

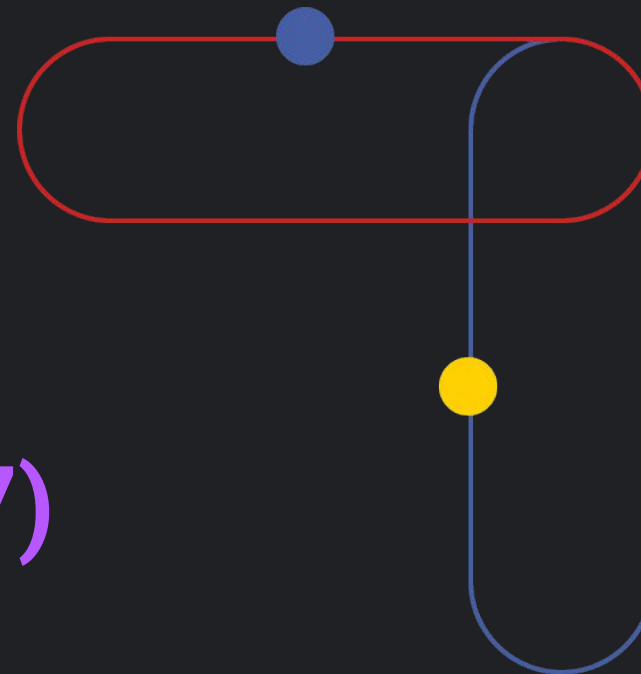


Contrast Analytics (0407)

Е. Мастер кадра (МТС)

Литвинов Вячеслав, МОВС НИУ ВШЭ

Бурлова Альбина, МОВС НИУ ВШЭ



На пути к решению

- Сложности парсинга
- Поиск картинок с такими же ракурсами и классами
- Добавление 16-го класса
- Описание данных ограничено
- Гарантии качества модели



Что попробовали в начале

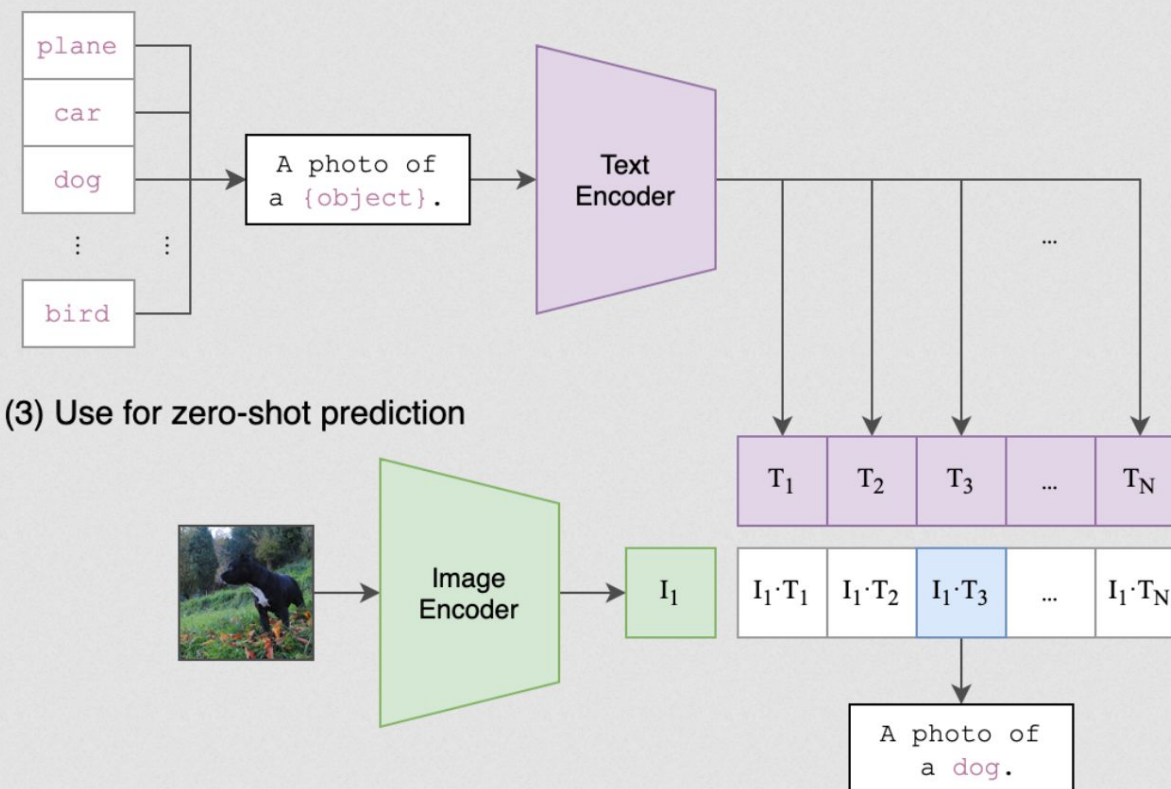
- Кластеризация с помощью ResNet
- Классификация на готовом датасете с GitHub












