

COMP
TECH

Зимняя школа
COMPTech



IMGFINDER

Image Finder



Кураторы:
Амир Утеуов, Владимир Киязов

Команда:
Александр Минин, Анна Глушкова, Влад Кузнецов, Дмитрий Москалев,
Кирилл Келлер, Максим Маштаков, Василий Дронов



Yandex Cloud



План презентации

- Идея
- gitflow
- Продуктовое исследование
- ML модель
- ML эксперименты
- Архитектура
- Демо
- Выводы

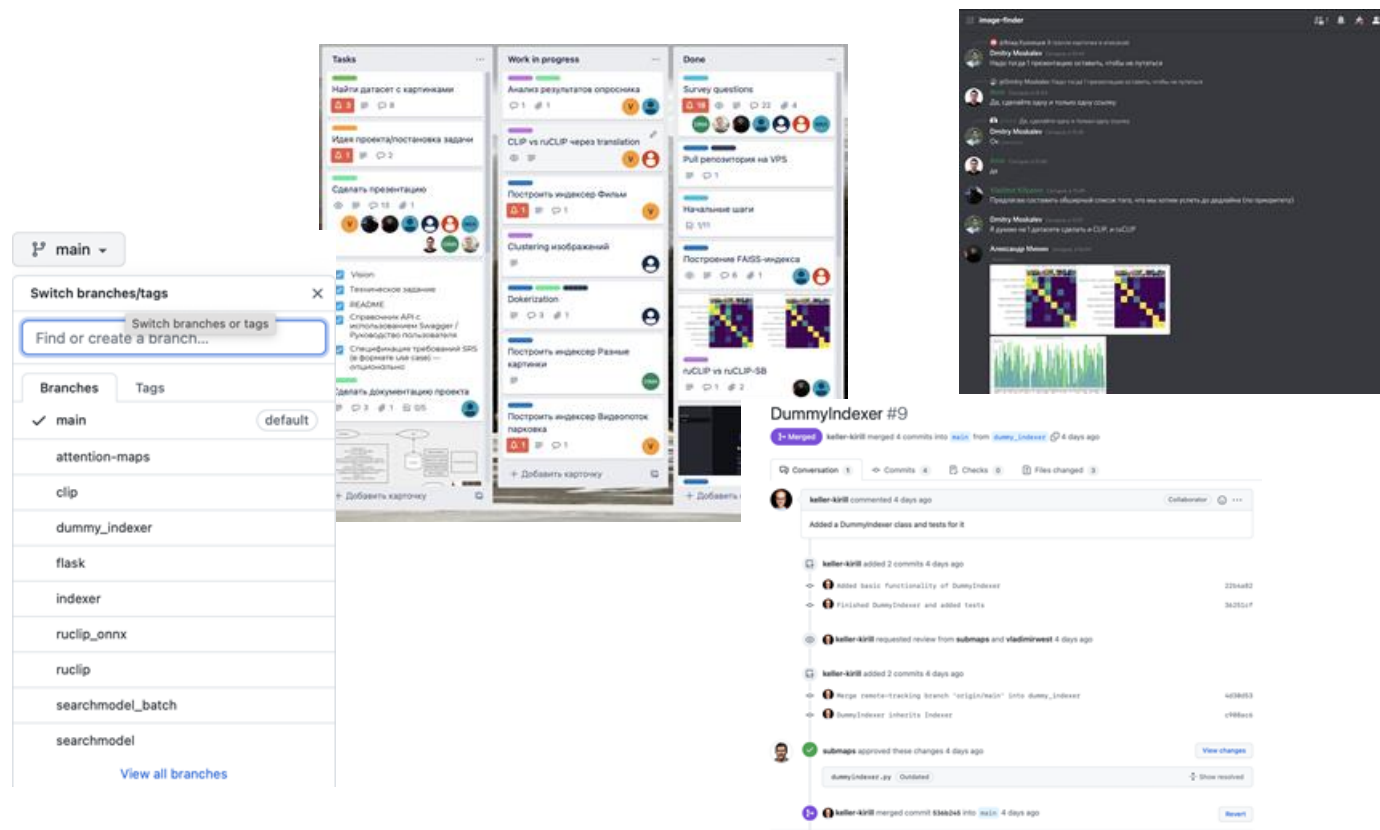
Идея

Разработать веб-сервис для поиска изображений из пользовательских коллекций по произвольному текстовому запросу или графическому изображению.

Решаемые задачи:

- Провести опрос, чтобы лучше понимать кто наш пользователь.
- Научиться работать с CLIP моделями и сравнить качество и производительность различных реализаций.
- Имплементировать поисковой индекс для хранения и поиска эмбеддингов изображений.
- Разработать MVP.

Платформы, которые мы использовали в процессе работы над проектом

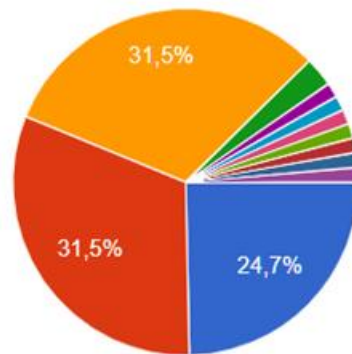


Опрос

В рамках проекта был проведен релевантный опрос, в котором приняло участие 73 человека.

Где храните фотографии? Выберите один вариант, в котором хранится большая часть.

73 ответа



- В облаке (Яндекс, google photos, ap...
- Локально в телефоне
- Локально на ПК/ноутбуке
- В социальных сетях (Instagram, VK, T...
- в телефоне/ПК/облако
- Все вышеперечисленное подходит
- Локально на ПК + в соцсетях
- Что значит в основном. Отобранные...

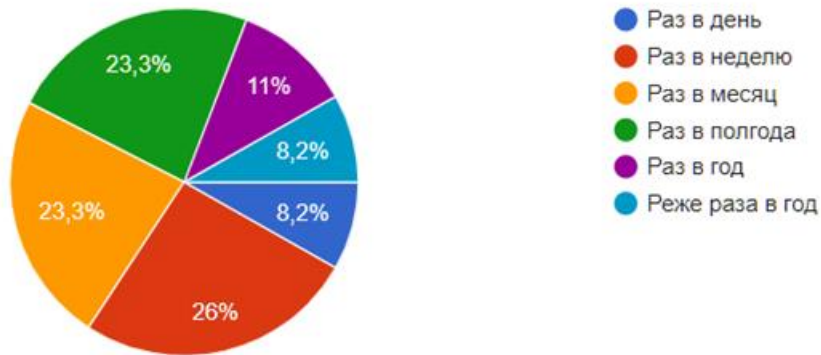
▲ 1/2 ▼

Вывод: большинство участников опроса хранят фотографии на ПК / телефоне

Опрос

Как часто вам нужно разбирать цифровые фотоархивы, искать какие-либо фотографии?

