

Анализ статических и динамических характеристик движения человека для идентификации по походке с помощью методов машинного обучения

Мавджуда Хакимова

Московский физико-технический институт

Май 2025

Научный руководитель: Голубинский А.Н. (ИППИ РАН)
Команда: Панкратов Евгений, Корнилов Константин, Игорь Нетай

План

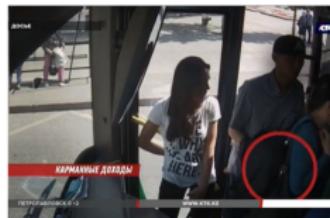
- ① Введение - формулировка задачи
- ② Анализ характеристик
- ③ Результаты на сегодня
- ④ Исследование новой модели для нашей задачи
- ⑤ Дальнейшие действия

Постановка задачи

Идентификация человека по походке представляет собой важный биометрический метод, который позволяет распознавать людей на расстоянии и без непосредственного контакта. В отличие от традиционных методов, таких как отпечатки пальцев или радужная оболочка глаза, распознавание по походке становится особенно актуальным в условиях современных систем видеонаблюдения.

Актуальность

Походка уникальна для каждого человека и практически не поддается фальсификации. Это делает её ценным идентификатором в сфере безопасности, где необходимо распознавать людей в реальном времени, например, для предотвращения преступлений или контроля доступа.



Цели и задачи проекта

Задача распознавания человека по походке:

Определить личность человека в видеозаписи на основе анализа его походки и сравнения с имеющейся галереей.

Научная постановка задачи

- **Разметка Данных:** Изучение инструментов разметки для идентификации по походке.
- **Предобработка Данных:** Сглаживание и фильтрация данных.
- **Анализ Характеристик:** Расчет и анализ статических и динамических характеристик движения.
- **Машинное Обучение:** Применение методов машинного обучения, включая нейросетевые модели.
- **Оценка Гиперпараметров:** Оценка и сравнительный анализ гиперпараметров нейросетевых архитектур.

Методы идентификации человека по походке в видео

Авторы: А. И. Соколова, А. С. Конушин

Обзор методов: В статье рассматриваются различные подходы к распознаванию походки, включая:

- **Классические методы:**

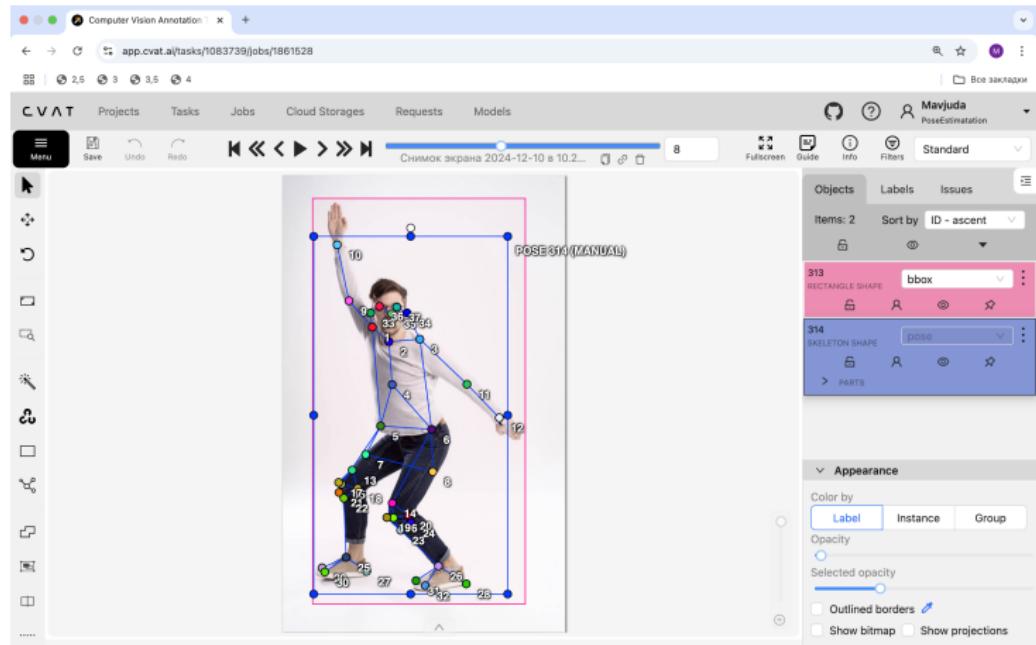
- Извлечение признаков вручную, таких как бинарные силуэты и изображения энергии походки (Gait Energy Image).
- Использование кинематических характеристик, таких как расстояния между суставами и их движения.

- **Нейросетевые методы:**

- Сверточные нейронные сети (CNN) для обработки изображений и извлечения абстрактных признаков.
- Комбинированные подходы, использующие как ручные признаки, так и обучаемые нейронные сети.

Ссылка на статью: <https://www.mathnet.ru/tisp399>

Computer Vision Annotation Tool (CVAT). Это open source инструмент, он используется для разметки изображений и видео, предназначенных для обучения модели компьютерного зрения.



Процесс обработки стереоскопического видео



Рис.: Пайплайн обработки данных

- Исходный стереоскопический видеофайл (.bag) преобразуется в двумерное видео (.mp4).
- Используется модель глубокого обучения **YOLOv11x-pose** для оценки позы человека.

Процесс обработки стереоскопического видео (продолжение)

- Модель возвращает координаты 17 ключевых суставных точек в пиксельной системе (x_pix, y_pix).
- Координаты сохраняются в текстовые файлы (.txt) для дальнейшей обработки.
- Стереофайл (.bag) используется повторно для извлечения трёхмерных координат (x, y, z).
- Координаты пикселей сопоставляются с пикселями стереоизображения с использованием глубинной информации от камеры **Intel RealSense D457**.
- Трёхмерные координаты сохраняются в массив формата (.pru).

