

ЛОКАЛИЗАЦИЯ ОБЪЕКТОВ НА ОСНОВЕ АНАЛИЗА СОСТОЯНИЯ КАНАЛА WI-FI

Борн Р.С.

Научный руководитель: Брагин А.Д.

Студент НИ ТПУ, ИШИТР, гр. 8В13, rsb8@tpu.ru

Ст. преподаватель НИ ТПУ, ИШИТР, bragin@tpu.ru

Аннотация

Статья представляет метод локализации объектов на основе анализа состояния канала Wi-Fi с использованием одной антенны. Предложенная нейросетевая архитектура включает: CSI энкодер, модуль слияния с механизмом внимания и обучается по парадигме учитель-ученик с псевдо-метками от YOLOv8-pose, обеспечивая определение ключевых точек человеческого тела без использования камер.

Ключевые слова: Wi-Fi сенсинг, информация о состоянии канала (CSI), локализация объектов, ключевые точки тела, дистилляция знаний.

Введение

В последние годы технологии беспроводной передачи данных получили широкое распространение, став неотъемлемой частью повседневной жизни. Одновременно с этим возрос интерес к использованию этих технологий не только для коммуникации, но и для задач сенсорного восприятия окружающей среды. Особое внимание привлекают системы, основанные на анализе состояния канала Wi-Fi (Channel State Information, CSI), которые позволяют определять положение и движения объектов в пространстве без использования традиционных средств наблюдения, таких как камеры видеонаблюдения или специализированных датчиков.

Традиционные методы локализации и отслеживания объектов в основном опираются на системы компьютерного зрения, использующие камеры. Однако такие системы имеют существенные ограничения: они требуют прямой видимости, чувствительны к условиям освещения, могут вызывать опасения, связанные с конфиденциальностью, и часто оказываются неэффективными в сложных условиях, например, при недостаточном освещении или наличии препятствий. В отличие от оптических систем, методы, основанные на анализе CSI Wi-Fi, свободны от большинства этих ограничений, что делает их перспективной альтернативой.

В настоящей статье представлен метод локализации объектов на основе анализа состояния канала Wi-Fi с использованием модели глубокого обучения. В отличие от существующих подходов, требующих нескольких точек доступа или множества антенн, предлагаемый метод позволяет эффективно определять ключевые точки человеческого тела с использованием CSI данных, полученных всего с одной антенны Wi-Fi. Это существенно упрощает практическое внедрение системы и расширяет спектр ее возможных применений.

Стенд

В данной работе использовался для сбора данных о состоянии канала Wi-Fi стенд, состоящий из двух микроконтроллерных модулей ESP32 с интерфейсом USB-C [1], оснащенных внешними антеннами (рис. 1).

Один из модулей выполняет функцию передатчика, а второй – приемника радиочастотного сигнала Wi-Fi в диапазоне 5 ГГц (стандарт IEEE 802.11ac). Антенны всенаправленного типа обеспечивают устойчивую передачу и прием сигнала в контролируемой экспериментальной среде. Конфигурация стенда позволяет осуществлять непрерывный мониторинг изменений состояния канала Wi-Fi с частотой дискретизации около 200 Гц.

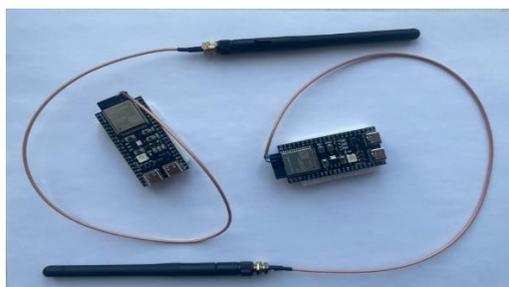


Рис. 1. Экспериментальный Стенд

Данные

Информация о состоянии канала Wi-Fi

В качестве основного источника информации используются данные о состоянии канала Wi-Fi. CSI представляет собой комплексное описание характеристик канала передачи между передатчиком и приемником, отражающее влияние физической среды на распространение радиосигнала. Эти данные содержат информацию о том, как сигнал трансформируется при прохождении через среду, включая затухание, фазовый сдвиг и другие параметры для каждой поднесущей в OFDM-модуляции (Orthogonal Frequency Division Multiplexing), применяемой в современных стандартах Wi-Fi.