73 ответа

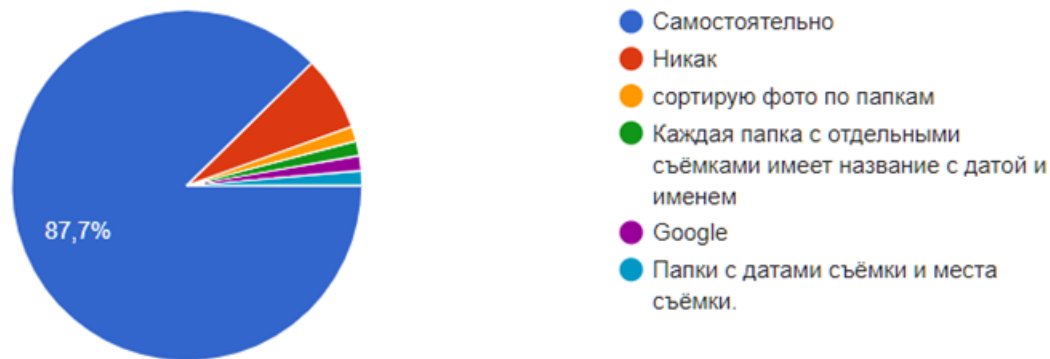


Вывод: пользователи довольно часто разбирают фотографии

Опрос

Как справляетесь с задачей нахождения нужных фотографий из определенного хранилища в настоящее время? Если используете сторонние сервисы, то напишите их названия в “Другое”.

73 ответа



Вывод: большинство пользователей самостоятельно ищут фотографии

Опрос

Вопрос: “Что в задаче нахождения нужных фотографий вам нравится и не нравится?”



Опрос

Вопрос: “Какие ещё инструменты вы пробовали применять, пока не остановились на том, что используете сейчас?”



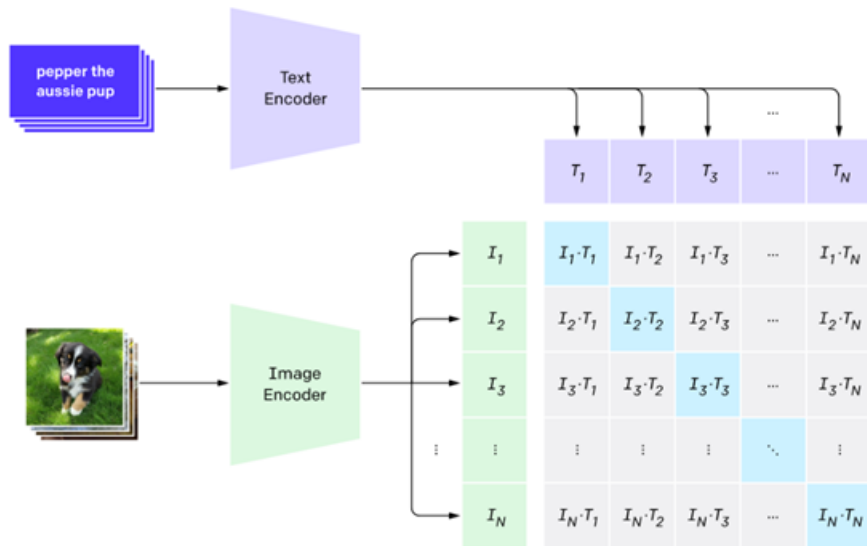
Проблемы

Некоторые результаты:

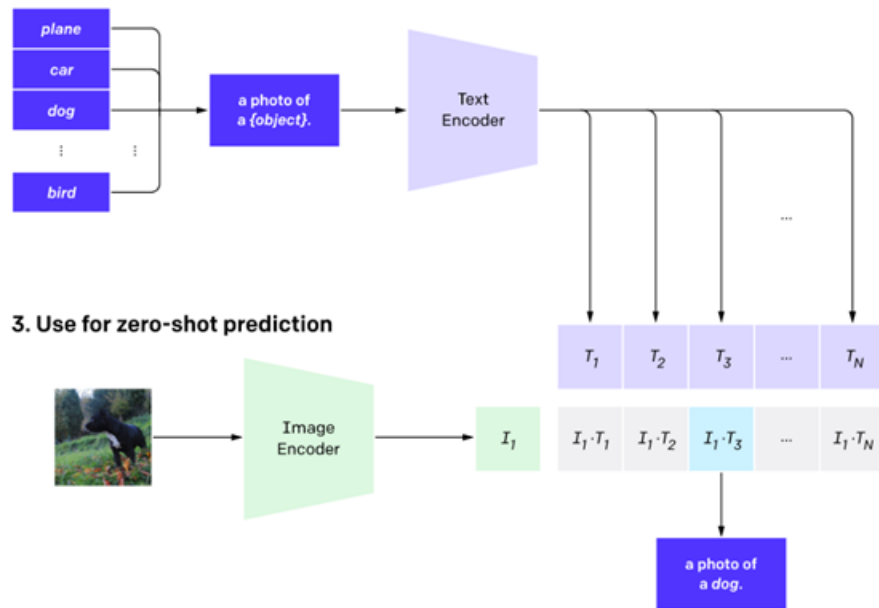
- Слишком большой объем фотографий, поэтому сложно иногда найти то, что нужно
- Не нравится то, что поиск изображений это обычно ручная работа, время выполнения которой напрямую зависит от исходного количества фотографий
- Нет “идеальных” бесплатных или готовых решений, которые пользователи знают

