Разработка быстрых методов 3D-трекинга объектов с использованием их расширенных векторных представлений

Титков Р.Е.
titkov.re@phystech.edu

Центр когнитивного моделирования МФТИ

2024 – Научно-исследовательский семинар

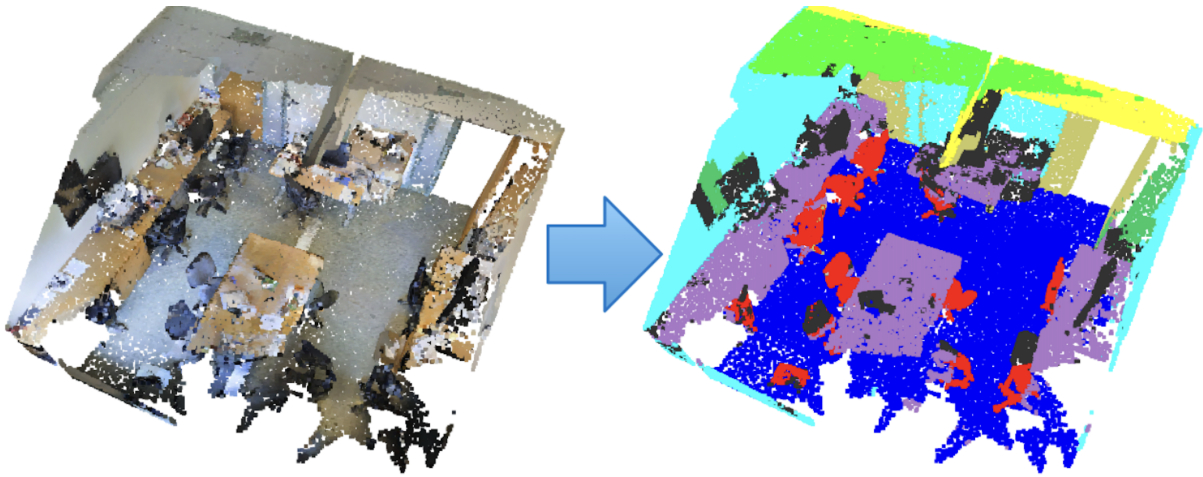
Введение

Мотивация

- Развитие автономных систем требует высокоточных и производительных алгоритмов трекинга.
- Современные методы 3D-трекинга имеют проблемы с:
  - низкой производительностью,
  - неточностью форм объектов в условиях шума,
  - интеграцией семантики в сложных сценах.

Цель работы

- Разработка **эффективного 3D-трекинга** с использованием визуальных и семантических представлений объектов.



Метрики оценки качества трекинга

- IDSW (ID Switches) - количество ложных обменов идентификаторов между кадрами.
- MOTA (Multiple Object Tracking Accuracy) - общая мера точности отслеживания, учитывающая ошибки детекции, ассоциации и потери объектов.

$$\text{MOTA} = 1 - \frac{\text{FN} + \text{FP} + \text{IDSW}}{\text{GT}}$$

Где:

- FN – количество пропущенных объектов (False Negatives)
- FP – количество ложных обнаружений (False Positives)
- IDSW – количество ложных обменов идентификаторов (ID Switches)
- GT – общее количество истинных объектов (Ground Truth)

Постановка проблемы

Дано: Целевые объекты S и их номера треков $Track_S$

Необходимо: Обнаружить объекты T и для каждого объекта определить номера треков $Track_T$

Или решить задачу оптимизации:

$$\max_{T, \text{Track}_T} \text{MOTA}(S, \text{Track}_S, T, \text{Track}_T)$$

Визуализация трекинга



Рис.: Пример последовательности кадров

Эволюция исследуемого подхода

1. **Closed-Vocabulary 2D Трекинг**
2. **Closed-Vocabulary 2D Трекинг + CLIP**: Добавление реидентификации
3. **3D Semantic Gaussian Splatting:** Open-Vocabulary трекинг в 3D
4. **Realtime 3D Semantic Gaussian Splatting:** Оптимизация для работы в реальном времени

