

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

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

Задача распознавания человека по походке:

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

- **Разметка Данных:** Изучение инструментов разметки для идентификации по походке.
- **Предобработка Данных:** Сглаживание и фильтрация данных.
- **Анализ Характеристики:** 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.

Исправлены временные и спектральные графики:

Для спектральных графиков использована формула преобразования амплитуды сигнала в децибелы:

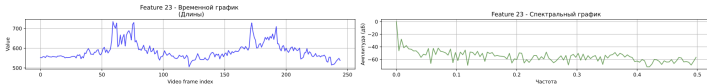
$$A[\text{дБ}] = 20 \log_{10} \left(\frac{A_{\max}}{A(f)} \right)$$

где

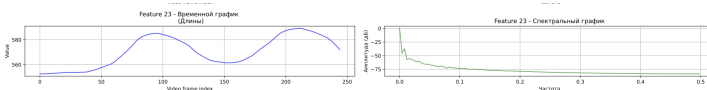
- $A(f)$ — амплитуда сигнала на частоте f ,
- A_{\max} — максимальная амплитуда сигнала.

Временные и спектральные графики

- Временной и спектральный графики для одного признака до фильтрации:



- Временной и спектральный графики после фильтрации:



Процесс обработки стереоскопического видео

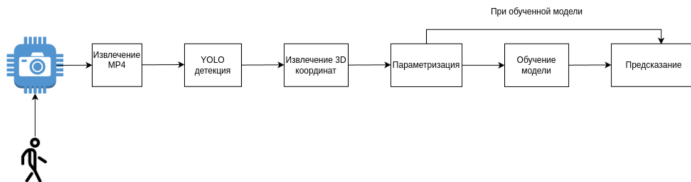


Рис.: Пайплайн обработки данных

- 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 (.npy).

В ходе исследования командой были применены следующие нейросетевые архитектуры:

- Многослойный перцептрон (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:

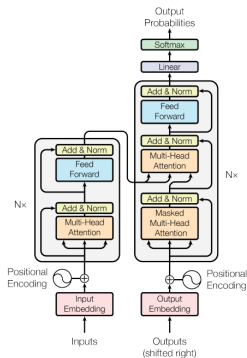


Figure 1: The Transformer - model architecture.

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

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](#)

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

Спасибо за внимание!