CLIP - Contrastive Language-Image Pre-training

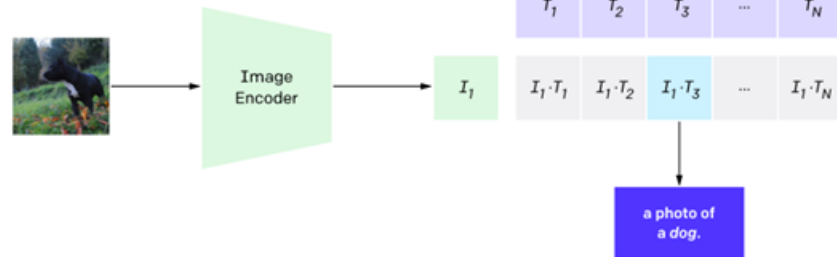
1. Contrastive pre-training



2. Create dataset classifier from label text

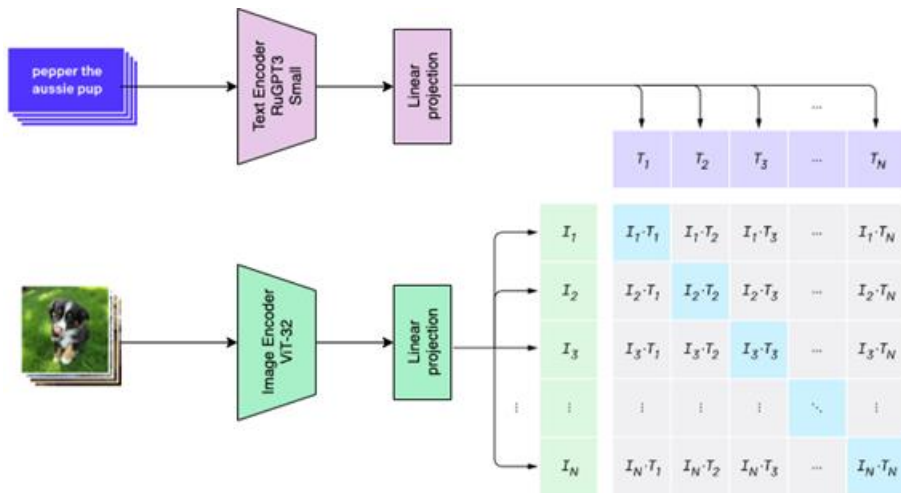


3. Use for zero-shot prediction



<https://openai.com/blog/clip/>

RuCLIP



Cosine similarity between text and image features RuCLIP

радостные студенты с закрытой сессией

друзья в парке на траве

свадьба в шатре

семейное фото из путешествий

тетрадь с формулами по матану

дизайн интерьера зала

кот астронавт в космосе

мем про продакшн



радостные студенты с закрытой сессией	0.42	0.23	0.10	0.19	0.07	0.06	0.02	0.16
друзья в парке на траве	0.28	0.38	0.12	0.24	0.01	0.02	0.02	0.09
свадьба в шатре	0.17	0.09	0.36	0.19	0.07	0.18	0.04	0.08
семейное фото из путешествий	0.10	0.17	0.08	0.29	0.05	0.05	0.05	0.10
тетрадь с формулами по матану	0.22	0.04	0.09	0.11	0.33	0.06	0.14	0.08
дизайн интерьера зала	0.15	0.05	0.08	0.11	0.06	0.22	0.11	0.06
кот астронавт в космосе	0.18	0.06	0.06	0.15	0.07	0.06	0.48	0.13
мем про продакшн	0.09	0.08	0.04	0.08	0.08	0.07	0.23	0.30

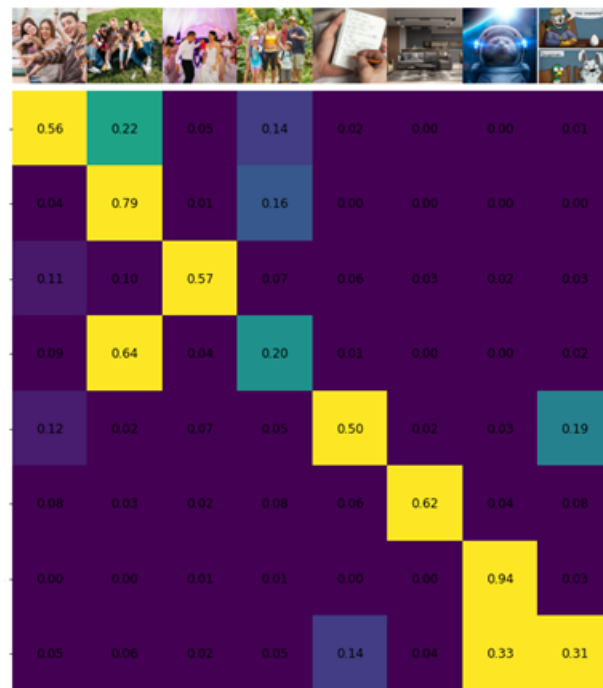
<https://habr.com/ru/company/sberdevices/blog/564440/>