2D Алгоритм BoT-SORT

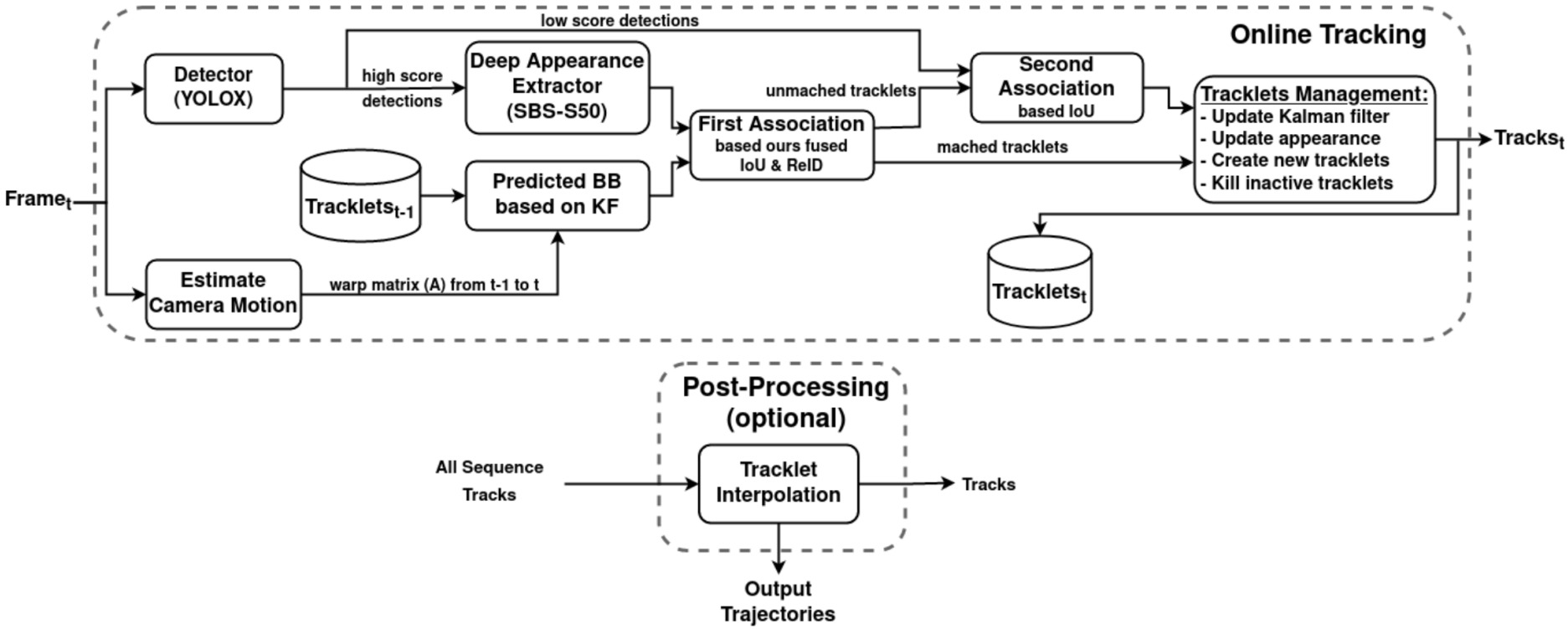
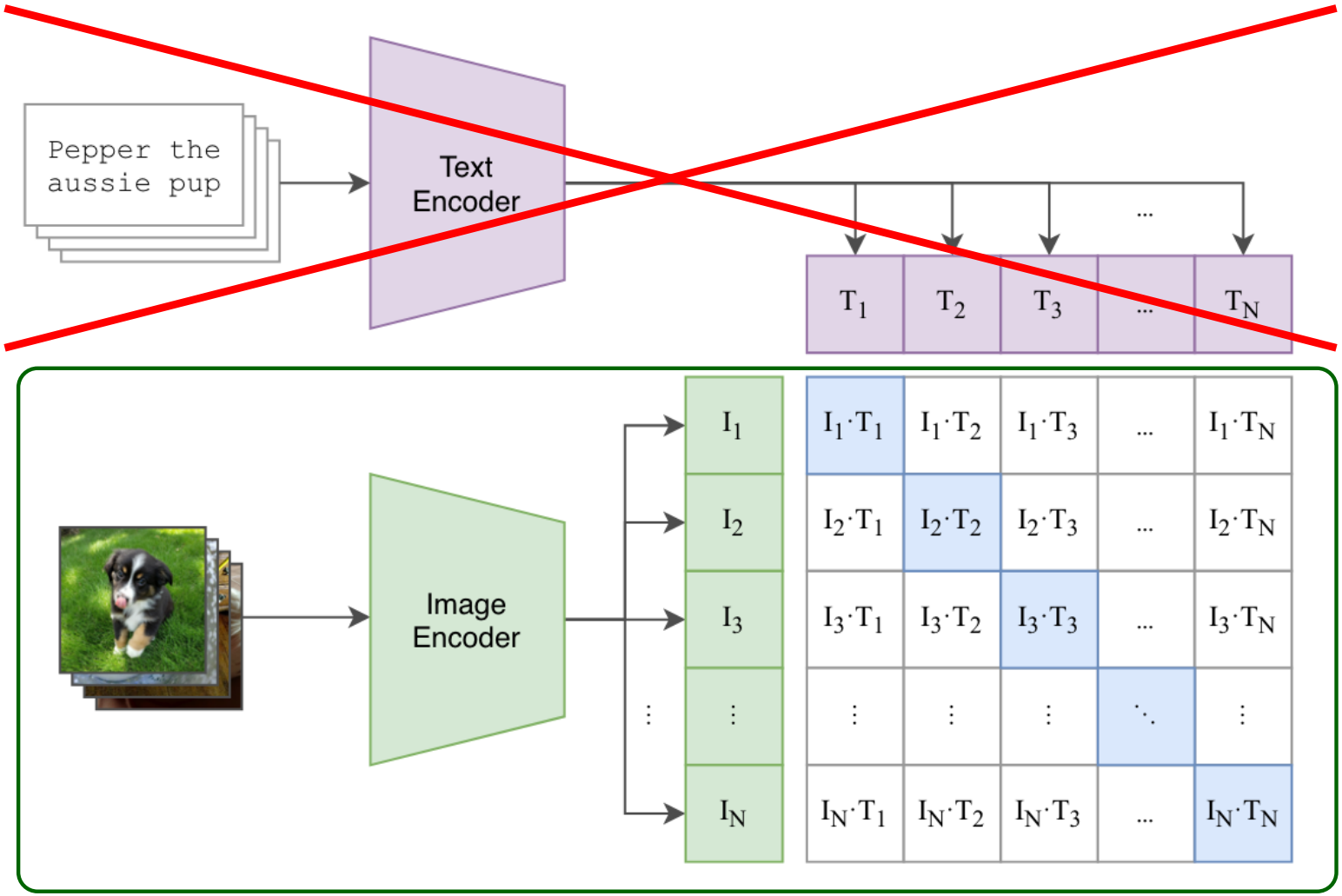


