

Санкт-Петербургский Государственный Университет

Кафедра системного программирования

Сегментация изображений человека в видеопотоке со статичным фоном

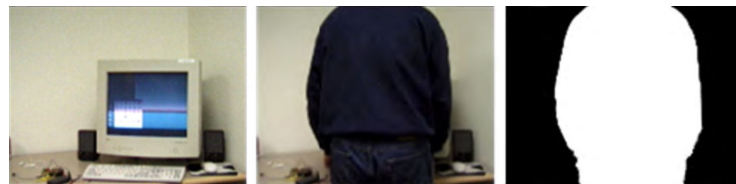
Автор: Немчинов Егор Игоревич, 16.Б11-мм

Научный руководитель: к. ф.-м. н., ст. преп. Д. В. Луцив

Санкт-Петербург, 2020

Сегментация видео со статичной камерой

- Задача отделения переднего плана от известного фона
- Применения
 - Смена фона
 - Системы видеонаблюдения
 - Техники “захвата движения”
 - Данные для нейронных сетей



K. Toyama et al., “Wallflower: Principles and practice of background maintenance,” (IEEE 1999)



Vacavant, Antoine, et al. “A benchmark dataset for outdoor foreground/background extraction” Springer, Berlin, Heidelberg, 2012.

Подходы для сегментации

- DeepLabv3+ — многоклассовая сегментация
 - Chen, L.C et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." (ECCV 2018).
- Pose2Seg — сегментация изображений людей по ключевым точкам
 - Работает стабильно хорошо в сложных сценариях, но не детализировано
 - Zhang, Song-Hai, et al. "Pose2seg: detection free human instance segmentation." (*IEEE*, 2019)



Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." *arXiv preprint arXiv:1812.08008* (2018).

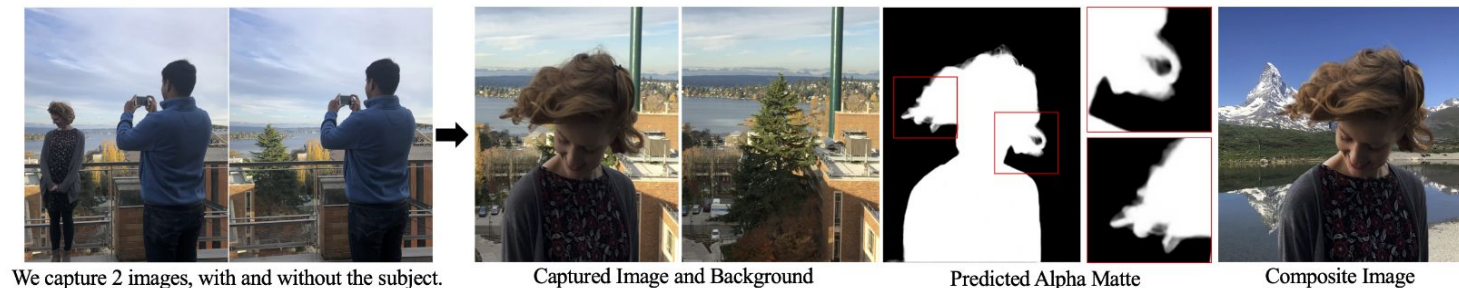
Подходы для удаления фона

- SemanticBGS

- Shahbaz et al. "Deep Foreground Segmentation using Convolutional Neural Network." (IEEE, 2019).

- Background-Matting — уточнение сегментации при известном фоне

- На вход требует маску (например, от DeepLabv3+) и фон



Sengupta, Soumyadip, et al. "Background Matting: The World is Your Green Screen." (arXiv preprint 2020).

Проблемы

- Сегментация изображений человека не использует информацию об известном фоне
- Методы удаления фона по видео с людьми не используют специфическую для класса человека информацию
- Методы сегментации видео с человеком труднодоступны

Цель и задачи

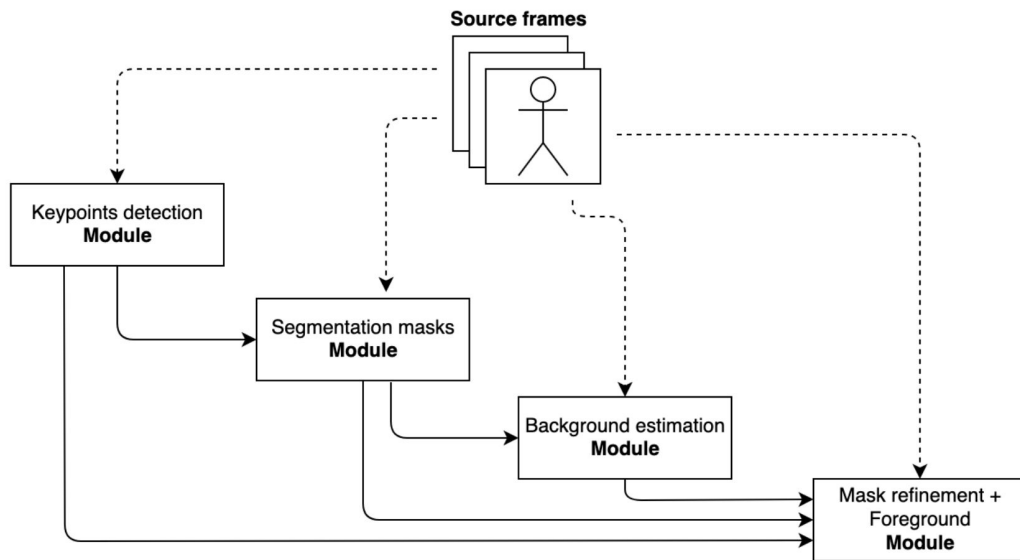
Цель: создать систему сегментации кадров видеопотока с человеком на статичном фоне, доступную широкой аудитории

