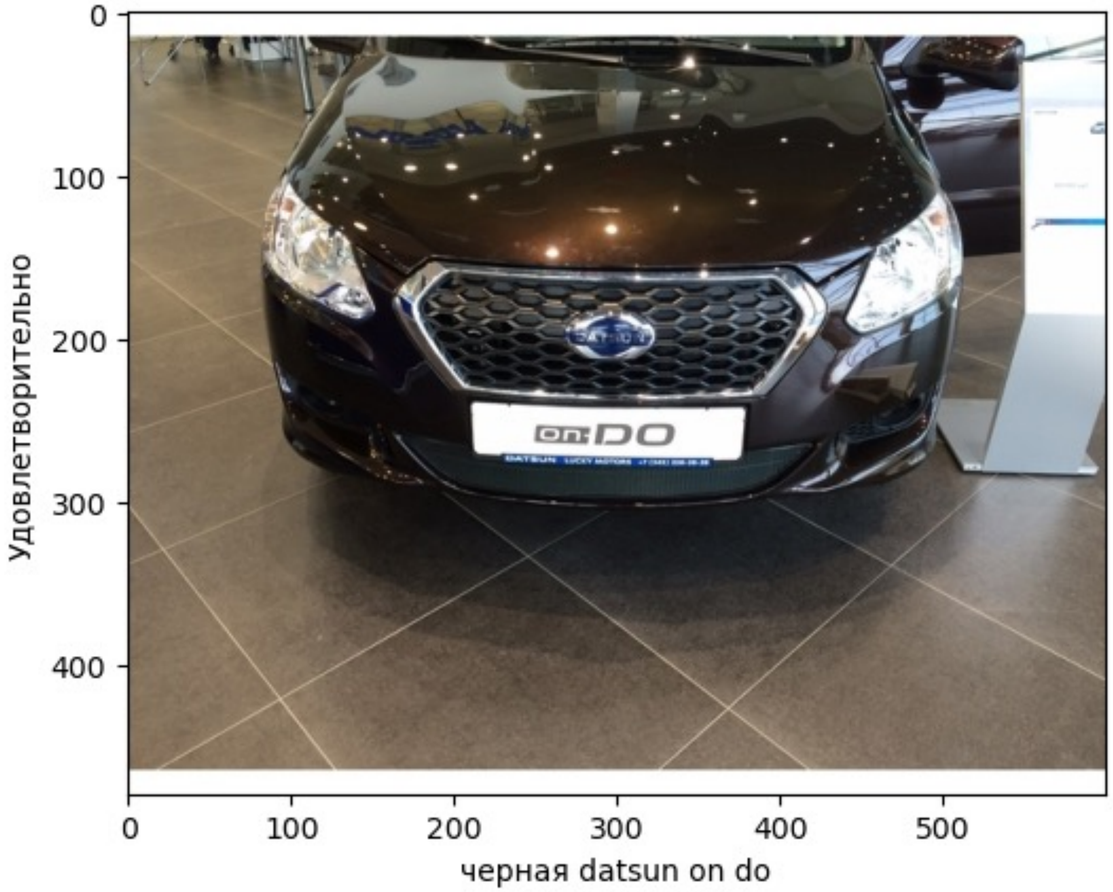
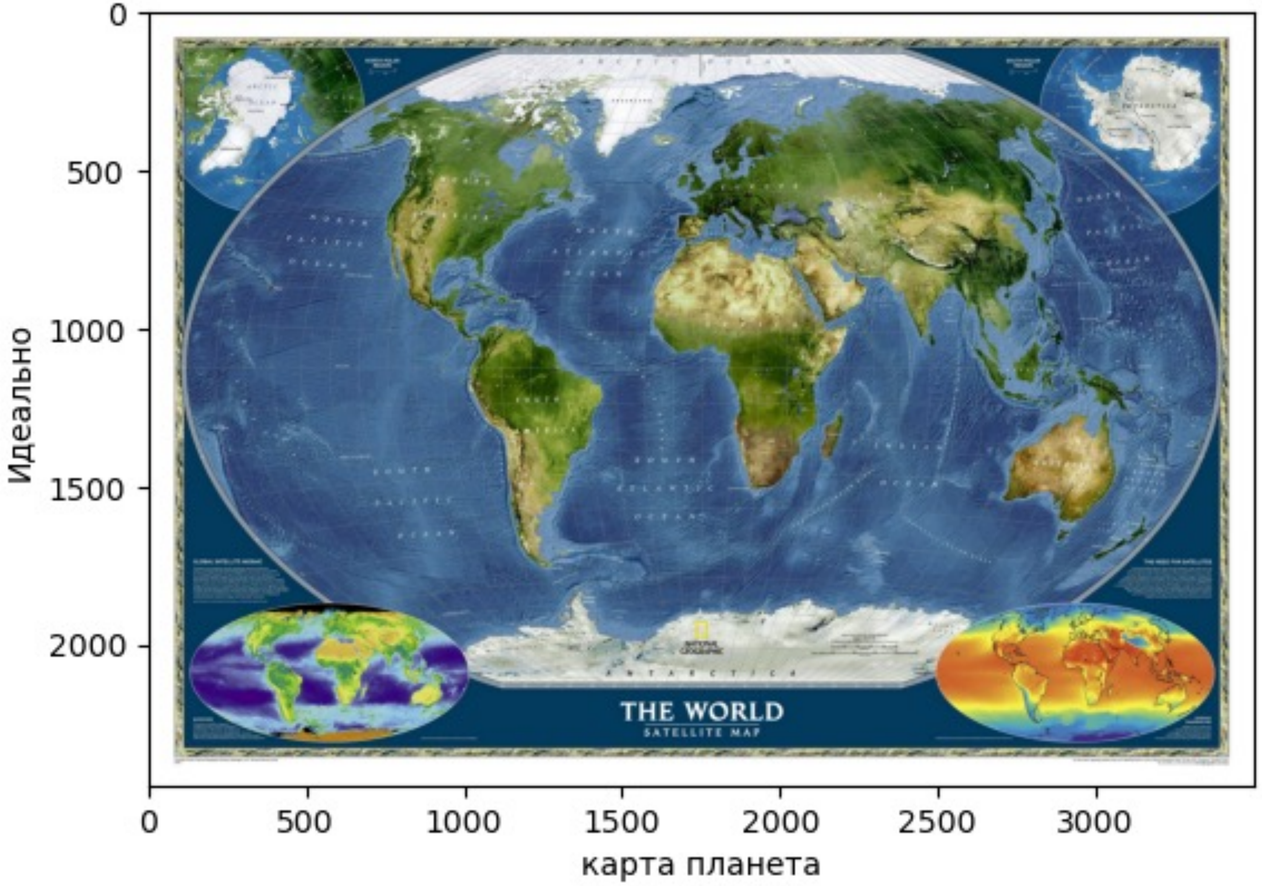