RuCLIP vs RuCLIP-SB

Cosine similarity between text and image features RuCLIP

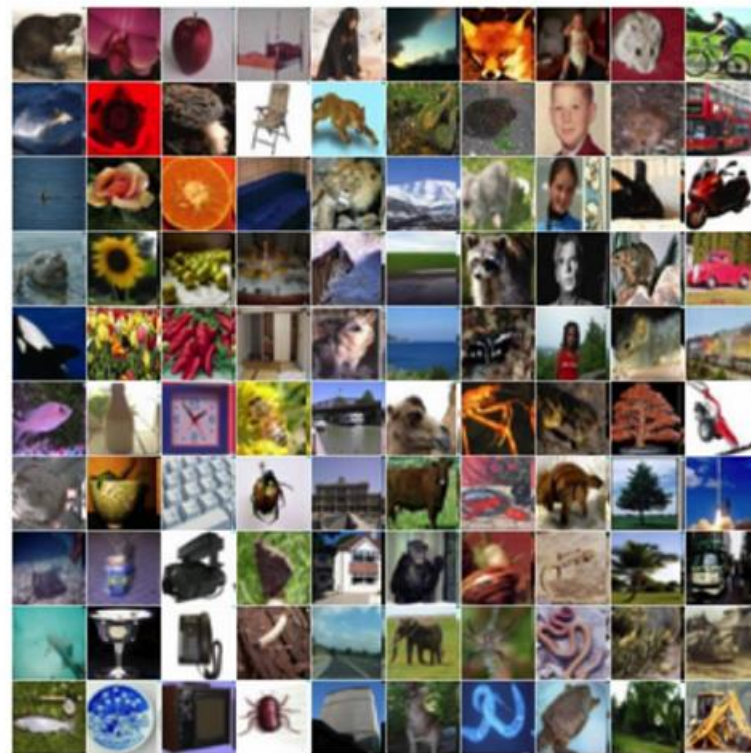


Cosine similarity between text and image features RuCLIP-SB

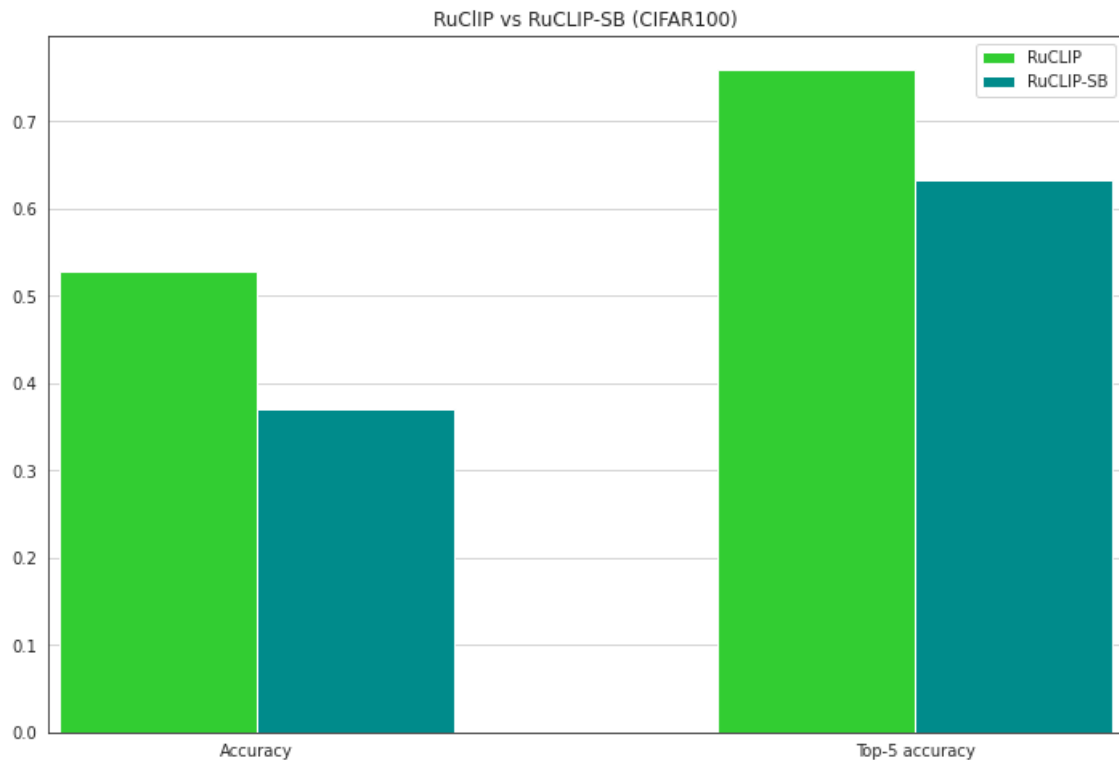


CIFAR-100

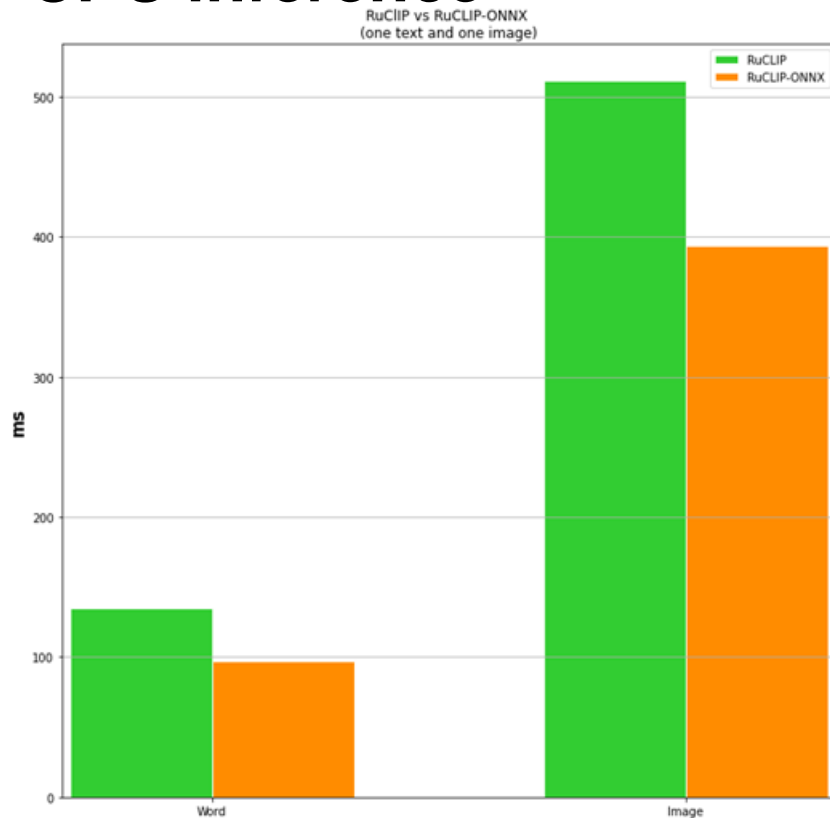
<https://www.cs.toronto.edu/~kriz/cifar.html>



Accuracy & Top-5 accuracy



RuCLIP CPU inference



Прирост скорости около 20% ↑



ONNX – это высокопроизводительный механизм вывода для развертывания моделей ONNX в рабочей среде.

Attention maps

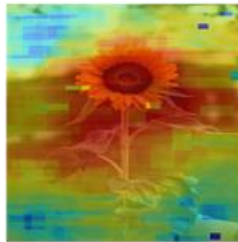
Подсолнух



CLIP



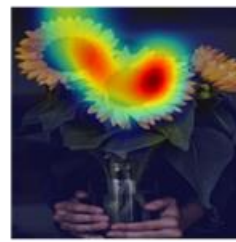
ruCLIP



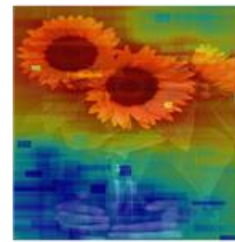
Подсолнух



CLIP



ruCLIP



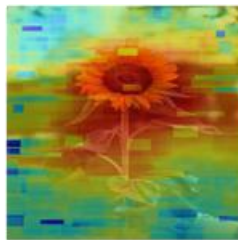
Подсолнухи



CLIP



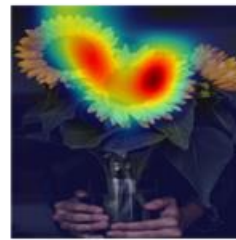
ruCLIP



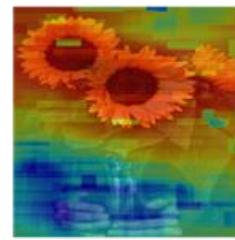
Подсолнухи



CLIP



ruCLIP



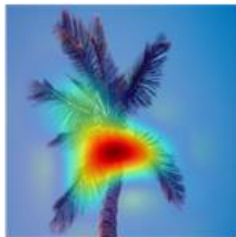
Вывод: результаты обработки для запросов в единственном и множественном числах почти не отличаются

