задачи является извлечение человека из фотографии. Чтобы качественно отделить человека от остального изображения, было принято использовать alpha matting. Следующим этапом обработки изображений является его детализированная сегментация, а именно сегментация лица и одежды человека. Дальнейшим шагом является генерация новой сегментации по отличительным признакам, указанным в описании. Конечным этапом является генерация нового изображения по полученной сегментации.

Данный подход позволит получать изображения людей в высоком разрешении, соответствующие данному текстовому описанию. Результаты можно будет использовать во многих сферах: игровая индустрия, киноиндустрия, научные исследования и многое другое.

ГЛАВА 1. ОБЗОР СОВРЕМЕННЫХ РЕЗУЛЬТАТОВ И ПОСТАНОВКА ЗАДАЧИ

1.1. Алгоритмы для сегментации изображений

Фундаментальной задачей компьютерного зрения является задача поиска групп пикселей, каждая из которых характеризует один смысловой объект, на изображениях и видео. Различают два типа подхода для данной задачи: сегментация и матирование.

Сегментация изображения — это задача кластеризации частей изображения на группы, посредством предсказания на уровне пикселей. То есть, сегментация изображения генерирует двоичное изображение, в котором пиксель либо принадлежит одной группе, либо другой.

Матирование изображения отличается от сегментации изображения тем, что некоторые пиксели могут принадлежать как одной группе, так и другой, такие пиксели называются частичными или смешанными пикселями. Чтобы полностью отделить передний план от фона в изображении, необходима точная оценка альфа-значений для частичных или смешанных пикселей.

В данной работе эта задача будет использована для выделения человека на изображении и детализированного выделения его одежды и лица.

1.1.1. Алгоритмы для сегментации человека полностью

Важной частью данной работы является выделение человека на изображении, чтобы впоследствии вырезать его и упростить задачу для генеративно-состязательных сетей, подавая на вход изображение без фона. Для сегментации и маттинга человека существует множество решений, которые были проанализированы, подробнее о них в следующей таблице 1.

Таблица 1 – Таблица существующих решений для выделения человека на изображении.

Статья	Комментарии
[5]	Используется tensorflow.
[4]	Делается сегментация, что грубее, чем маттинг, а значит, конечные результаты будут хуже. Отдельно для своих данных нужно выделять keypoints.
[3]	Нужна дополнительно сгенерированная тримапа.
[7]	Недавно полученные хорошие результаты, предобученная модель есть.

1.1.2. Алгоритмы для детализированной сегментации человека

Для сегментации одежды на человеке был найден хороший датасет DeepFashion2[6], который и был использован для обучения модели впоследствии.

Для сегментации лица человека была найдена предобученная модель[1] с хорошими результатами, которая была использована дальше в реализации алгоритма.

1.2. Алгоритмы для трансляции изображения

Важной задачей компьютерного зрения является также задача трансляции изображения, цель которой состоит в том, чтобы научиться строить соответствия между входным и выходным изображениями, используя тренировочные данные.

Для решения этой задачи используются такие архитектуры, как генеративно-сопоставительные сети. Генеративно-сопоставительные сети — алгоритм машинного обучения, входящий в семейство порождающих моделей и построенный на комбинации из двух нейронных сетей: генеративная модель G , которая строит приближение распределения данных, и дискриминативная модель D , оценивающая вероятность, что образец пришел из тренировочных данных, а не сгенерированных моделью G . Обучение для модели G заключается в максимизации вероятности ошибки дискриминатора D .

В данной работе этот подход используется для генерации нового изображения человека по сегментации и отличительным признакам, указанным в тексте.

1.3. Постановка задачи

В качестве достижения поставленной цели диплома были поставлены следующие подзадачи:

- а) Разработка алгоритма для детализированной сегментации изображения человека.
- б) Разработка алгоритма генерации новой позы человека.
- в) Разработка алгоритма генерации изображения по текстовому описанию и сегментации, основанной на новой позе человека.
- г) Генерация набора данных для оценки качества работы алгоритма

ГЛАВА 2. ОПИСАНИЕ РАЗРАБОТАННЫХ АЛГОРИТМОВ

2.1. Подробное описание предполагаемого решения

Для решения поставленной задачи был придуман следующий алгоритм:

- а) Вырезать человека из изображения с помощью alpha mating.
- б) Сегментировать лицо и одежду.
- в) Реализовать генерацию новой позы человека по тексту (человек сидит/-стоит).
- г) Сгенерировать новую сегментацию по позе человека.
- д) По сегментации сгенерировать изображение человека по отличительным признакам, релевантным тестовому описанию.

