

Помоги роботу навести порядок

15 October 2023

Презентация решения для кейса
“Помоги роботу навести порядок” в
хакатоне “Цифровой прорыв 2023”

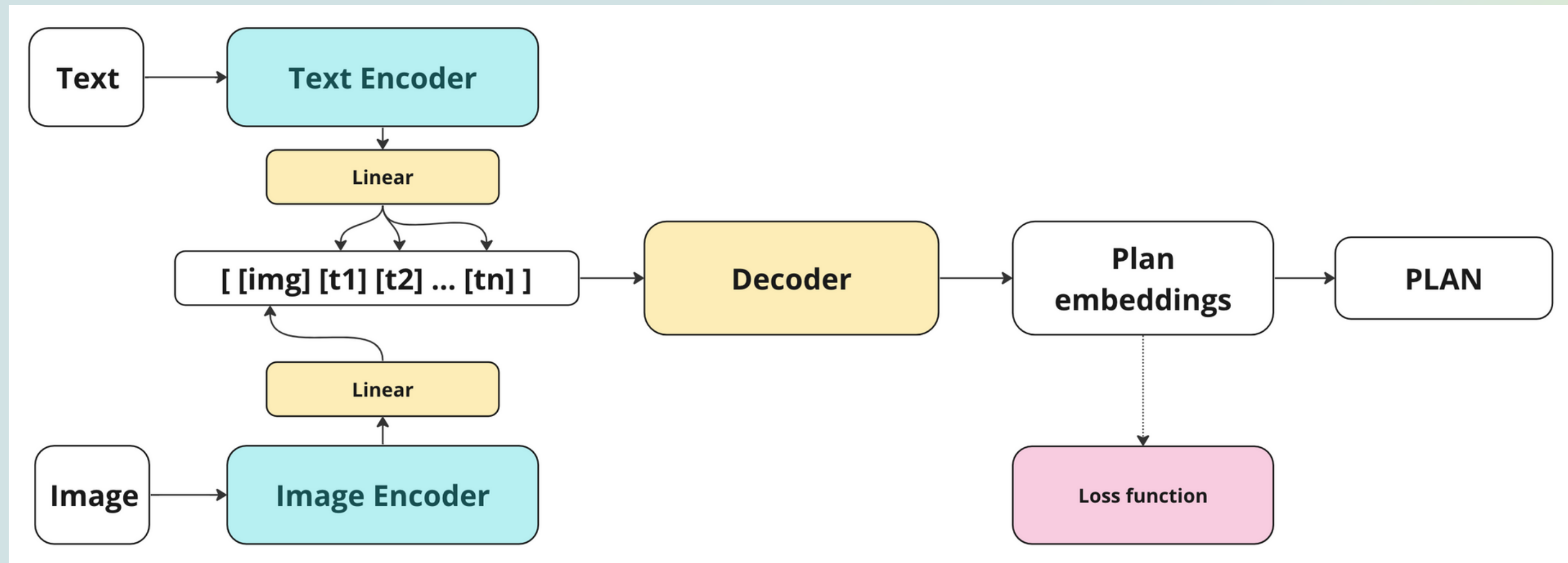


Команда “Бойцовский клуб”

Решаемая проблема

Существующие модели недостаточно хорошо решают задачу преобразования неоднозначного естественного (текст + картинка) языка в строгую последовательность интерпретируемых команд.

Архитектура решения



Здесь:

- Синий - замороженный предобученный на общей задаче элемент
- Желтый - обучаемый элемент
- Красный - функция потерь

Модули и инструменты

Генератор текстовых эмбеддингов

Bert-подобная архитектура предобученной на общей задаче LLM модели

Компромисс между временем работы и качеством генерации

Использовано:

- bert-base-uncased

Генератор эмбеддингов картинок

Использует модель, предобученную на наборе данных из картинок с подписями.