Attention maps

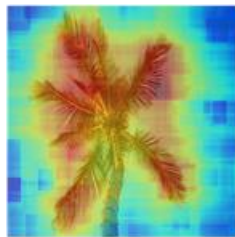
Пальма



CLIP



ruCLIP



Пальма



CLIP



ruCLIP



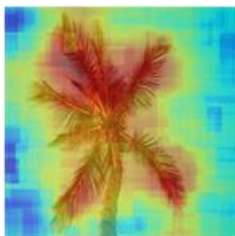
Пальмы



CLIP



ruCLIP



Пальмы



CLIP



ruCLIP



Вывод: результаты обработки для запросов в единственном и множественном числах почти не отличаются

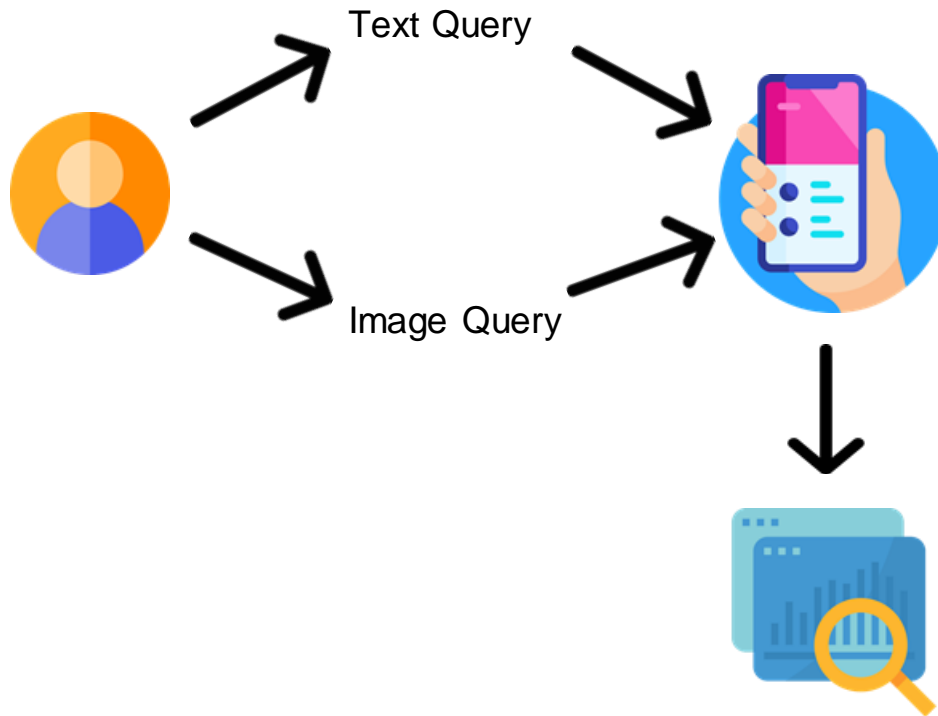
Архитектура проекта

System operation:

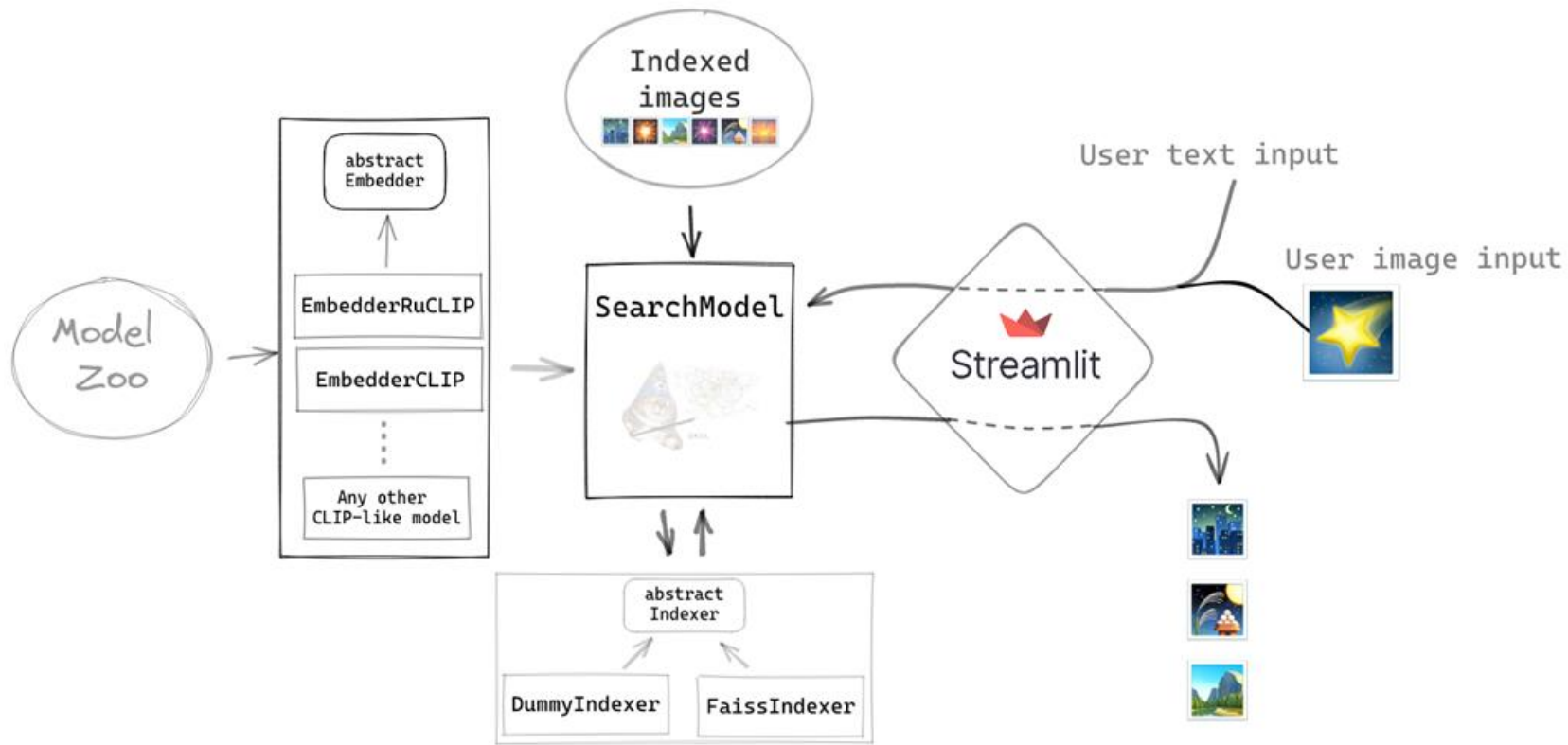
- Image search;
- Search by text query (Russian/English);
- Video search.

