Разработка методов навигации для роботехнических систем на основе обучения с подкреплением и оптимального управления

Осколков Н.С.
oskolkov.ns@phystech.edu

Научный руководитель: канд. физ. наук Макаров Д.А

Центр когнитивного моделирования МФТИ

2025 – Научно-исследовательский семинар



Мотивация



Рис.: Колесный робот помощник "Temi Personal Robot"

Цель исследования

Целью исследования ставится используя оптимальное управление обучить с помощью обучение с подкреплением колесного робота доезжать до выбранного текстовой командой объекта из ограниченного множества объектов, избегая столкновения с случайно расставленными статическими препятствиями на сцене.

Формальная постановка задачи

Мы моделируем задачу навигации как частично наблюдаемую марковскую задачу принятия решений (POMDP): (S, O, A, P, R, γ). S — реальное состояние, O — наблюдаемое состояние, A — действия, R — вознаграждения, P — динамика переходов $(O \times A \to S)$, $\gamma \in (0, 1]$ — коэффициент дисконтирования.
Цель: найти оптимальную стратегию π, максимизирующую вознаграждение:

$$G_\pi = \mathbb{E}_{\pi,\tau}\left[\sum_{t=0}^{T-1} \gamma^t r(O_t, a_t)\right] \tag{1}$$

Задача: робот должен достичь цели, избегая препятствий , используя данные сенсоров. Эпизод завершается при достижении цели.

Формальная постановка задачи

1. **пространства наблюдений** $\mathcal{O}$:
   a) Данные с камеры - тензор формата $(400 \times 400 \times 3)$ - X_{cam},
   b) Текстовая команда X_{goal},
   c) Граф сцены G (наименования, положения и пространственные характеристики),
   d) Скорость базы $V_{current}$(2 dof): скорость состоит из двух компонент: линейной и угловой.
2. **пространства действий** A: скорость $V_{control} \in \overline{V}$, где $\overline{V}$ - класс допустимого управления,

$$V_{control} = f(X_{cam}, X_{goal}, G, V_{current}).$$

Выделенные подзадачи

1. реализовать обучение с расписанием ✓
2. реализовать имитационное обучение ✓
    1. реализовать глобальное планирование пути (оптимальное планирование) ✓
    2. реализовать алгоритм следования (эврестическое управление) ✓
3. реализовать граф сцены и локализацию внутри графа
4. реализовать кратковременную память
5. реализовать долговременную память
6. реализовать адаптивное управление