Сбор данных Wi-Fi CSI

В рамках исследования был собран экспериментальный набор данных, состоящий из информации о состоянии канала Wi-Fi и соответствующих изображений. Сбор CSI данных осуществлялся путем захвата пакетов Wi-Fi в реальном времени с частотой дискретизации приблизительно 200 Гц. Для каждого временного интервала фиксировались как метаданные пакетов, так и детальная информация о состоянии канала. Метаданные включали в себя следующие параметры (таблица 1):

Таблица 1. Метаданные CSI

Обозначение	Расшифровка
type	Тип пакета
seq	Порядковый номер пакета
mac	MAC-адрес устройства-отправителя
rsqi	Уровень принимаемого сигнала
rate	Скорость передачи данных
sig_mode	Режим сигнала
mcs	Индекс модуляции и кодирования
bandwidth	Ширина канала
smoothing	Флаги сглаживания
not_sounding	Звучание
aggregation	Агрегации
stbc	Пространственно-временного блочного кодирования
fec_coding	Кодирования с прямым исправлением ошибок
sgi	Использования короткого защитного интервала
noise_floor	Уровень шума
ampdu_cnt	Количество подпакетов в A-MPDU
channel, secondary_channel	Основной и вторичный каналы
local_timestamp	Временную метку
ant	Индекс антенны
sig_len	Длина сигнала
rx_state	Состояние приемника
len	Длина полезной нагрузки CSI
first_word	Служебная информацию
data	Массив данных CSI

Наиболее важной составляющей сырых данных является массив data, содержащий комплексные значения поднесущих Wi-Fi, которые представляют собой квадратуры (I/Q-компоненты) сигнала для каждой поднесущей. В исследовании использовались данные одной антенны в частотной полосе 40 МГц, что соответствует 56 поднесущим.

Параллельно с CSI данными производилась видеосъемка с синхронизированной временной меткой, что позволило создать согласованный набор данных, где каждой CSI записи соответствует кадр видеозаписи. Эти изображения были использованы для генерации псевдо-меток ключевых точек, необходимых для обучения модели.

Временные окна и агрегация данных

Поскольку движения человека представляют собой динамический процесс, для их эффективного анализа используется подход на основе временных окон. В данном исследовании применяются окна размером 5 последовательных CSI записей, что позволяет учитывать временную динамику изменений канала связи при перемещении объектов в пространстве. Такой подход обеспечивает возможность отслеживания движений и более точной локализации ключевых точек, по сравнению с использованием изолированных CSI записей. Данные в пределах временного окна подвергаются нормализации и центрированию для устранения влияния различных аппаратных факторов и условий среды, не связанных с наличием и движением объектов. Это позволяет сконцентрироваться на изменениях CSI, вызванных непосредственно присутствием и перемещением человека в зоне действия Wi-Fi сигнала.

Преобразование CSI данных

Комплексные значения CSI, полученные из сырых данных, преобразовывались в амплитудные и фазовые компоненты. Полученные амплитудные и фазовые характеристики нормализовались и структурировались в виде тензора размерности $(1, N, 1)$, где N – количество используемых поднесущих. Первое измерение соответствует одной используемой антенне, а последнее – одному временному отсчету.

Формирование окон наблюдения

Для учета временной динамики CSI сигналов и повышения устойчивости модели к шумам был использован подход на основе скользящего окна. Для каждого момента времени t формировалось окно из W последовательных CSI наблюдений $(t-W+1, t-W+2, \dots, t)$, где $W = 5$ – размер окна. Таким образом, входные данные для модели представляли собой тензоры размерности (W, N) для амплитуд и фаз соответственно, где $W = 5$ – размер окна, а $N = 30$ – количество поднесущих. В процессе формирования окон также производилась дополнительная валидация данных и обработка краевых случаев. При недостаточном количестве поднесущих в исходных данных применялось дополнение нулями до требуемого размера.