Решение

- Решаем задачу zero-shot классификации
- Используем CLIP модель из HF
- Для каждого класса подобрали prompt
- Назначаем класс картинке в соответствии с самым близким промптом
- Если модель не уверена, назначаем 16-ый класс

(2) Create dataset classifier from label text



Примеры промптов

Классы		
1. Exterior (Внешний вид)		
		
2. Room (Номер)		
		
3. Pool (Бассейн)		
		

→ Hotel exterior, outdoor area, or building facade

→ Hotel room, living space, or bedroom with furniture

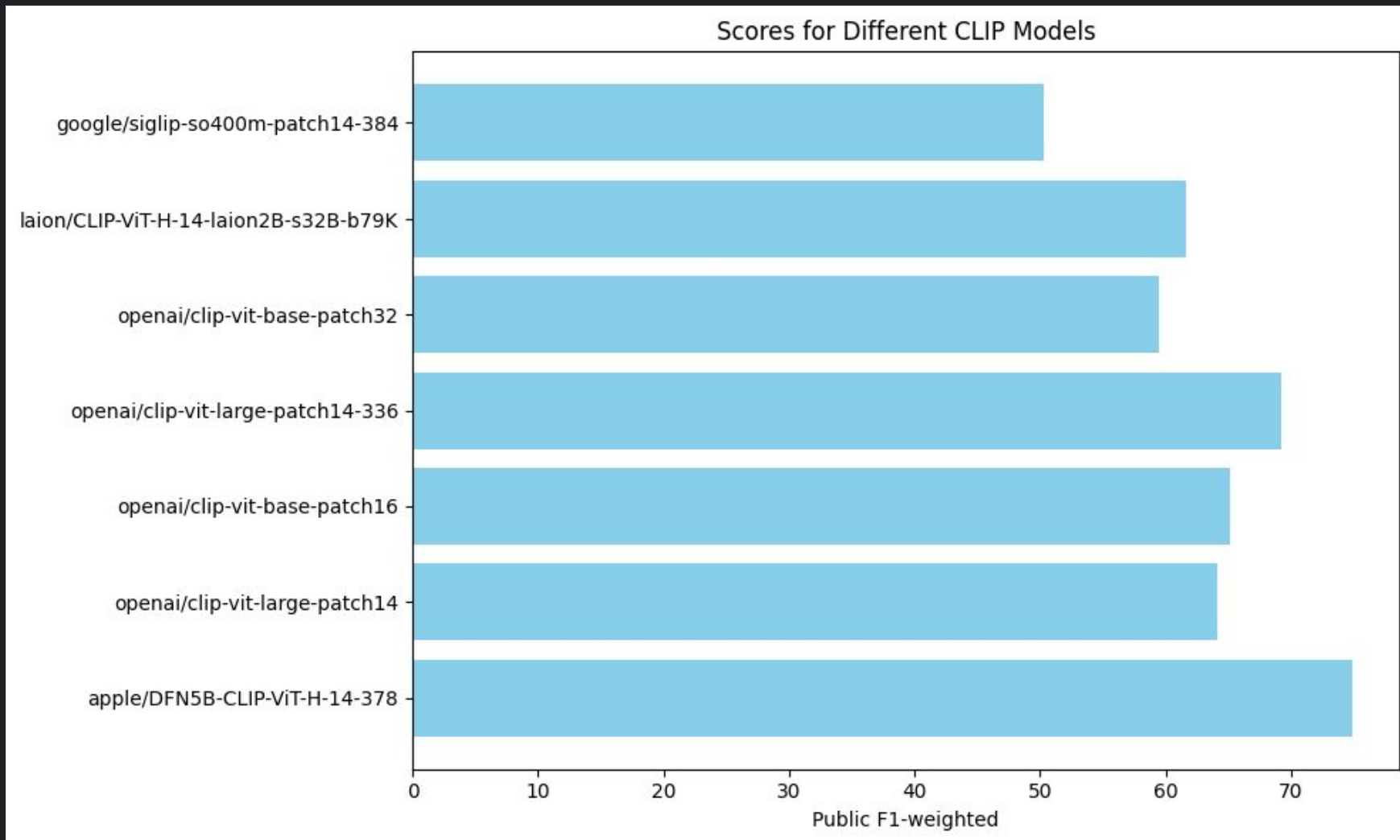
→ Swimming pool or hotel pool area

Сложности, которые преодолели

- Перебирали различные prompt'ы
- Добавляли дополнительные классы
- Пробовали различные модели
- Лучший результат
 - model: **apple/DFN5B-CLIP-ViT-H-14-378**
 - public: 74.88
 - private: 71.53



Метрики различных моделей



Выводы

- CLIP - лучший по соотношению качество/время
- Быстрый инференс без обучения
- Можно изменять промпты
- Можно использовать ансамбли CLIP





Спасибо за внимание!