Описание используемого метода

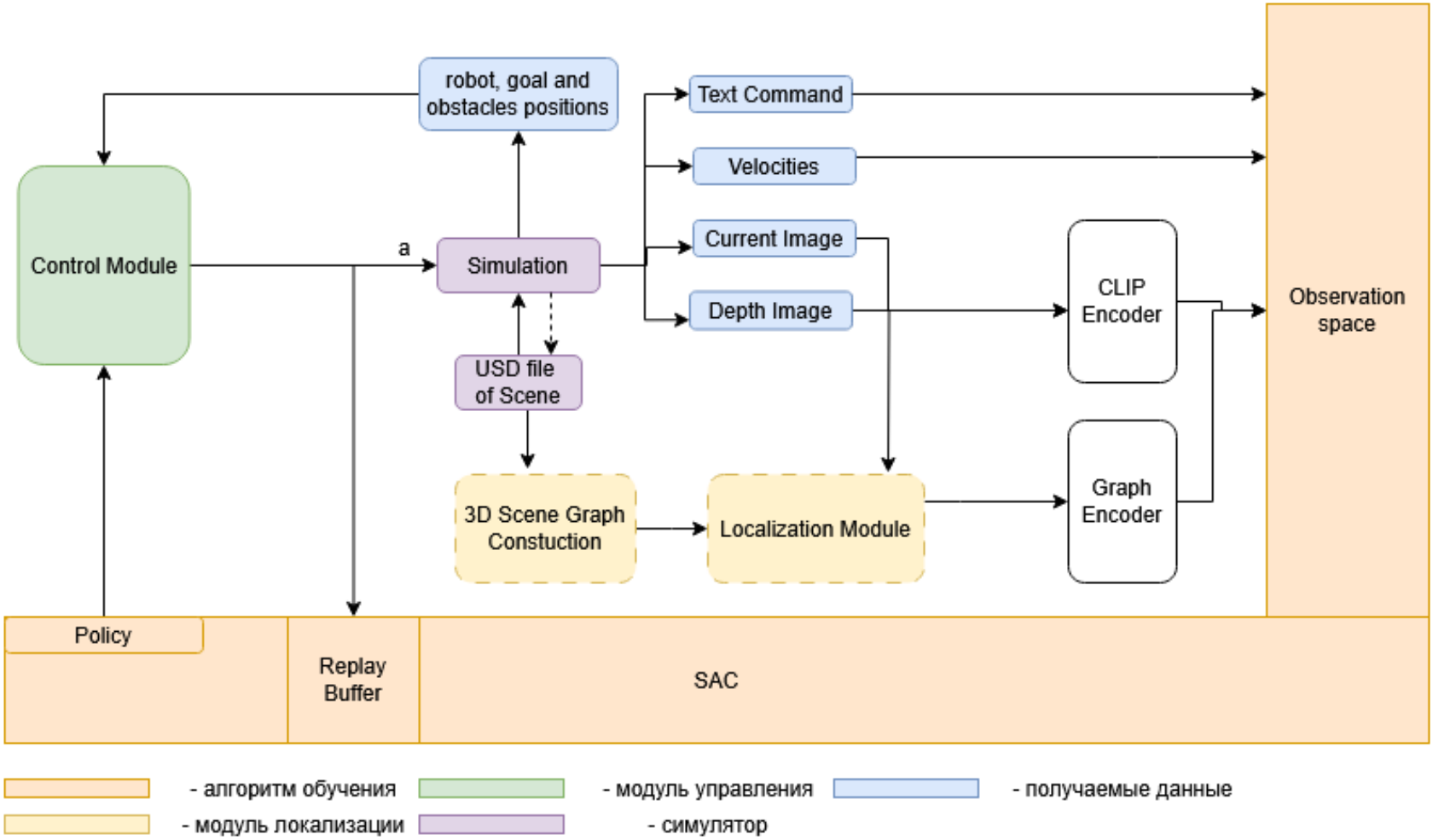

Рис.: Общая схема пайплайна

Обзор используемых методов

Способы взаимодействия MPC и обучения с подкреплением:[1]

a) MPC как актор
b) MPC как модельное предсказательное управление
c) MPC как критик

[1] Synthesis of Model Predictive Control and Reinforcement Learning: Survey and Classification (2025)

1: Обучение с подкреплением

Выбранный алгоритм: SAC

Функция вознаграждения.

1. Подзадача 1: ошибка по расстоянию < 1 м.
2. Подзадача 2: ошибка по углу $< \frac{\pi}{20}$

5 если выполнены **все** подзадачи, -1 если **одна** из подзадач, иначе -2.

2: Модуль контроля (имитационное обучение)