Генерация псевдо-меток

Поскольку непосредственная разметка ключевых точек человеческого тела на основе CSI данных является сложной задачей, в исследовании был применен подход с использованием псевдо-меток. Для этого на синхронизированных по времени изображениях применялась предобученная модель YOLOv8-pose [2], которая позволяет детектировать 17 ключевых точек человеческого тела (включая нос, глаза, уши, плечи, локти, запястья, бедра, колени и лодыжки). Полученные ключевые точки сохранялись в формате плоского вектора $[x_1, y_1, x_2, y_2, \dots, x_{17}, y_{17}]$, где (x_i, y_i) – координаты i -й ключевой точки в пространстве изображения. Эти координаты нормализовались относительно размеров изображения, что обеспечивало инвариантность к масштабу. Синхронизация между CSI данными и изображениями осуществлялась на основе временных меток, что позволяло сопоставить каждому окну CSI данных соответствующие координаты ключевых точек. Таким образом, для каждого временного окна формировался обучающий пример, состоящий из входных данных (амплитуды и фазы CSI) и целевых значений (координаты ключевых точек).

Структура набора данных

Итоговый набор данных, использованный для обучения и валидации модели, имел следующую структуру:

1. Входные данные:

- Амплитуды CSI: тензор размерности (B, W, N)
- Фазы CSI: тензор размерности (B, W, N), где B – размер обучающего батча, W = 5 – размер временного окна, N = 30 – количество поднесущих.

2. Целевые значения:

- Координаты ключевых точек: тензор размерности (B, 34), где 34 соответствует 17 ключевым точкам по 2 координаты (x, y).

Общий объем экспериментального набора данных составил несколько временных последовательностей, каждая содержащая в среднем 100 CSI записей, что после применения метода скользящего окна позволило сформировать обучающий набор для разработки и валидации модели.

Архитектура модели и ее обучение

Предлагаемая архитектура, представляет собой композицию нескольких нейросетевых модулей: энкодера для CSI данных (CSIEncoder), энкодера для ключевых точек (KeypointsEncoder) и модуля мультимодального слияния (MultimodalFusion).

Общая архитектура модели состоит из следующих основных компонентов:

1. CSI энкодер: извлекает признаки из амплитудных и фазовых компонент CSI данных.
2. Энкодер ключевых точек: используется только на этапе обучения для обработки псевдо-меток.
3. Модуль мультимодального слияния: объединяет признаки из различных источников с использованием механизма внимания.

4. Выходной слой: преобразует объединенные признаки в предсказания координат ключевых точек.

Для обучения предложенной модели используется парадигма учитель-ученик (Teacher-Student Training), которая позволяет эффективно передавать знания от предварительно обученной модели компьютерного зрения к модели, работающей с CSI данными.

В данном подходе роль учителя выполняет предварительно обученная модель YOLOv8-pose, которая анализирует визуальные данные (изображения) и предоставляет высокоточные оценки положения ключевых точек человеческого тела. Эти оценки используются в качестве псевдо-меток для обучения модели-ученика, работающей исключительно с CSI данными.

Процесс обучения включает следующие этапы:

1. Генерация псевдо-меток: модель-учитель (YOLOv8-pose) применяется к набору изображений, синхронизированных с CSI данными, для получения координат ключевых точек. Только оценки с уверенностью выше порогового значения (0.5) принимаются во внимание.
2. Дистилляция знаний: модель-ученик обучается минимизировать расхождение между своими предсказаниями и псевдо-метками учителя. В качестве функции потерь используется среднеквадратичная ошибка (MSE).

Такой подход позволяет модели научиться извлекать релевантную информацию о позе человека непосредственно из CSI данных, без необходимости доступа к визуальным данным во время инференса.