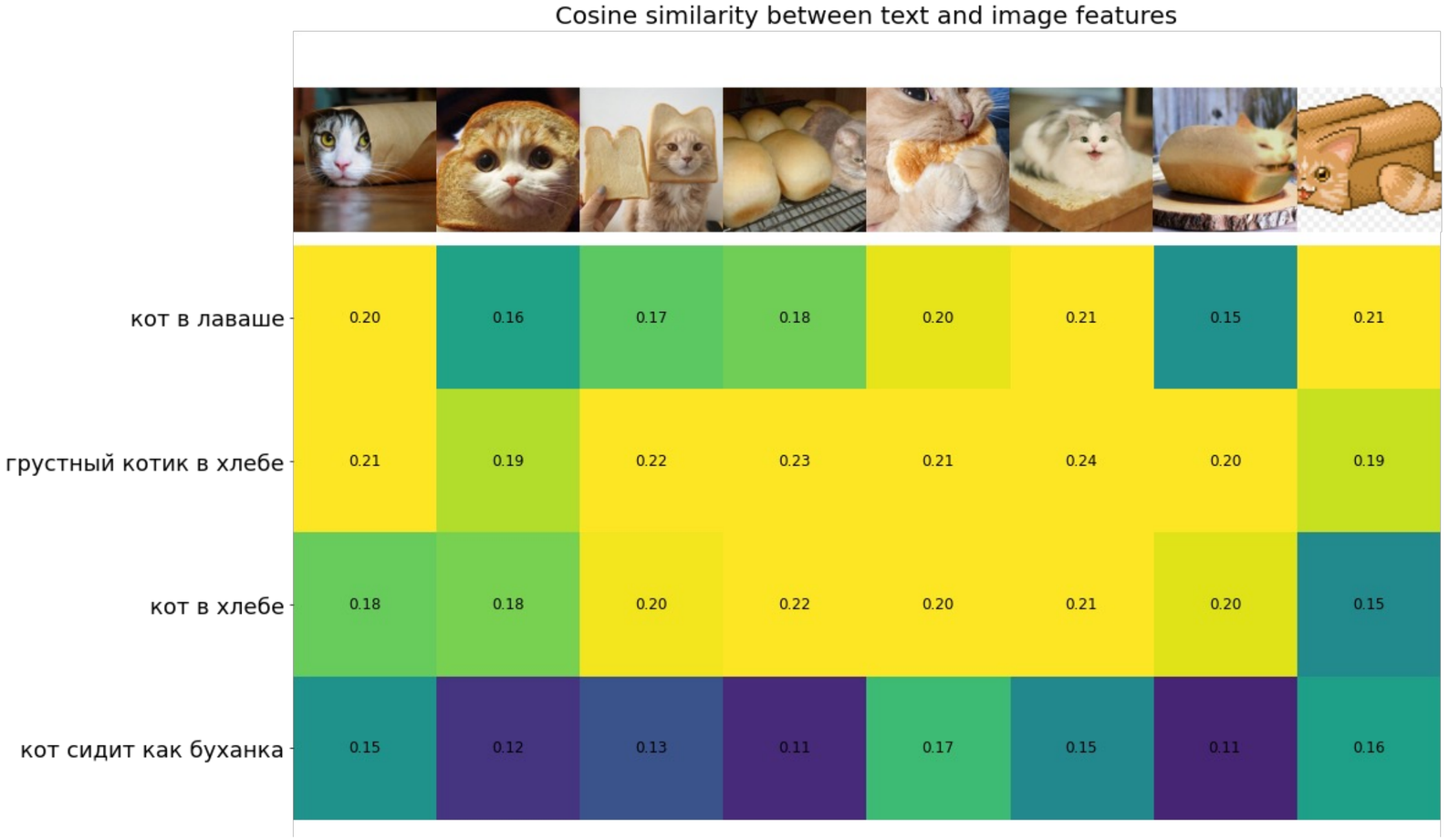
CLIP