Дейкстра

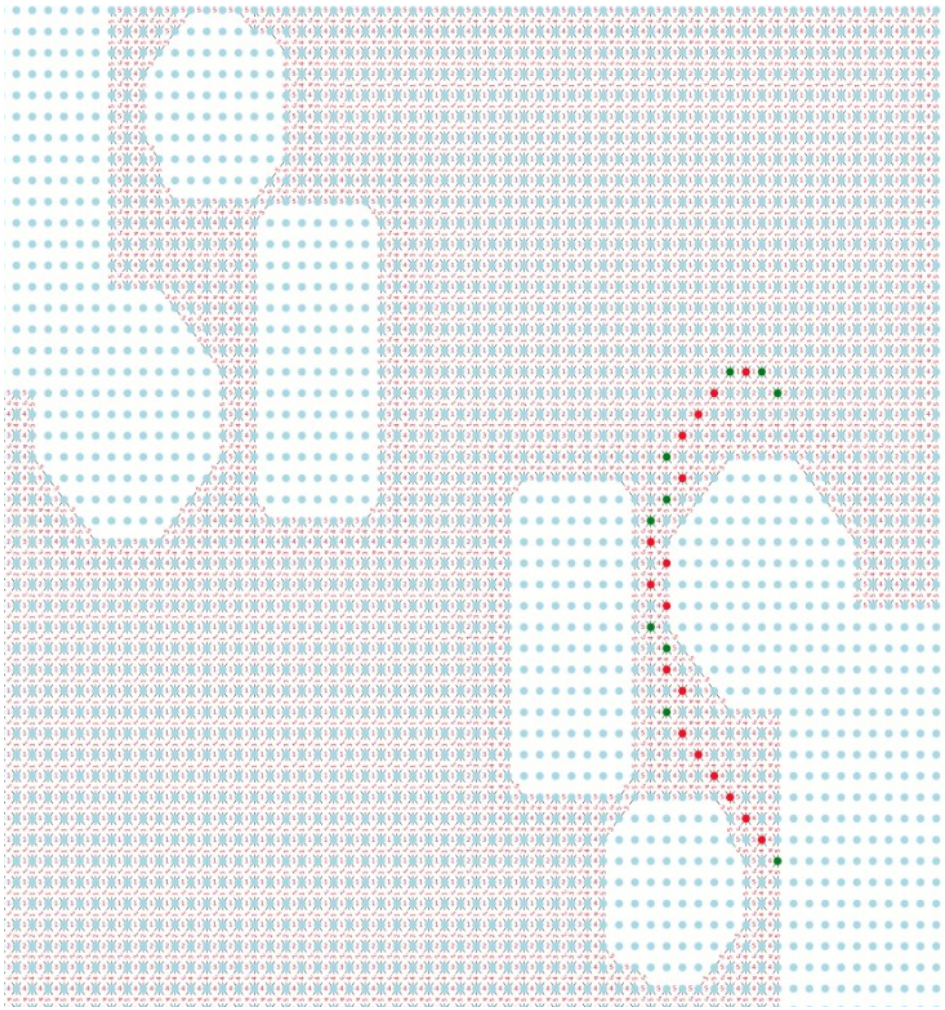

Рис.: Пример построенного алгоритмом Дейкстры и спрямленного пути

2: Модуль контроля (имитационное обучение)

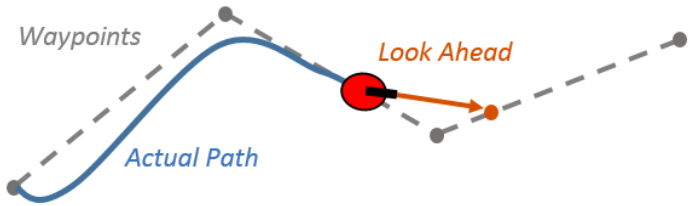

Рис.: Алгоритм следования Pure Persuit[2]

1. **Кривизна траектории**:

$$\gamma = \frac{2 \cdot \sin(\alpha)}{L_d}; \delta = f(\gamma) \tag{2}$$

- γ: Кривизна траектории (обратный радиус поворота, $\gamma = \frac{1}{R}$).
- δ: Угол курса.
- α: Угол между текущей ориентацией транспортного средства и направлением на look-ahead точку.
- L_d: Расстояние до look-ahead точки.

[2] Coulter RC. Implementation of the pure pursuit path tracking algorithm.