Рис.: Принцип работы алгоритма BoT-SORT

Clip Encoder

В этом эксперименте используется только Image Encoder для получения векторного представления изображения.



Результаты использования CLIP и CLIP Encoder

Method	Same object		Different object		Diff - Same
	Mean	Std	Mean	Std	
Encoder, cropped	0.036	0.016	0.088	0.018	0.052
Clip, cropped	0.114	0.080	0.115	0.102	0.001
Encoder, masked	0.044	0.018	0.080	0.024	0.036
Clip, masked	0.090	0.071	0.180	0.146	0.09
Encoder, offset 15	**0.049**	**0.016**	**0.129**	**0.037**	**0.08**
Clip, offset 15	0.182	0.141	0.173	0.106	0.009

Таблица: Среднее косинусное расстояние для различных способов получения визуальных эмбеддингов

Сравнение количества созданных треков

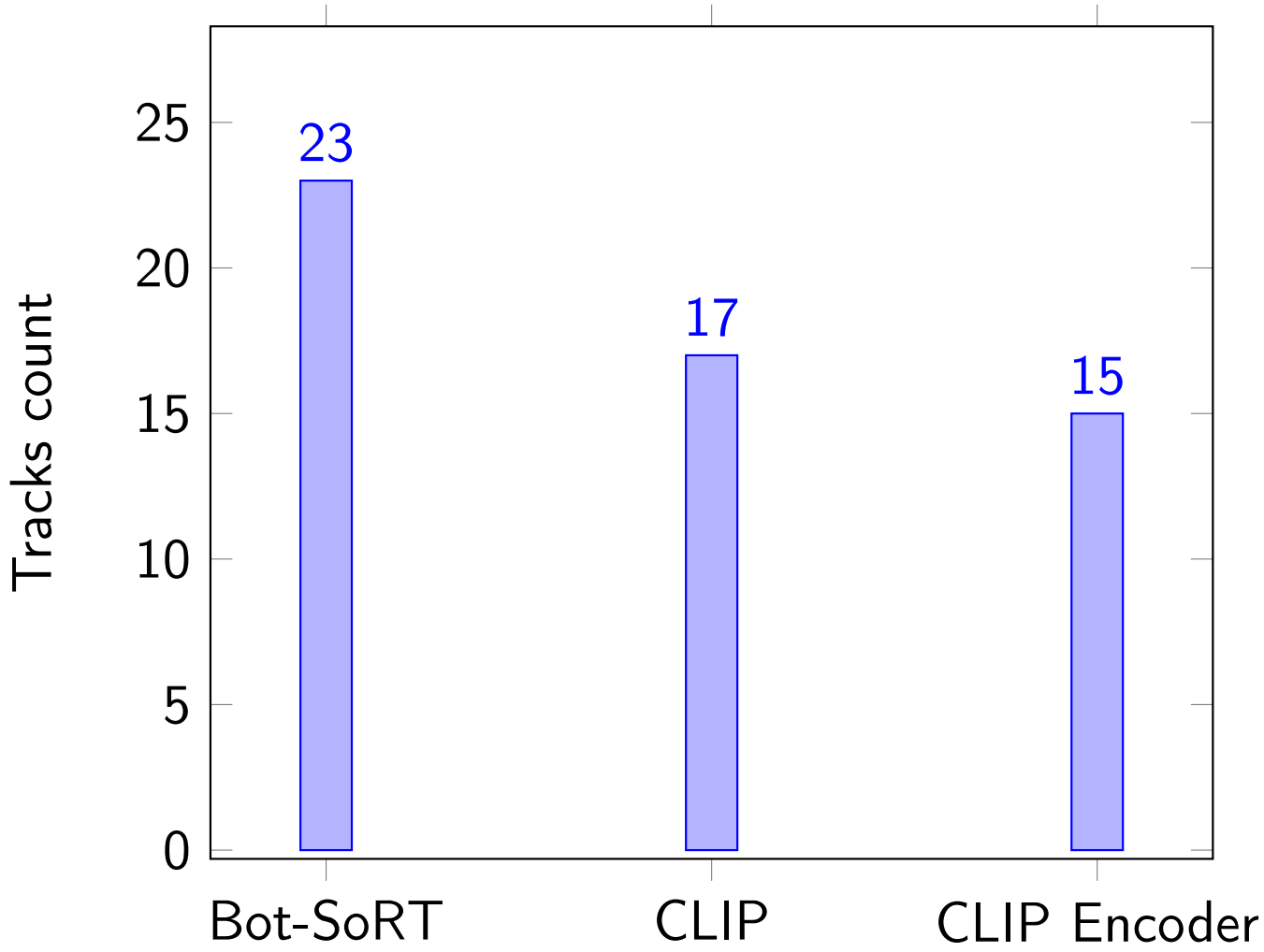
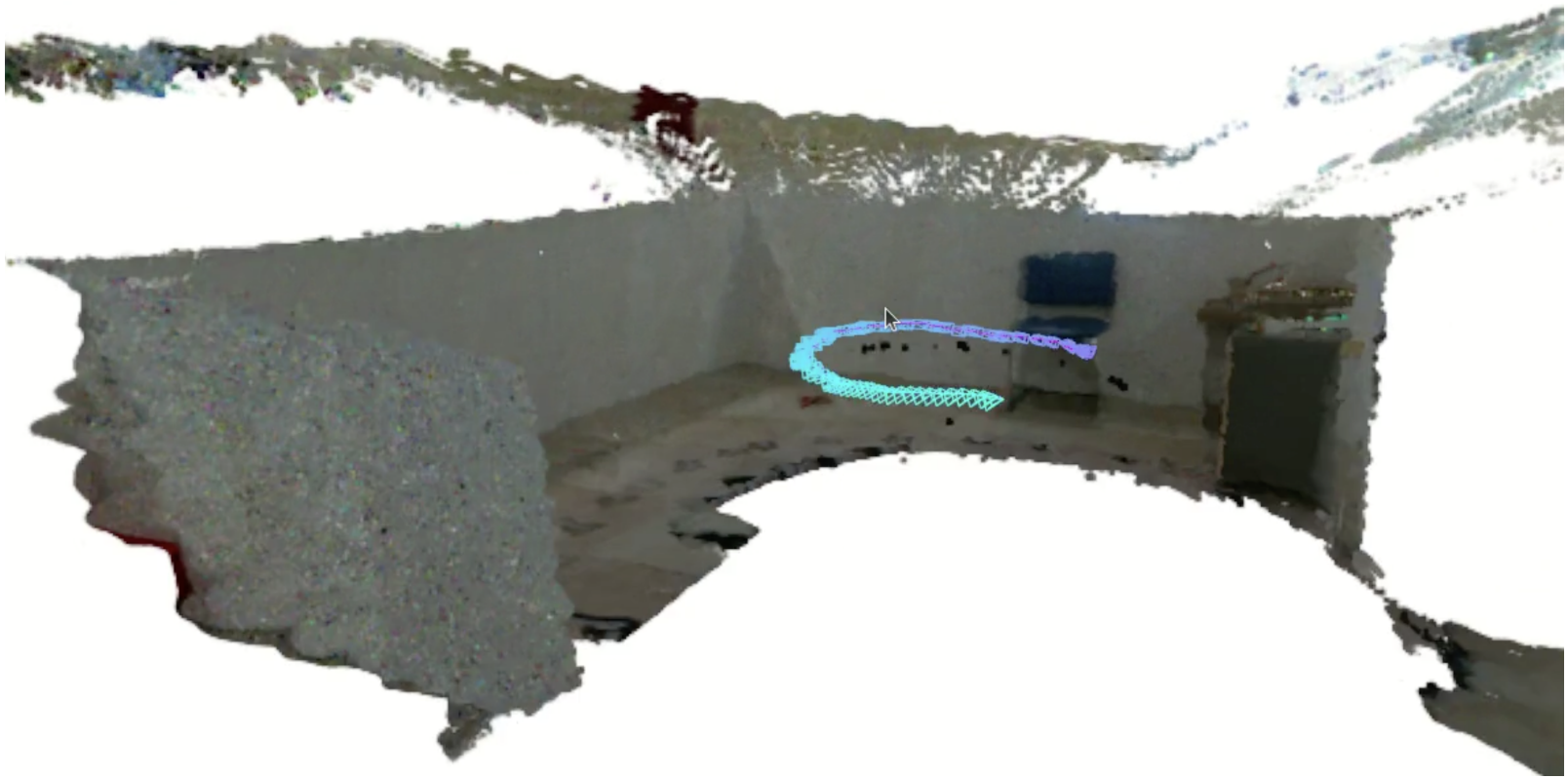