Image-text matching





CLIP



https://habr.com/ru/companies/sberdevices/articles/564440/

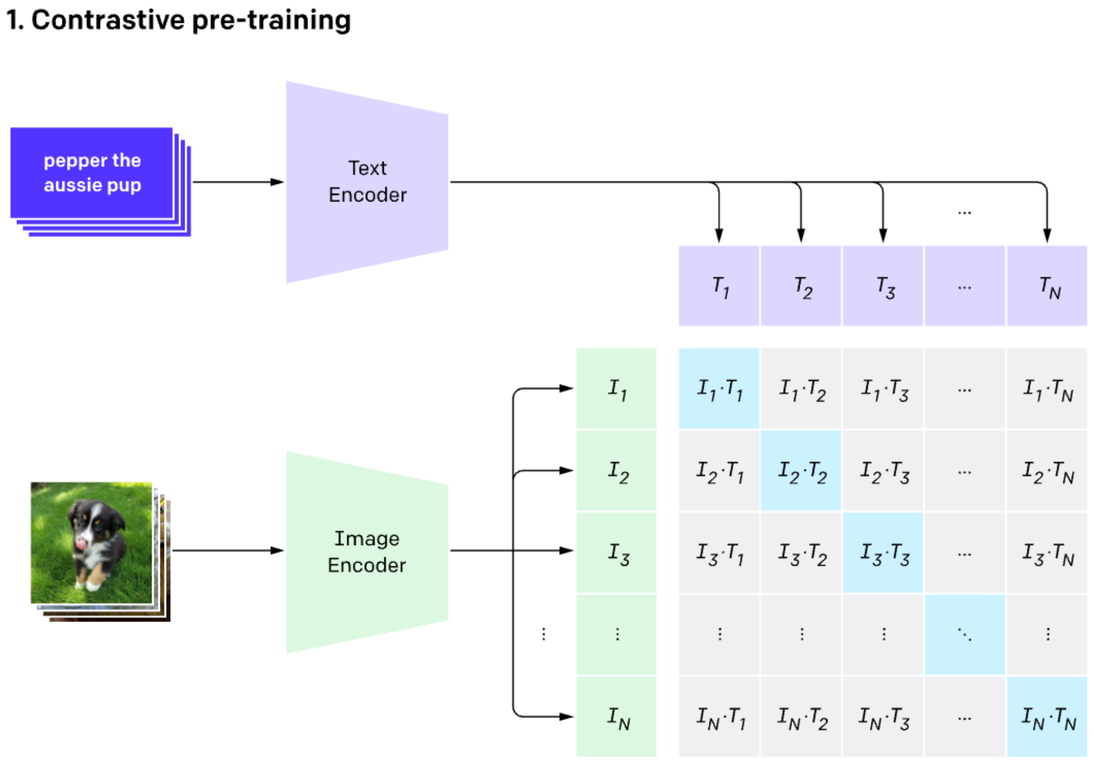
CLIP

CLIP — это модель, состоящая из двух частей (или нейронных сетей):

1. **Image Encoder** — часть для кодирования изображений и перевода их в общее векторное пространство. В качестве архитектуры в оригинальной работе берутся ResNet разных размеров и Visual Transformer — тоже разных размеров.
2. **Text Encoder** — часть для кодирования текстов и перевода их в общее векторное пространство. В качестве архитектуры в оригинальной работе используется небольшой текстовый Transformer.