3: Обучение с расписанием

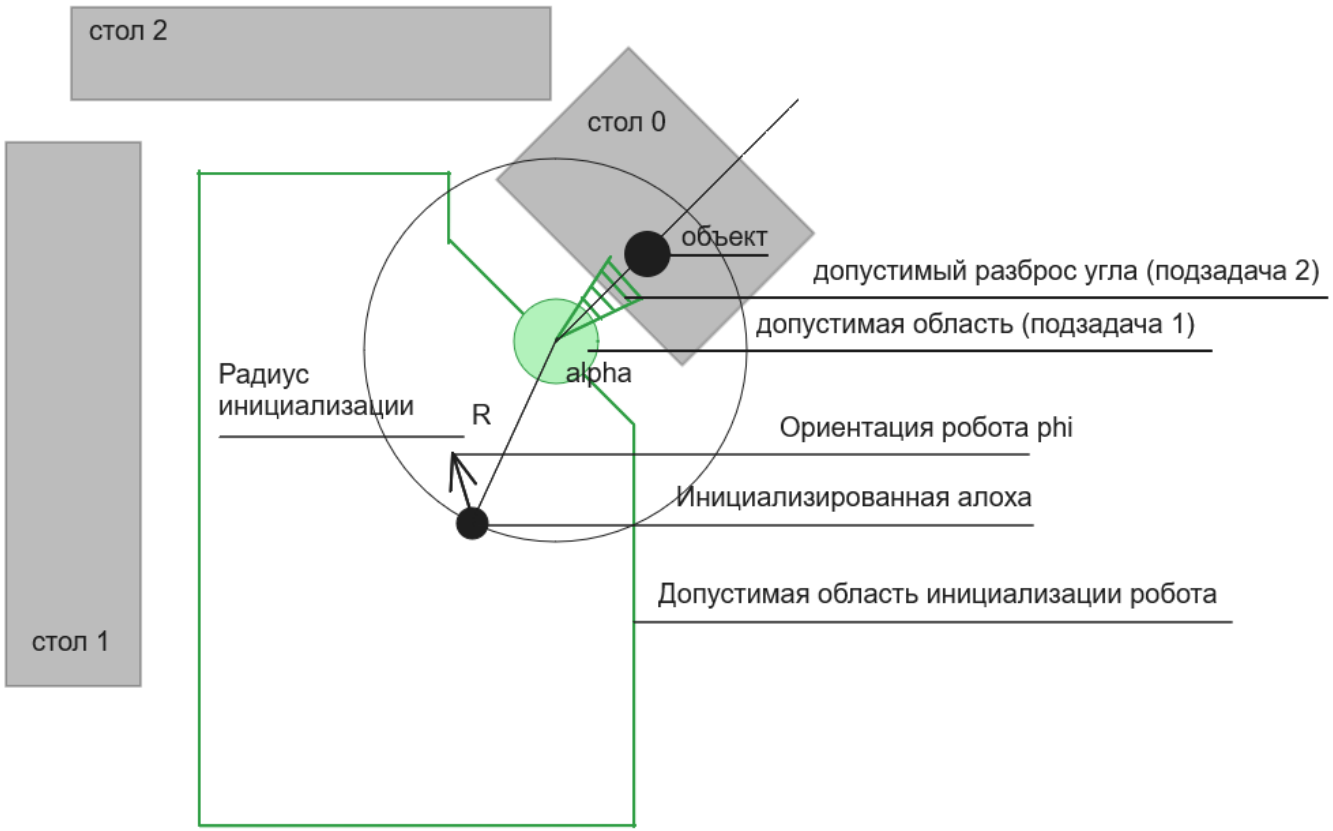

Рис.: Иллюстарция уровня сложности

4: Графы сцены, BBQ

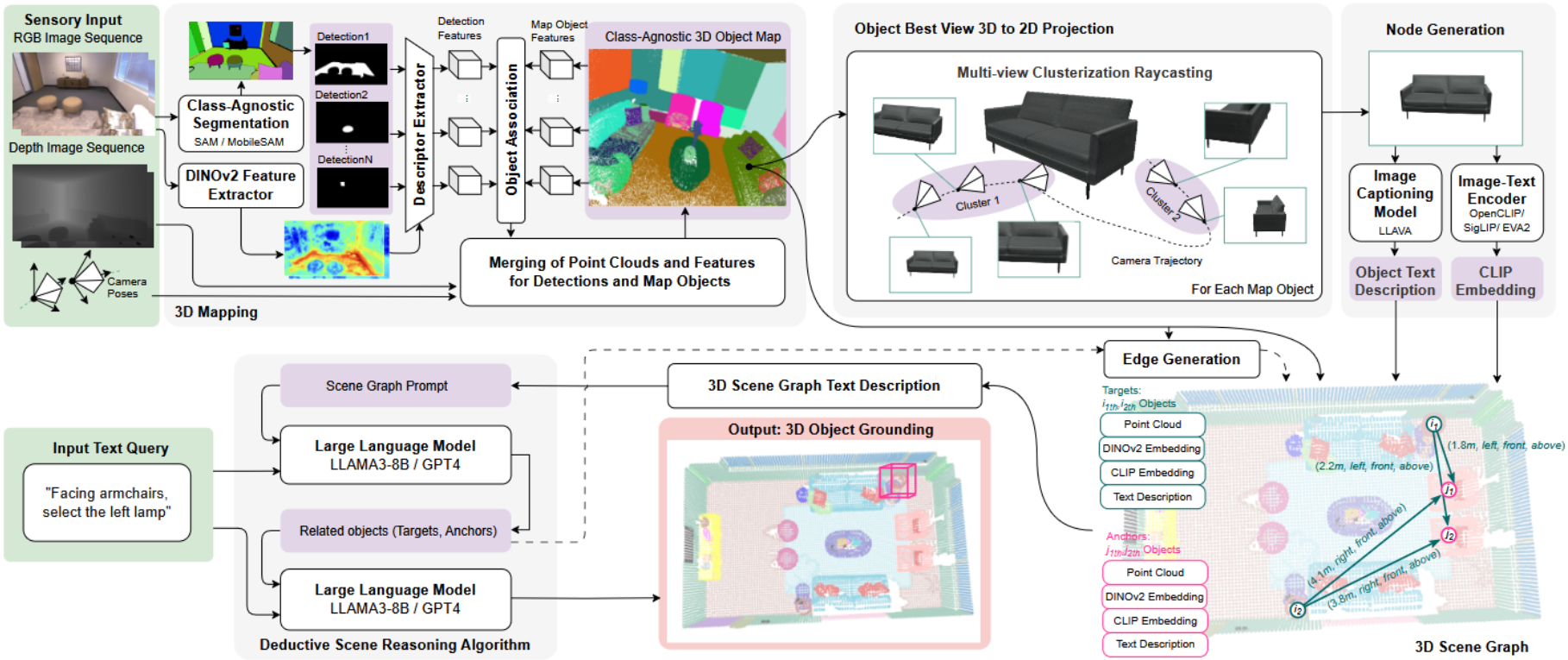

Рис.: схема BBQ[3]

[3] Beyond Bare Queries: Open-Vocabulary Object Grounding with 3D Scene Graph