Задачи:

- Сделать обзор предметной области
- Разработать метод сегментации кадров видеопотока с человеком на статичном фоне
- Провести апробацию метода и проанализировать результаты
- Разработать веб-сервис с интерфейсом для алгоритма сегментации видео

Общая схема

- Собран единый процесс сегментации видео на статичном фоне из различных компонентов на языке Python

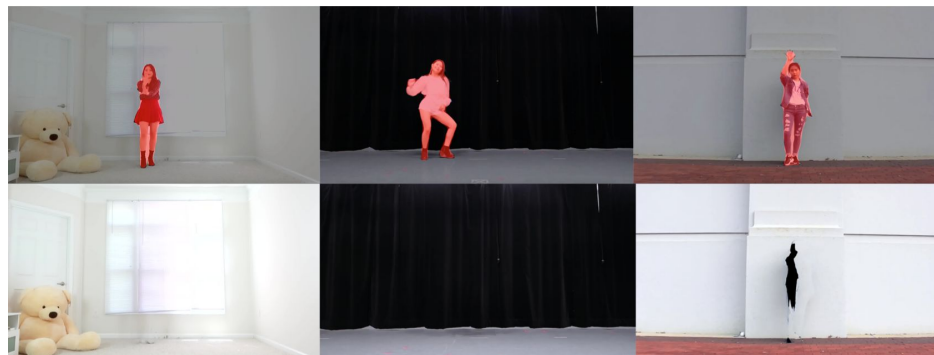


Реализация

- В качестве метода сегментации выбран **Pose2Seg** с ключевыми точками из метода **AlphaPose**
- Реализован метод извлечения фона на основе **Pose2Seg**

Метод	F-мера
DeepLabv3+	0.95
Pose2Seg (OpenPose)	0.92
Pose2Seg (AlphaPose)	0.96

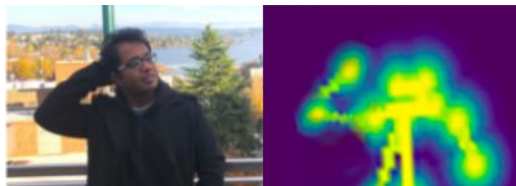
Сравнение подходов для сегментации на подмножестве LASIESTA



Верхняя строчка: примеры работы Pose2Seg, нижняя строчка: извлеченный с видео фон

Реализация

- Для уточнения маски взята модификация метода **BG Matting**
- На вход добавлены ключевые точки из метода **AlphaPose**
- Добавлена функция потерь, которая штрафует за сильное отклонение от изначальной маски
- Собрано 40 видео с Youtube с людьми в полный рост



Пример добавленных ключевых точек

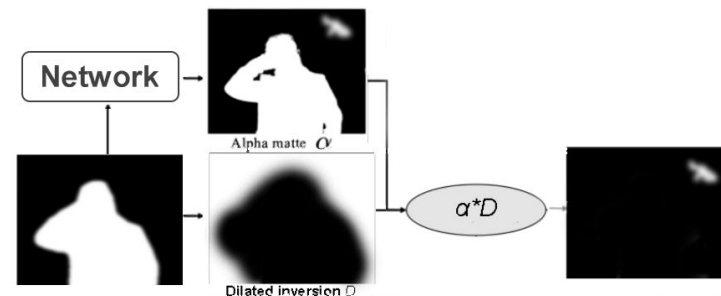


Схема вычисления добавленной функции потерь, *Divergence loss*

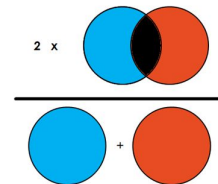
Датасет и метрики

● Метрика

- **F-мера** — гармоническое среднее точности (precision) и полноты (recall)

● Датасеты для апробации

- CDNet 2014. Наиболее популярный набор данных для сравнения методов удаления фона. Отобраны видео с людьми
 - Yi Wang et al., “CDnet 2014: An expanded change detection benchmark dataset” (IEEE, 2014).
- LASIESTA. Содержит видео и сегментацию для различных случаев: легкая тряска, изменение освещения, одежда цвета фона, разные погодные условия в уличной съемке
 - C. Cuevas et al., “Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA” (Comput. Vis. Image Understand., 2016).