Функции сглаживания данных

- **Decay фильтр (decay_filt):** Функция `decay_filt` предназначена для сглаживания данных с использованием экспоненциального затухания. Это может быть полезно в ситуациях, когда мы хотим, чтобы значения постепенно уменьшались в зависимости от предыдущих значений, что может помочь в устраниении резких изменений в данных.
- **Smoothen фильтр (smoothen):** Функция `smoothen` использует окно Ханна для сглаживания данных. Это метод, который помогает уменьшить шум в данных, применяя взвешенное усреднение значений.

Получение характеристик

Получаем длины ребер и углы между ними, получаем 12 длин рёбер, характеризующих расстояния между суставными точками, и 16 углов, описывающих ориентацию соединений в пространстве. Далее проводится анализ динамических свойств данных путём вычисления первых и вторых производных. Всего получаем 84 признака.

Получение графиков характеристик

Нарисованы временные и спектральные графики:

Для спектральных графиков использована формула преобразования амплитуды сигнала в децибелы:

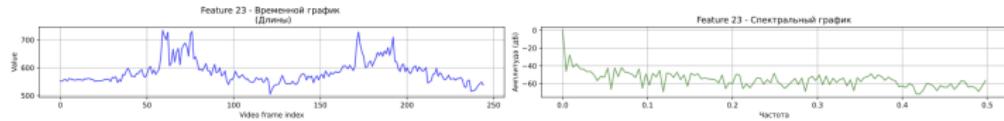
$$A[\partial B] = 20 \log_{10} \left(\frac{A_{\max}}{A(f)} \right)$$

где

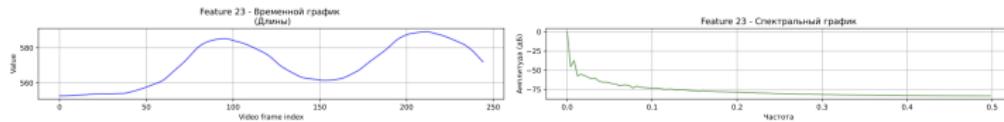
- $A(f)$ — амплитуда сигнала на частоте f ,
- A_{\max} — максимальная амплитуда сигнала.

Временные и спектральные графики

- Временной и спектральный графики для одного признака до фильтрации:



- Временной и спектральный графики после фильтрации:



Применение нейросетевых архитектур

В ходе исследования командой были применены следующие нейросетевые архитектуры:

- Многослойный перцепtron (MLP)
- Сети с долгосрочной краткосрочной памятью (LSTM)
- Свёрточные нейронные сети (CNN)

На текущий момент модель MLP демонстрирует способность распознавать 12 из 21 класса, при этом значение метрики F1 составляет 0.46.

Объяснение метрики F1

Метрика F1 сочетает в себе как точность, так и полноту, обеспечивая более полное представление о качестве модели.

- **Точность (Precision):** отражает долю правильно предсказанных положительных классов среди всех предсказанных положительных классов.
- **Полнота (Recall):** показывает долю правильно предсказанных положительных классов среди всех истинных положительных классов.

Формула для вычисления метрики F1:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Значение F1 варьируется от 0 до 1, где 1 указывает на идеальную точность и полноту.

Изучение трансформеров

Трансформеры состоят из следующих компонентов:

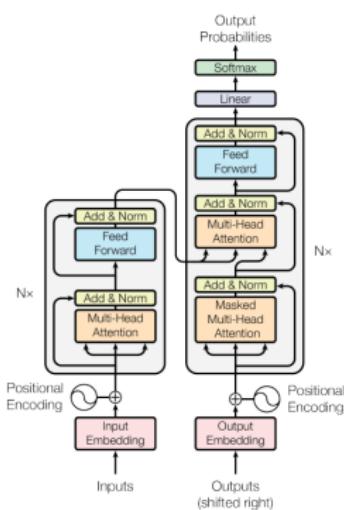


Figure 1: The Transformer - model architecture.

- Входные данные и эмбеддинги
- Позиционные эмбеддинги
- Механизм внимания
- Многоголовое внимание
- Нормализация и активация
- Полносвязные слои
- Слои кодировщика и декодировщика
- Выходной слой
- Обучение

Изучение трансформеров

Основная идея трансформеров была изучена по этим статьям:

- [Attention is All You Need](#)
- [Multi-head Attention](#)
- [Что такое трансформеры?](#)

Также был разработан простой трансформер для языковой идентификации, который позволяет определить язык текста по предложению.

Вы можете ознакомиться с кодом на моем GitHub: [GitHub репозиторий](#)

Проблемы с запуском кода трансформеров

- Попытка запустить код внутри контейнера командой:

```
docker compose exec -u root -it rsvid bash
```

- Это позволило войти в контейнер с правами root и выполнить необходимые команды.
- Однако такой подход оказался нежелательным, т.к. повышение привилегий может привести к ошибкам или нарушениям безопасности.
- В результате пришлось разобраться с Docker более подробно:
 - Настройка Dockerfile и docker-compose.yml для правильной изоляции и прав.
 - Использование правильных пользовательских настроек для запуска приложений.
 - Обеспечение безопасности и стабильности работы контейнера.

Дальнейшие действия

- Применение трансформеров к нашей задаче.
- Исследование точности.
- Изучение других нейросетевых архитектур.

Благодарность

Спасибо за внимание!