5: Модуль локализации

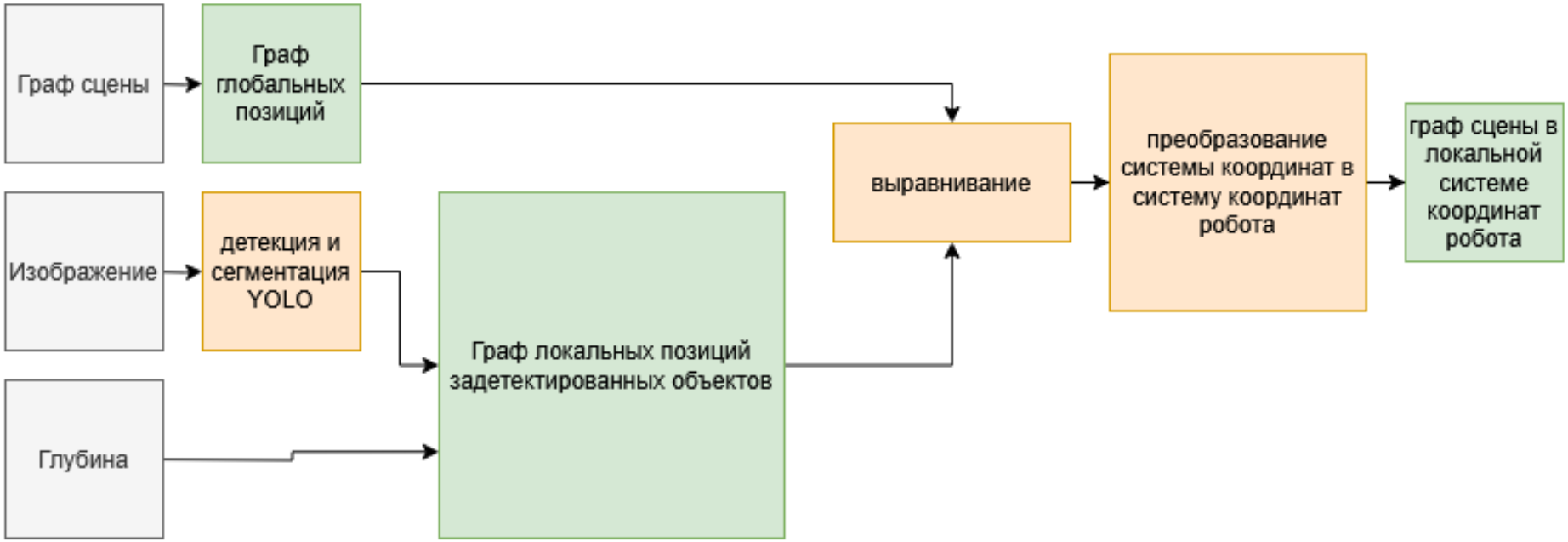

Рис.: Схема модуля локализации

Функция ошибки в методе минимизации Пауэлла:

$$\text{ошибка} = \sum_i \min_{j:\text{name}_i=\text{name}_j} \|(R \cdot \text{local}_i + [tx, ty]) - \text{global}_j\|$$