Structure:

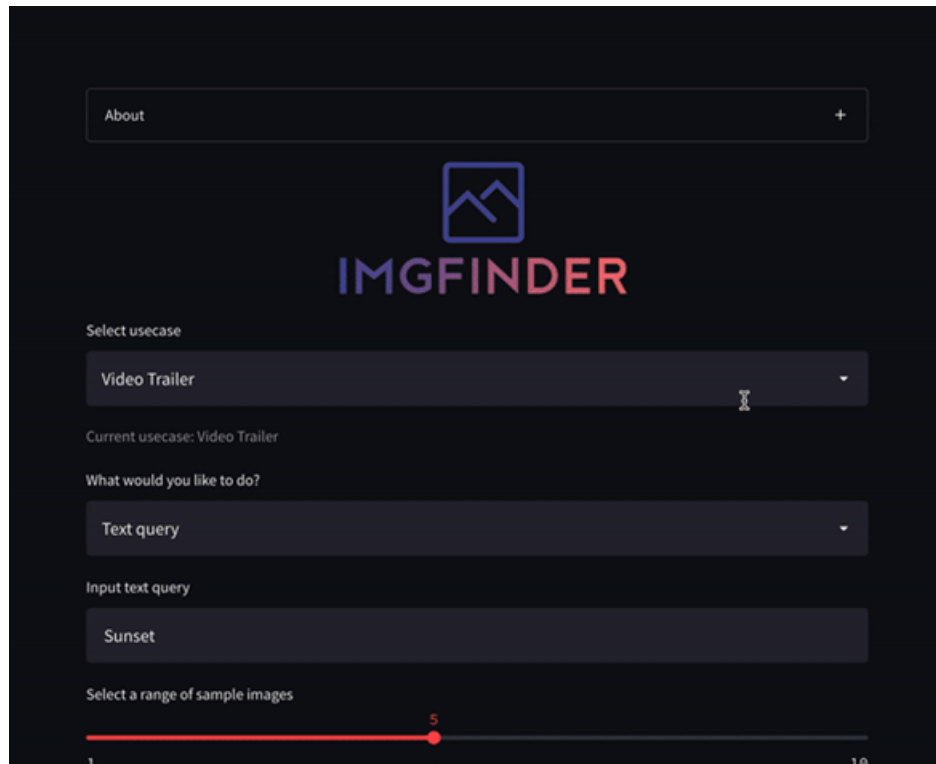
- colabs
- test
- assets
- main.py
- faissindexer.py
- dummyindexer.py
- embedder.py
- searchmodel.py
- CLIP_attention_maps.py
- ruCLIP_attention_maps.py
- requirements.txt



Архитектура приложения



Демо



You can watch presentation on YouTube:

<https://youtu.be/wJorcAXURz8?t=8203>

Работает на CPU (VPS)

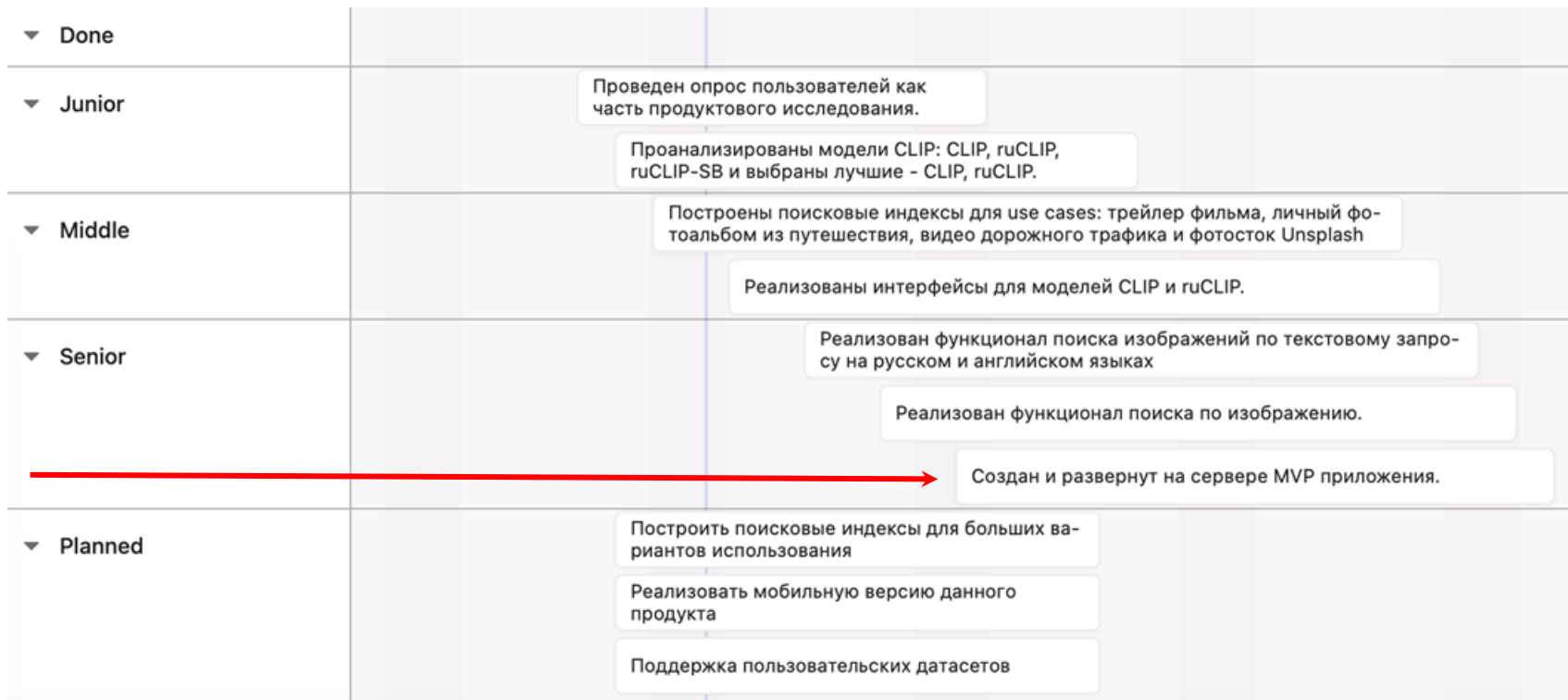
Use cases:

1. Личный фотоальбом из путешествия
2. Трейлер фильма
3. Фотосток Unsplash
4. Видео дорожного трафика

Заключение

- Мы разработали ML сервис от идеи до MVP и задеплоили его
- Провели опрос пользователей как часть продуктового исследования.
- Проанализировали модели CLIP, ruCLIP, ruCLIP-SB и выбрали лучшие - CLIP, ruCLIP.
- Построили поисковые индексы для use cases: трейлер фильма, личный фотоальбом из путешествия, видео дорожного трафика и фотосток Unsplash (~1.8 млн изображений).
- Реализовали функционал поиска изображений по текстовому запросу на русском и английском языках и также по изображению
- Создали и развернули на сервере MVP приложения.

Roadmap



Команда

Mentors:

- Amir Uteuov
- Vladimir Kilyazov

Team Lead:

- Dmitry Moskalev

Разработчики:

- Aleksandr Minin
- Anna Glushkova
- Vladislav Kuznetsov
- Dmitry Moskalev
- Kirill Keller
- Maxim Mashtakov
- Vasiliy Dronov

Ссылка на репозиторий:

<https://github.com/comptech-winter-school/image-finder>



COMP
TECH

Зимняя школа
COMPTech



Спасибо за внимание



Datasets

- Cifar100; URL: <https://www.cs.toronto.edu/%7Ekriz/cifar.html>
- Unsplash; URL: <https://unsplash.com/data>

Materials

[Статья об оригинальной модели CLIP для аннотирования изображений](#)

[Статья о реализации ruCLIP от SberDevices](#)

[Применение FAISS для поиска изображений](#)

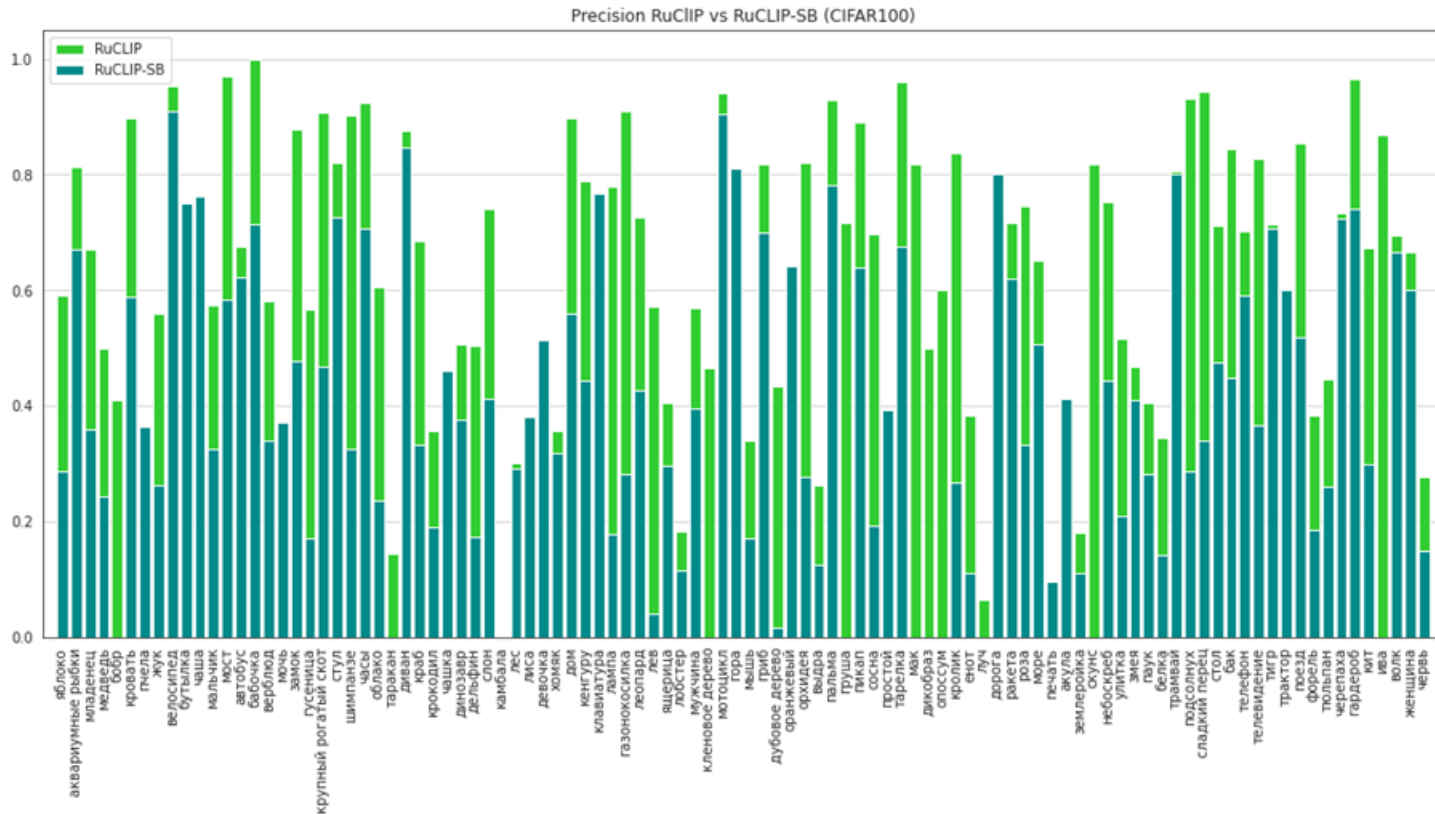
[Репозиторий мультимодальной модели ruDOLPH](#)