<https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2>

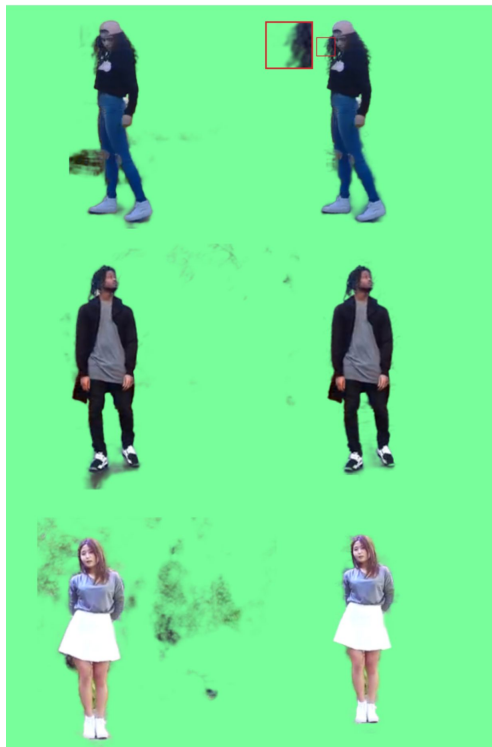


<http://www.gti.ssr.upm.es/data/LASIESTA>

Апробация

Модель	Средняя точность	Средняя полнота	Средняя F-мера
SemanticBGS	0.93	0.97	0.95
Pose2Seg	0.96	0.94	0.95
Background Matting	0.93	0.96	0.95
<i>Эта работа, kpts</i>	0.95	0.98	0.96
<i>Эта работа, div. loss</i>	0.96	0.97	0.97
<i>Эта работа, kpts + div. loss</i>	0.98	0.97	0.98

Визуальное сравнение



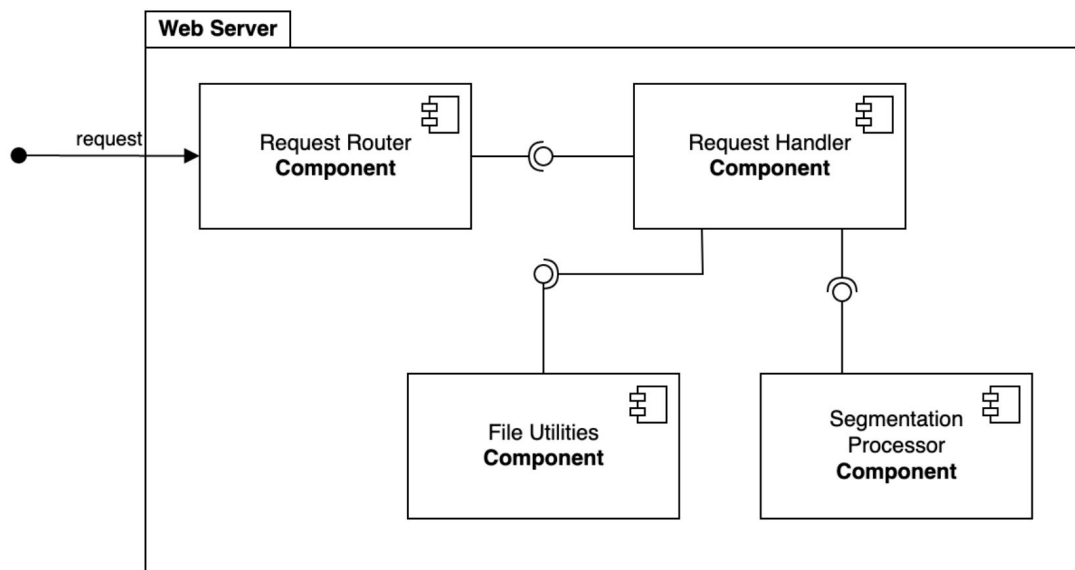
Сравнение вариаций метода с использованием ключевых точек. Слева отсутствует добавленная функция потерь, справа присутствует



Сравнение трех вариаций метода на подмножестве LASIESTA. Слева направо: исходное изображение; только ключевые точки; только новая функция потерь; и ключевые точки, и функция потерь

Веб-сервис

- Интерфейс для загрузки видео и выгрузки результатов
- Python + JavaScript для асинхронной загрузки файлов



Результаты

- Проведен обзор предметной области
 - Рассмотрены методы сегментации изображений человека
 - Рассмотрены методы вычисления и вычитания фона
- Разработан метод для сегментации видео человека со статичным фоном
 - Опробованы различные архитектуры нейросетей
 - Изменена и обучена сеть, дорабатывающая маски вычитанием фона
 - Реализована система для сегментации видео человека из нескольких частей
- Проведены апробация и анализ метода
 - Измерено влияние модификации архитектуры сети на качество сегментации
 - Произведено сравнение с другими аналогичными системами
- Реализован простой веб-сервис с реализацией метода