Результаты

Для оценки качества работы модели используется процентное соотношение корректно определенных ключевых точек (РСК), т.е. доля ключевых точек, предсказанных с ошибкой не более заданного порога.

Модель демонстрирует следующие показатели эффективности на тестовом наборе данных:

1. Среднеквадратическая ошибка (MSE) составила 0.0274 для нормализованных координат, что соответствует средней абсолютной ошибке приблизительно 4.8 см в реальном физическом пространстве при условиях экспериментальной среды размером 3×4 м.

2. Метрика РСК@0.5 составила 76.3 %, что указывает на то, что более трех четвертей ключевых точек локализованы с точностью до 50% от размера ограничивающей рамки человека.

Анализ результатов по отдельным ключевым точкам выявил закономерность в точности их локализации: наиболее точно определялись крупные суставы (плечи, бедра) и центральные точки тела (например, центр туловища), в то время как дистальные точки конечностей (запястья, лодыжки) и лицевые ориентиры (нос, глаза) локализовались с относительно большей погрешностью. Данное наблюдение коррелирует с физическими характеристиками отражения и дифракции радиоволн от объектов различного размера и свойств.

1. Сравнение с существующими методами

Сопоставление результатов с аналогичными исследованиями в области Wi-Fi сенсинга [3] показало, что предложенная модель достигает сопоставимой точности при существенно меньших аппаратных требованиях. В частности, в сравнении с методом «DensePose From WiFi» [4], требующим множественных точек доступа, предложенное решение с одной антенной демонстрирует лишь незначительное снижение точности (приблизительно на 8-12 % по метрике РСК), при этом значительно упрощая практическую реализацию системы.

2. Вычислительная эффективность

Оценка вычислительной эффективности модели показала, что время инференса на современном CPU составляет в среднем 23 мс на образец, что позволяет осуществлять обработку в режиме, близком к реальному времени, при частоте дискретизации CSI данных около 200 Гц. При использовании графического ускорителя время инференса сокращается до 5 мс, что обеспечивает возможность применения модели в системах, требующих обработки данных в реальном времени.

Размер модели в сериализованном виде составляет около 12 МБ, что делает её пригодной для развертывания на устройствах с ограниченными вычислительными ресурсами, включая встраиваемые системы и устройства Интернета вещей.

Заключение

Проведенная экспериментальная оценка подтверждает эффективность предложенного подхода к локализации ключевых точек человеческого тела на основе анализа CSI данных, полученных с одной антенны Wi-Fi. Модель демонстрирует достаточно хорошую точность в определении общей позы и положения основных суставов, сохраняя работоспособность в различных сценариях, включая отсутствие прямой видимости.

Полученные результаты свидетельствуют о потенциальной применимости разработанной технологии в практических системах умного дома, здравоохранения и человеко-машинного взаимодействия, где требуется неинвазивное отслеживание положения и движений человека без использования оптических средств наблюдения.

Список использованной литературы

1. Espressif. esp-csi: CSI (Channel State Information) Tools for ESP32 [Электронный ресурс]. – URL: github.com/espressif/esp-csi (дата обращения 15.01.2025).

2. Ultralytics. YOLOv8: Supported Tasks and Modes [Электронный ресурс]. URL: <https://docs.ultralytics.com/ru/models/yolov8/#supported-tasks-and-modes> (дата обращения 15.01.2025).

3. Кучин И.Ю. Разработка системы позиционирования и контроля объектов с помощью беспроводной технологии Wi-Fi / Кучин И.Ю., Иксанов Ш.Ш., Рождественский С.К., Коряков А.Н. // Системы анализа и обработки данных. – 2015. – № 3 (60).

4. Jiaqi Geng, Dong Huang, Fernando De la Torre // arXiv:2301.00250 [Электронный ресурс]. – URL: arxiv.org/abs/2301.00250 (дата обращения 15.01.2025).