[Фреймворк для индексации эмбеддингов](#)

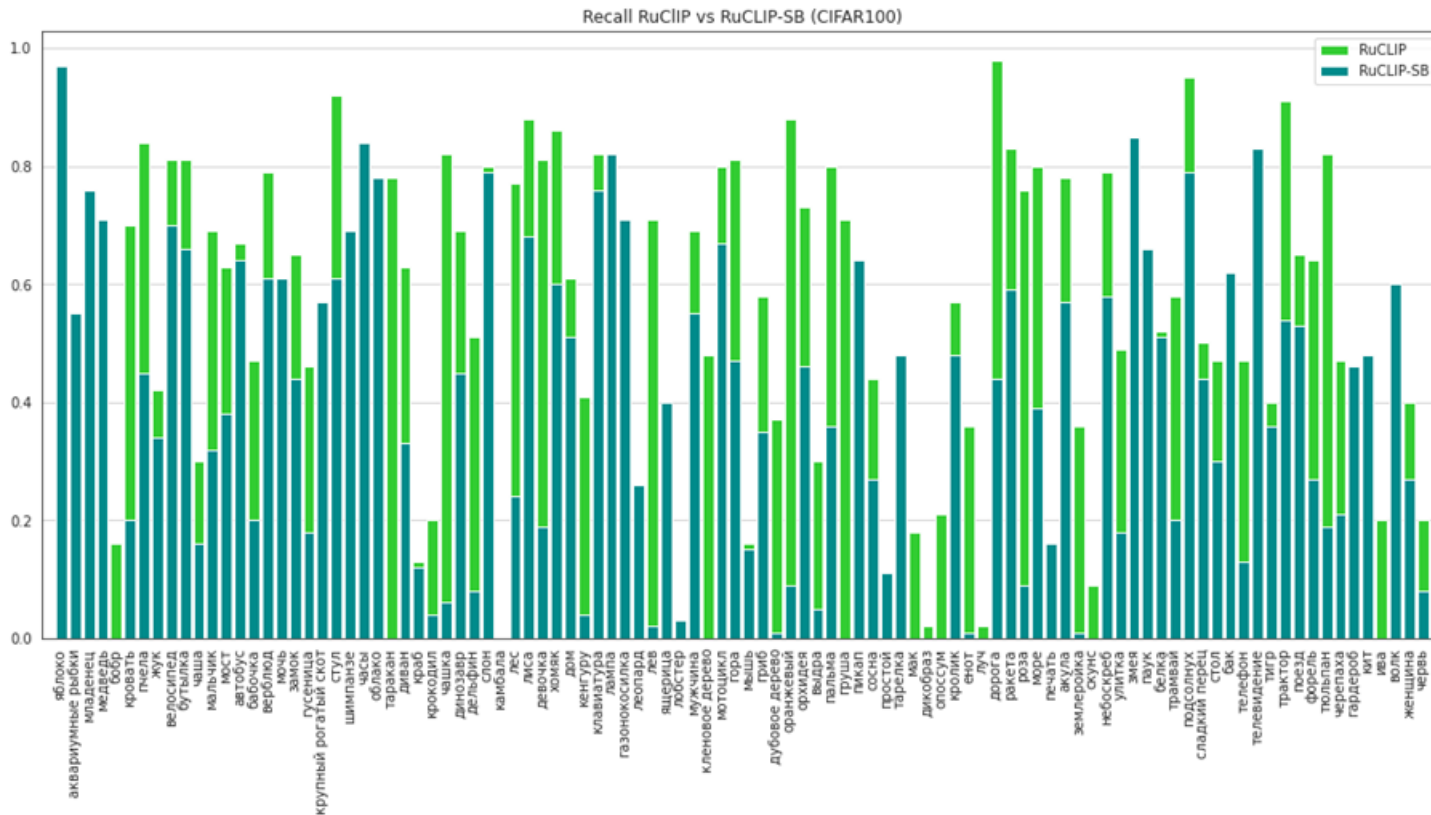
[CV фреймворк Detectron2](#)

[Лекция DeepMind о Geometric Deep Learning](#)

Precision (CIFAR-100)

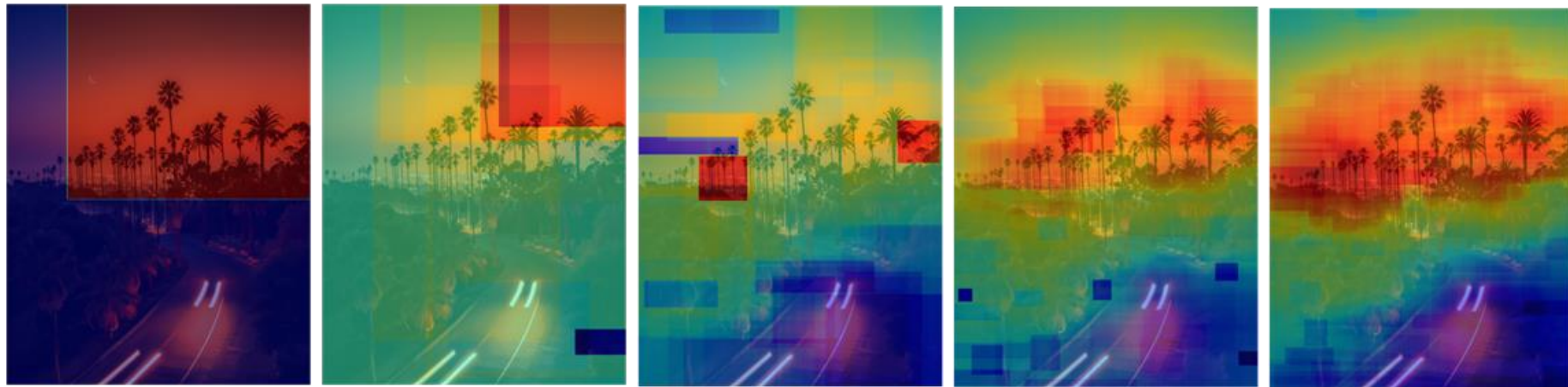


Recall (CIFAR-100)



Attention maps

ruCLIP, training: 1, 10, 100, 1000, 2000 iterations



Вывод: результаты обработки в зависимости от количества итераций