CLIP

CLIP выучивает мультимодальное пространство путём совместного обучения Image Encoder и Text Encoder, чтобы максимизировать косинусную близость эмбеддингов изображения и текста реальных пар и минимизировать косинусную близость эмбеддингов неправильных пар. Авторы оптимизируют симметричную кросс-энтропийную функцию потерь, над полученными оценками близости.



RUCLIP

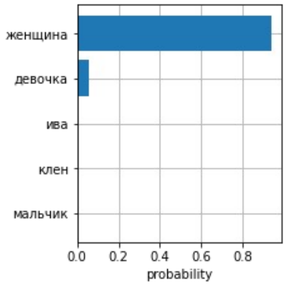
Модель CLIP дообучена на русскоязычных данных - RuCLIP.

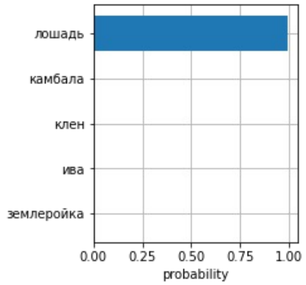
Дообучение модели для русского языка происходило на собранных нами датасетах.

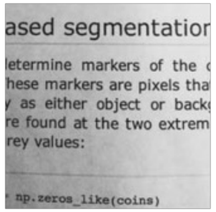
Вот некоторые из них:

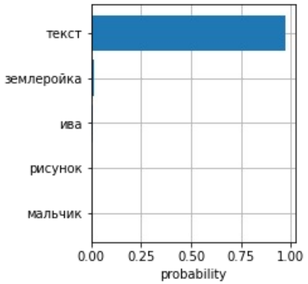
- ImageNet — переведённый на русский язык;
- Flickr — картинки с русскими описаниями с фотостока;
- Ru-wiki — часть картинок из русской Википедии с описаниями.

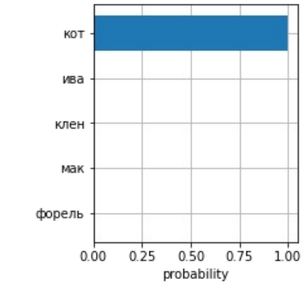
RUCLIP

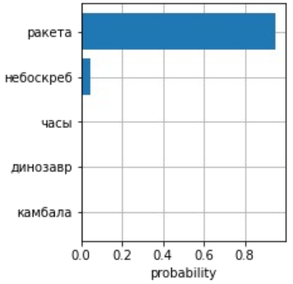

женщина
девочка
ива
клен
мальчик
0.0 0.2 0.4 0.6 0.8
probability

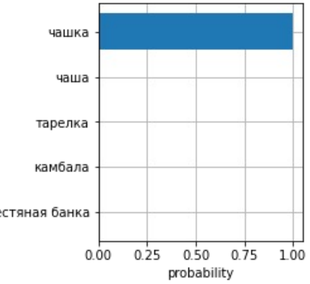

лошадь
камбала
клен
ива
землеройка
0.00 0.25 0.50 0.75 1.00
probability

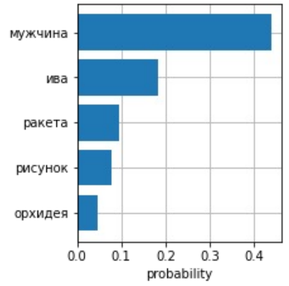

ased segmentatior
letermine markers of the
hese markers are pixels tha
y as either object or back
re found at the two extrem
rey values:
np.zeros_like(coins)

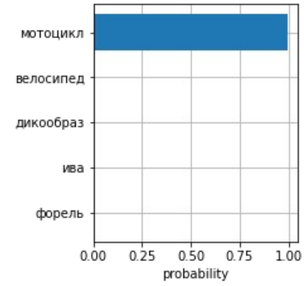

текст
землеройка
ива
рисунок
мальчик
0.00 0.25 0.50 0.75 1.00
probability


кот
ива
клен
мак
форель
0.00 0.25 0.50 0.75 1.00
probability


ракета
небоскреб
часы
динозавр
камбала
0.0 0.2 0.4 0.6 0.8
probability


чашка
чаша
тарелка
камбала
жестяная банка
0.00 0.25 0.50 0.75 1.00
probability


мужчина
ива
ракета
рисунок
орхидея
0.0 0.1 0.2 0.3 0.4
probability


мотоцикл
велосипед
дикообраз
ива
форель
0.00 0.25 0.50 0.75 1.00
probability

CLIP Demo

https://sachinruk.github.io/blog/2021-03-07-clip.html