2.2. Алгоритм для детализированной сегментации изображения человека

Для осуществления детализированной сегментации нужно поэтапно сделать следующее:

- а) Вырезать человека полностью из изображения с помощью матирования
- б) Сегментировать лицо и одежду

Для сегментации человека была выбрана модель с лучшими показателями[7] из таблицы 1.

Для сегментации лица человека была использована предобученная модель[1].

Для сегментации одежды человека была обучена Mask-RCNN[2] с использованием датасета DeepFashion2[6].

2.3. Алгоритм генерации новой сегментации по текстовому описанию

Для генерации новой сегментации по текстовому описанию:

- а) Определить текущую позу человека
- б) По тексту определить новую позу человеку
- в) Сгенерировать сегментацию для новой позы человека

ГЛАВА 3. ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ АЛГОРИТМОВ

В результате была обучена Mask-RCNN, результаты ниже на графике:

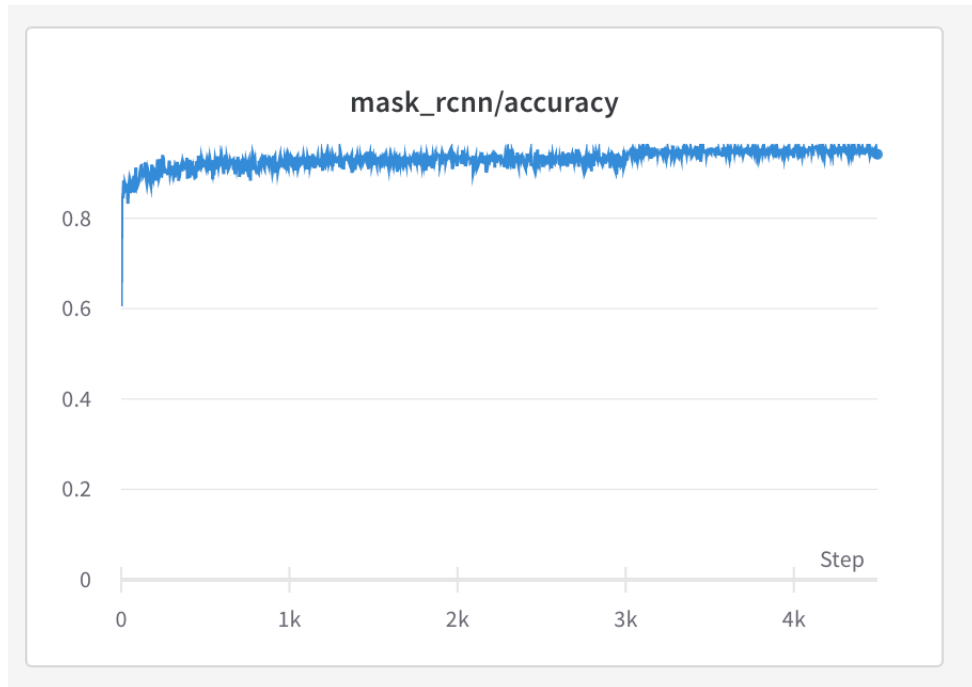


Рисунок 1 – Mask-RCNN результаты

Был предобработан датасет Fashionpedia, в результате чего был получен датасет с 31188 качественными изображениями с человеком, а 17695 изображений были удалены.

ЗАКЛЮЧЕНИЕ

В процессе.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Jian Li Y. W.* DSFD: Dual Shot Face Detector [Электронный ресурс]. — 2019. — URL: <https://arxiv.org/abs/1810.10220>.
- 2 *Kaiming He G. G.* Mask R-CNN [Электронный ресурс]. — 2018. — URL: <https://arxiv.org/abs/1703.06870>.
- 3 *Marco Forte F. P. F, B.* Alpha Matting [Электронный ресурс]. — 2020. — URL: <https://arxiv.org/abs/2003.07711>.
- 4 *Song-Hai Zhang R. L.* Pose2Seg: Detection Free Human Instance Segmentation [Электронный ресурс]. — 2019. — URL: <https://arxiv.org/abs/1803.10683>.
- 5 *Yunke Zhang L. G.* A Late Fusion CNN for Digital Matting [Электронный ресурс]. — 2019. — URL: https://openaccess.thecvf.com/content_CVPR_2019/papers/Zhang_A_Late_Fusion_CNN_for_Digital_Matting_CVPR_2019_paper.pdf.
- 6 *Yuying Ge R. Z.* DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images [Электронный ресурс]. — 2019. — URL: <https://arxiv.org/abs/1901.07973>.
- 7 *Zhanghan Ke K. L.* Is a Green Screen Really Necessary for Real-Time Human Matting? [Электронный ресурс]. — 2020. — URL: <https://arxiv.org/abs/2011.11961v1>.