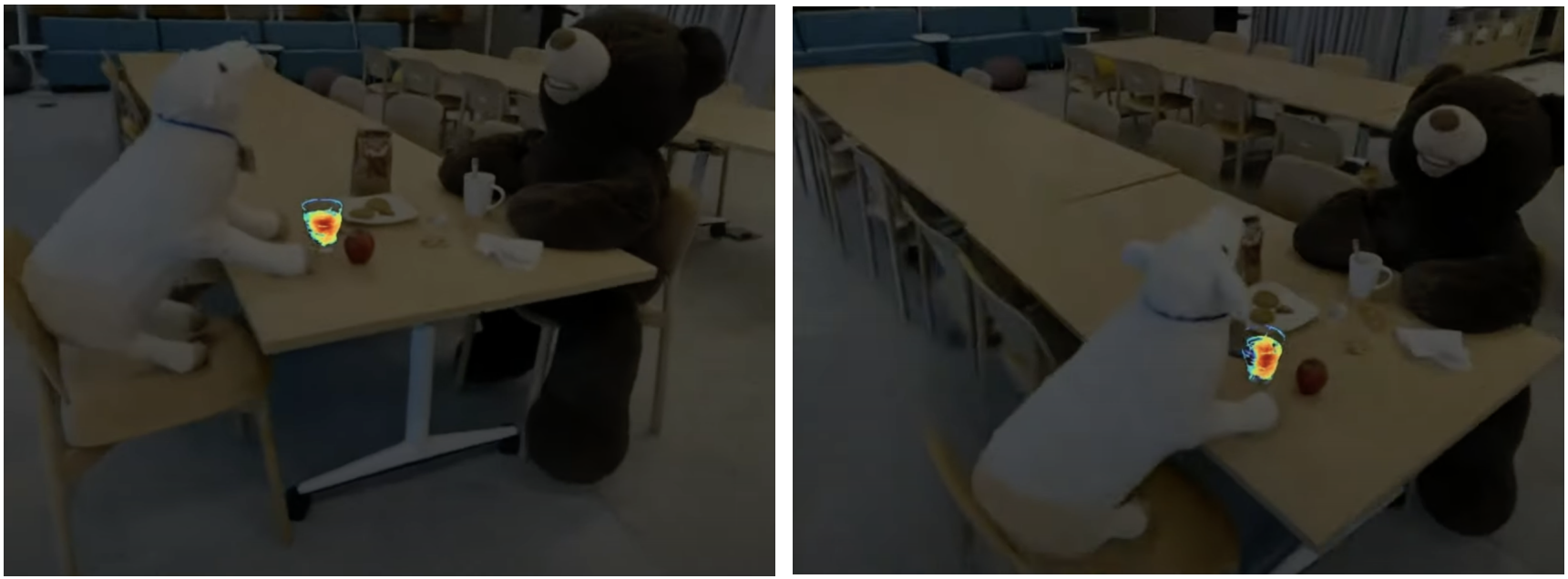
Рис.: Количество созданных треков для сцены с 13 объектами