Использовано:

- openai/clip-vit-large-patch14

Декодер

Трансформер со сложной структурой

Принимает в себя все эмбеддинги
Выдает как результат последовательность токенов из разрых словарей

На каждый токен можно использовать свою функцию потерь

Использовано:

- transformerdecoderlayer
- softmax

Основные особенности решения

- Решение состоит из независимых модулей
- Используются преднатренированные модели, хорошо решающие свои общие задачи
- Хорошее качество модели за счет построения архитектуры под задачу

Сравнение с аналогами

Baseline (llama)

Обычная llama
Не учитывает фото

Результат на примере: 0.1

miniGPT-4

Простая LLM модель, принимающая
фото и текст

Результат на примере: 0.3

llama-2-7B-LoRA-assemble

Дообученная llama-2
Не учитывает фото

Результат на примере: 0.7

Результат на тестовых данных: 0.1



Пример - это фиксированные 10 элементов набора данных,
на которых мы быстро проверяли теории о работе моделей

Выводы

Сложности с решением:

- Технически сложно реализовать, мало времени на тестирование
- Очень маленький набора данных даже с аугментацией (около 1000 эл.)

Масштабирование:

- Абстракция на более общую задачу
- Дообучить генерацию эмбеддингов под данную задачу
- Увеличить обучающую выборку
- Добавить больше параметров в модель
- Сделать на основе этой модели RL модель

Скринкаст

Команда

Маслов Михаил

ML engineer at Yandex



Попов Александр

Software engineer at VK



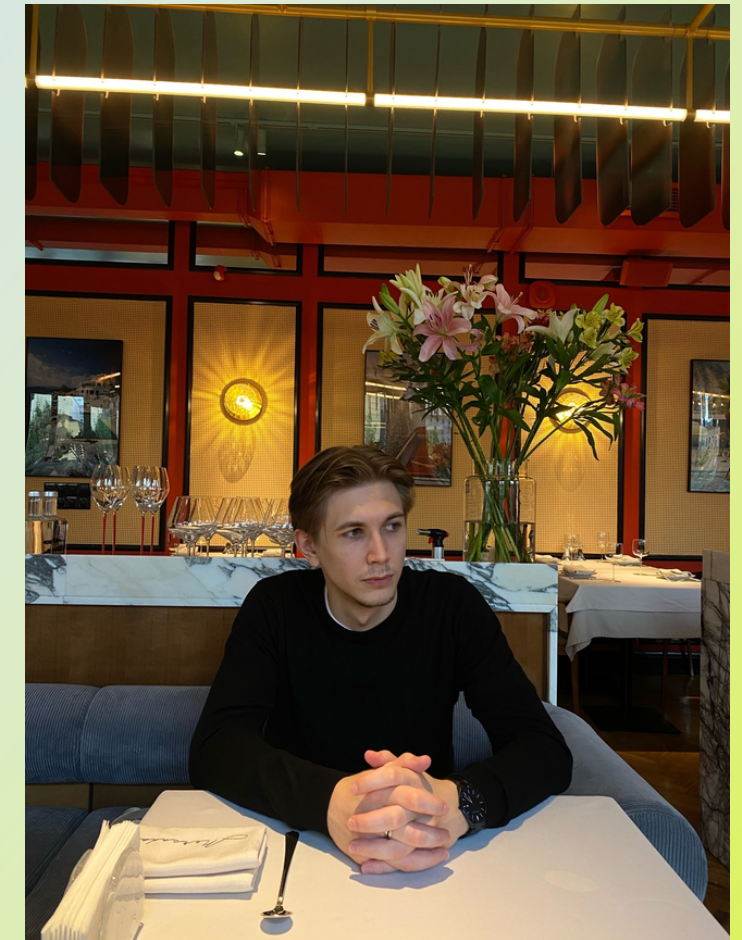
Никита Мишанин

Software engineer at VK



Кирилл Танковский

Software engineer at Yandex



**СПАСИБО ЗА
ВНИМАНИЕ**