где R- матрица поворота, tx, ty - смещение локальной с.к. local относительно глобальной global).

Эксперименты

Технические характеристики: эксперементы производились в среде Isaac Sim на колесном роботе Mobile Aloha[4]





Рис.: Изображение с камеры робота в эксперимента с множеством выборов

[4] Fu Z, Zhao TZ, Finn C. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation.

Эксперименты

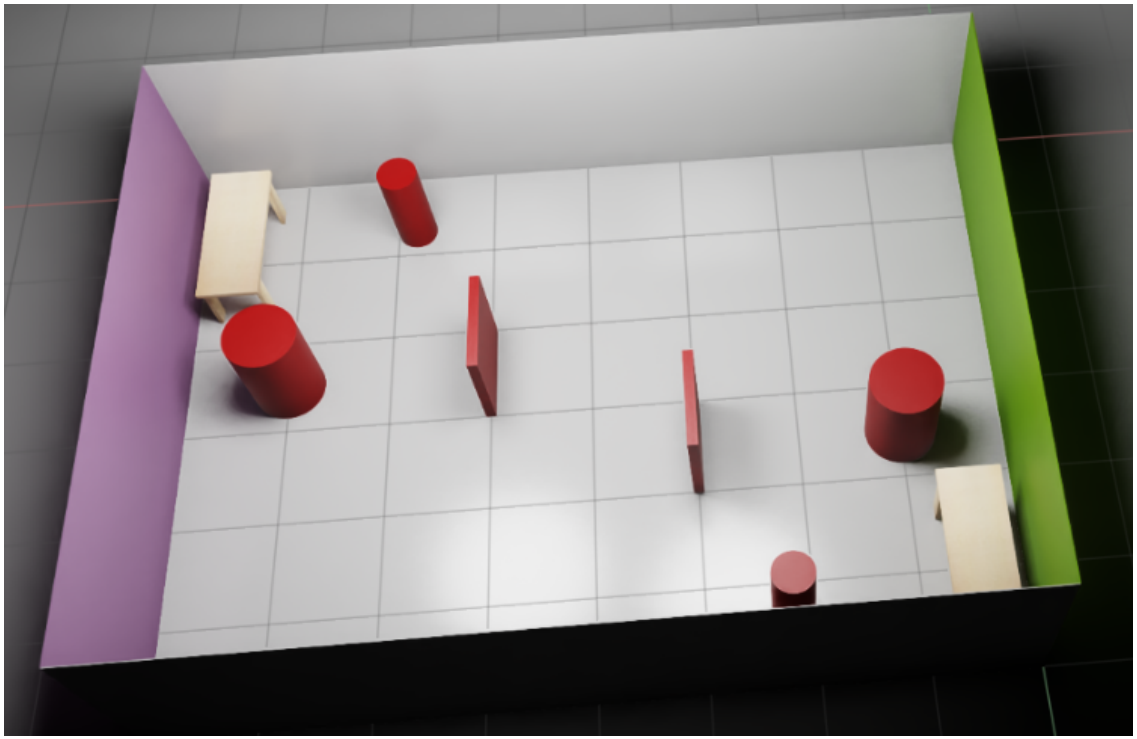

Рис.: Эксперимент со статическими препятствиями

Эксперименты (Со статическими препятствиями)