3D Реконструкция сцены с помощью Gaussian Splatting



Трекинг с использованием Semantic Gaussian Splatting

- Использование CLIP-эмбеддингов позволяет хранить **семантическую информацию** в 3D-реконструкции.
- Локализация объектов на основе сравнения эмбеддингов изображения и текста.
- Эмбеддинги позволяют реидентифицировать объекты при движении или повторном появлении.



Локализации объекта “tea mug”

LangSplat: 3D Language Gaussian Splatting

Особенности LangSplat

- \+ Использование семантики в 3D-реконструкции
- \+ **Open-Vocabulary** запросы к сцене
- – Медленное построение реконструкции
- – RGB и семантика строятся разными моделями



Визуализация работы LangSplat

LangSplat: 3D Language Gaussian Splatting

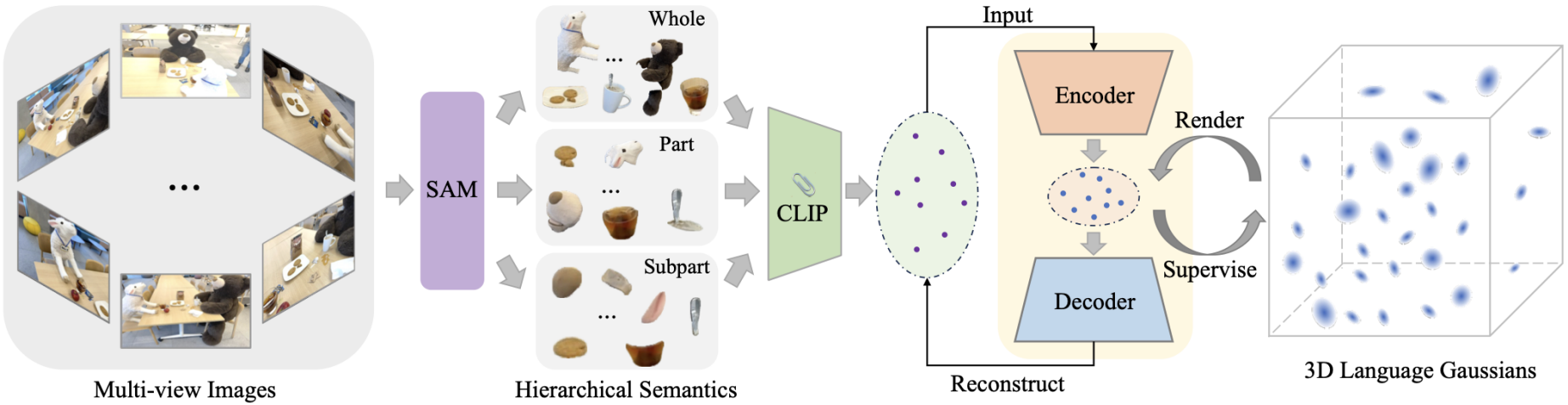


Схема работы LangSplat

