

## 2 слайд

Цель данной работы — исследовать применимость модели машинного обучения, способную определять ключевые точки человеческого тела на изображениях и видео.

Задачи, подлежащие решению в работе:

- Анализ литературы по методам захвата движений и преобразованию 2D ключевых точек в 3D координаты.
- Сравнительный анализ алгоритмов машинного обучения для задач захвата движения.
- Реализация алгоритма для определения 2D ключевых точек на основе выбранных моделей и оптимизаторов.
- Проведение экспериментов и анализ результатов для моделей с различными архитектурами.

## 3 слайд

Маркерный захват движения использует специальные маркеры, крепящиеся на тело актера или спортсмена, которые отслеживаются камерами, фиксирующими их положение в пространстве.

Безмаркерный захват движения анализирует видеоизображения с помощью алгоритмов машинного обучения. Этот метод требует только видеозапись с одной или нескольких камер и распознает форму тела, определяя положение суставов и ключевых точек.

Оба метода применяются для анимации, анализа спортивных движений и в медицине для диагностики и реабилитации.

- Задача алгоритма машинного обучения – максимизировать вероятность правильного определения координат.
- Процесс обучения модели заключается в минимизации функции потерь.

## 4 слайд

В данной работе использовались свёрточные нейронные сети (CNN) для пространственного распознавания и точного определения поз на отдельных

кадрах. Основой решения стала архитектура, включающая несколько компонентов для извлечения признаков, генерации тепловых карт и предсказания ключевых точек:

- Backbone сеть, извлекающая первичные признаки из входного изображения.
- RPN (Region Proposal Network) генерирующая предложения регионов интереса (RoI) на основе карты признаков.
- RoI Align - выравнивает предложения RPN и извлекает фиксированные размеры признаков для каждой области интереса.
- RoI Head состоит из дополнительных сверточных слоев (3x3 Conv) и транспонированных сверточных слоев (Transposed Conv). Эти слои генерируют тепловые карты для каждой ключевой точки. Эти тепловые карты представляют собой вероятностные распределения, где высокая интенсивность указывает на высокую вероятность нахождения ключевой точки в данном месте.
- Для классификации используется алгоритм Softmax, который оценивает вероятность принадлежности объекта к классу. Для регрессии используется билинейная интерполяция, которая корректирует параметры ограничивающей рамки.

Для обучения модели использовался оптимизатор Stochastic Gradient Descent (SGD) в сочетании с расписанием скорости обучения MultiStepLR. Функция потерь включала Smooth L1 Loss для задач регрессии и Cross-Entropy Loss для задач классификации.

## 5 слайд

В работе рассмотрены 3 метода преобразования двумерных точек в трёхмерные координаты:

Мультивидовая стереоскопия основана на анализе снимков объекта с разных ракурсов для создания трехмерных моделей, применяя алгоритмы, такие как SIFT и SURF для сопоставления точек и последующей триангуляции.

Второй метод использует стереокамеры, LiDAR или алгоритмы глубокого обучения для получения карт глубины из одиночного изображения.

Третий метод основан на использовании аннотированных данных, содержащих информацию о глубине каждого пикселя на изображении.

Для собственной реализации выбран метод использования одиночного изображения с картой глубины, полученной от модели глубокого обучения MiDaS.

## 6 слайд

На данном слайде представлены графики, демонстрирующие метрики обучения модели, основанной на архитектуре ResNet50. Эти метрики включают изменение средней потери по ключевым точкам и изменения средней точности и средней полноты по мере обучения модели.

AP (Average Precision) и AR (Average Recall) используются как метрики, потому что они позволяют оценивать модель как по точности (доля правильных положительных предсказаний среди всех положительных предсказаний), так и по полноте (доля правильных положительных предсказаний среди всех реальных положительных случаев).

Эти графики демонстрируют, что в процессе обучения модель значительно улучшила свою точность, что подтверждается снижением средней потери и ростом средней точности и полноты.

## 7 слайд

На данном слайде представлены результаты экспериментов по генерации двумерных ключевых точек с использованием различных архитектур backbone сетей, а также визуализация работы модели на изображении.

**Таблица** демонстрирует метрики точности (AP) и полноты (AR) для моделей с архитектурами ResNet18, ResNet50 и ResNet101.

**ResNet18** показала наименьшую точность среди всех трех моделей. Это объясняется ее простой архитектурой и небольшим количеством слоев, что ограничивает способность извлекать сложные признаки из изображений.

**ResNet50** продемонстрировала наилучшие результаты среди всех моделей. Значительное улучшение точности объясняется большей глубиной сети, что позволяет более эффективно извлекать признаки.

**ResNet101** показала результаты несколько хуже, чем ResNet50, несмотря на наибольшую глубину. Хотя она превзошла ResNet18 по всем метрикам, увеличение числа слоев не всегда приводит к улучшению результатов, если модель становится слишком сложной для обучения.

**Рисунок справа** иллюстрирует работу модели для определения двумерных ключевых точек.

## 8 слайд

На данном слайде представлена информация о получении карт глубины и построении трехмерного скелета на основе этих данных.

**Рисунок слева** демонстрирует карту глубины, полученную с помощью модели MiDaS. Карта глубины является визуализацией, где более темные области обозначают большую глубину (более удаленные от камеры), а светлые области находятся ближе к камере.

**Рисунок справа** представляет трехмерный скелет, созданный на основе карты глубины и определения ключевых точек. Общая картина трехмерного скелета достаточно правдоподобна и соответствует позе на исходном изображении.

## 9 слайд

В ходе работы был проведен аналитический обзор научной литературы по методам захвата движений с маркерами и без маркеров, исследованы и сравнены алгоритмы машинного обучения, а также реализован алгоритм для определения двумерных ключевых точек с использованием различных архитектур моделей. Кроме того, были изучены методы преобразования двумерных ключевых точек в трёхмерные координаты.

Эксперименты с архитектурами различной глубины показали, что ResNet50 продемонстрировала наилучшие результаты, достигая  $AP_{50} = 85,5$ ,  $AP_{75} = 69,7$ ,  $AR_{50} = 90,5$ ,  $AR_{75} = 76,0$ .

Также, была произведена апробация научного доклада по теме данной работы в рамках Молодежной научной конференции университета.