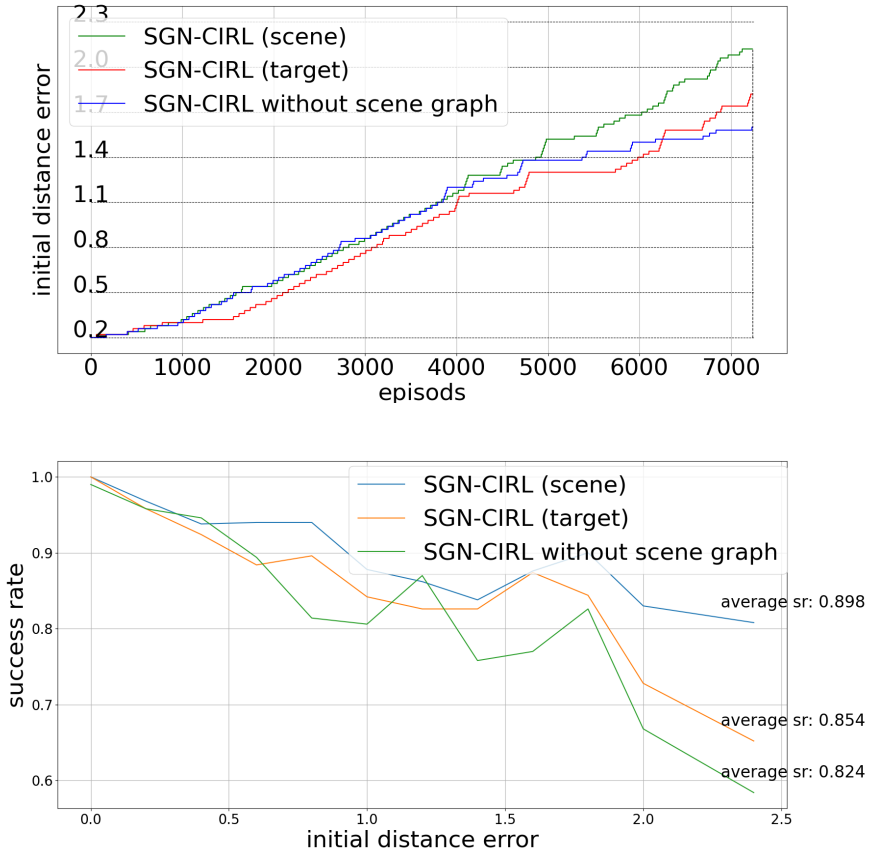

Рис.: Графики достигнутого уровня сложности от кол. эпизодов во премя обучения и процента успеха от стартовой дситанции между агентом и целевым объектом во время инференса (задействованы все модули кроме локализации)

Эксперименты (С случайными позициями препятствий)

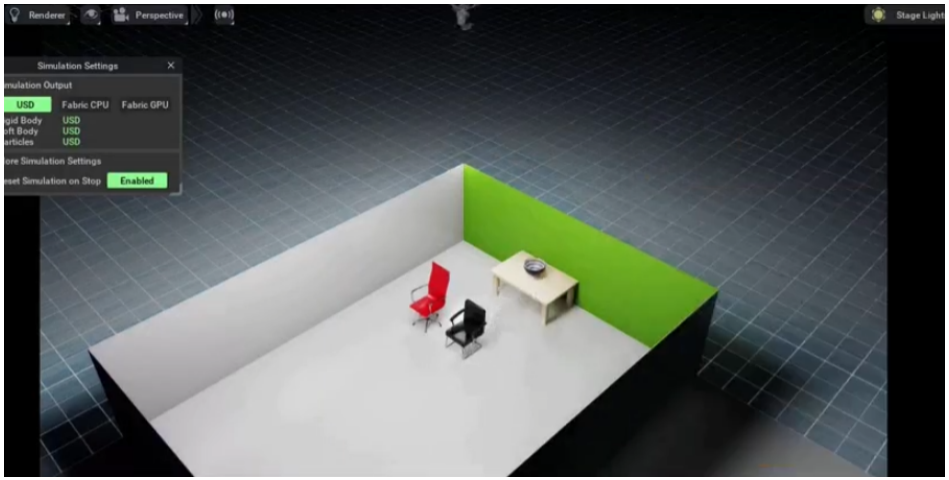

Рис.: Случайные препятствия

Эксперимент (Наличие модулей)	**Процент успеха**
CL	0.06
CL, IL	0.20
CL, IL, G	0.92
CL, IL, G, L	0.36

где, CL - обучение с расписанием, IL - модуль контроля, G - граф сцены, L - модуль локализации