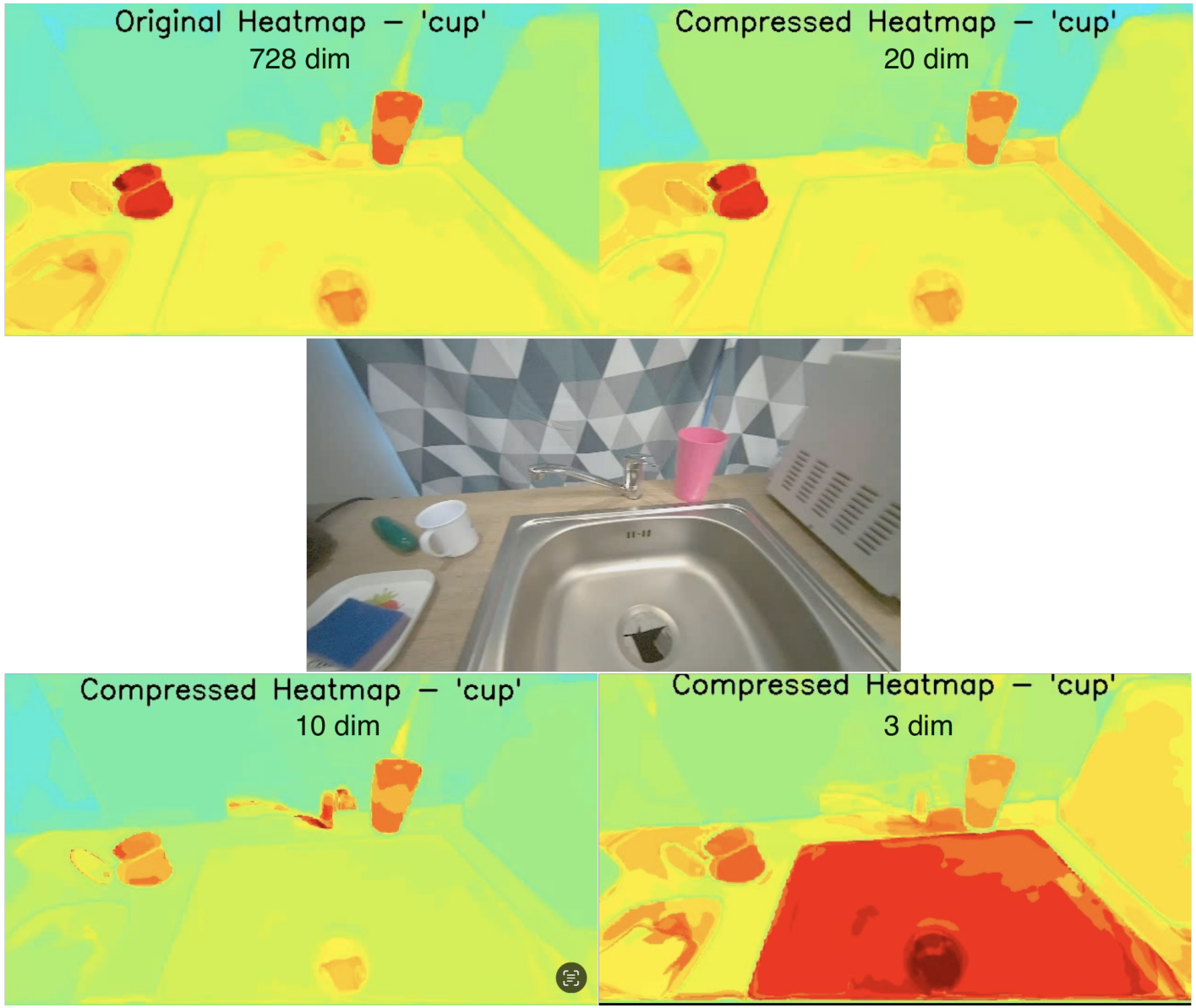
Language-Enhanced Gaussian Splatting for Real-Time SLAM

Ключевые особенности предлагаемого решения:

- \+ Использование семантики в 3D-реконструкции
- \+ **Open-Vocabulary** запросы к сцене
- \+ Реконструкция в реальном времени
- – Сжатое представление эмбеддингов может снижать точность семантического анализа



Эксперименты с разной степенью сжатия эмбеддинга



Input 768 dim -> Encoder N dim -> Decoder 768 dim

Дальнейшие работы

- Исследовать пробелмы трекинга в сценах, реконструированных с помощью Semantic Gaussian Splatting.
- Интегрировать семантический энкодер в пайплайн сплаттинга.
- Завершить реализацию растеризатора с поддержкой произвольного числа каналов для нанесения семантики на сцену.

Спасибо за внимание!

titkov.re@phystech.edu