

Перенос обучения

Урок 2

Эмбеддинги. Использование предобученных моделей для трансформации признаков



О чём будет лекция

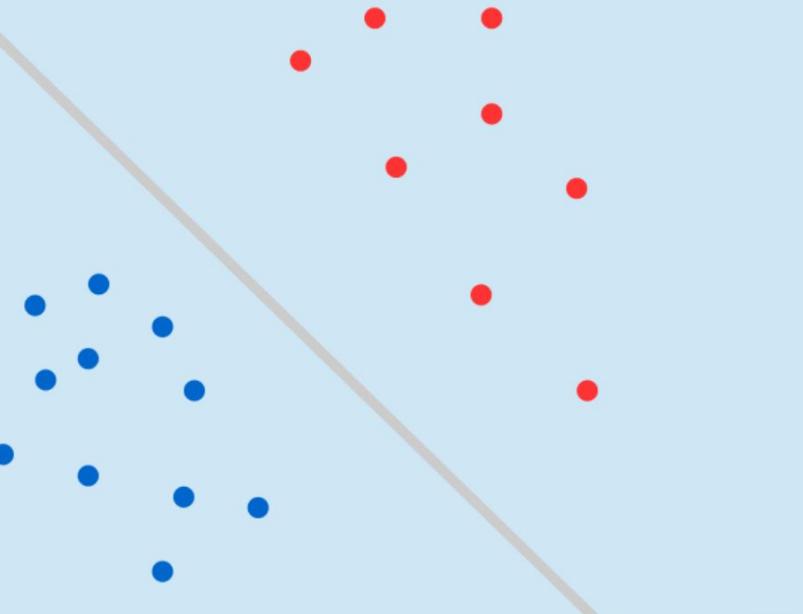
Пример отзыва после удаления пунктуации и приведения символов к нижнему регистру	Оценка тональности с помощью библиотеки <i>dostoevsky</i>
отличный курс большое вам спасибо артем	Отзыв отнесен к категории «речь» (благодарность) с оценкой 0,99683732
по русски читать невозможно ошибка на опечатке но для тех кому химия никак не дается курс будет полезен в задачках непосредственно показывается откуда что брать и куда подставлять	Отзыв отнесен к категории «негативный» с оценкой 0,44553956
лучший курс для подготовки к егэ по математике с отличным учителем ever учеба с тимуром сотовка баллов за экзамен отлично проведенное время	Отзыв отнесен к категории «позитивный» с оценкой 0,65842754
курс хорош и не только для начинающих для тех кто знаком с python не первый год так же есть немало полезного я бы особенно отметил некоторые полуолимпиадные задачки по программированию а так же минипроекты по matplotlib seaborn и plotly которые хоть и не простые и многие не любят их делать но приносят реальную пользу	Отзыв отнесен к категории «нейтральный» с оценкой 0,62978464

miro

Пример: несколько строк из выборки отзывов о русскоязычных онлайн-курсах



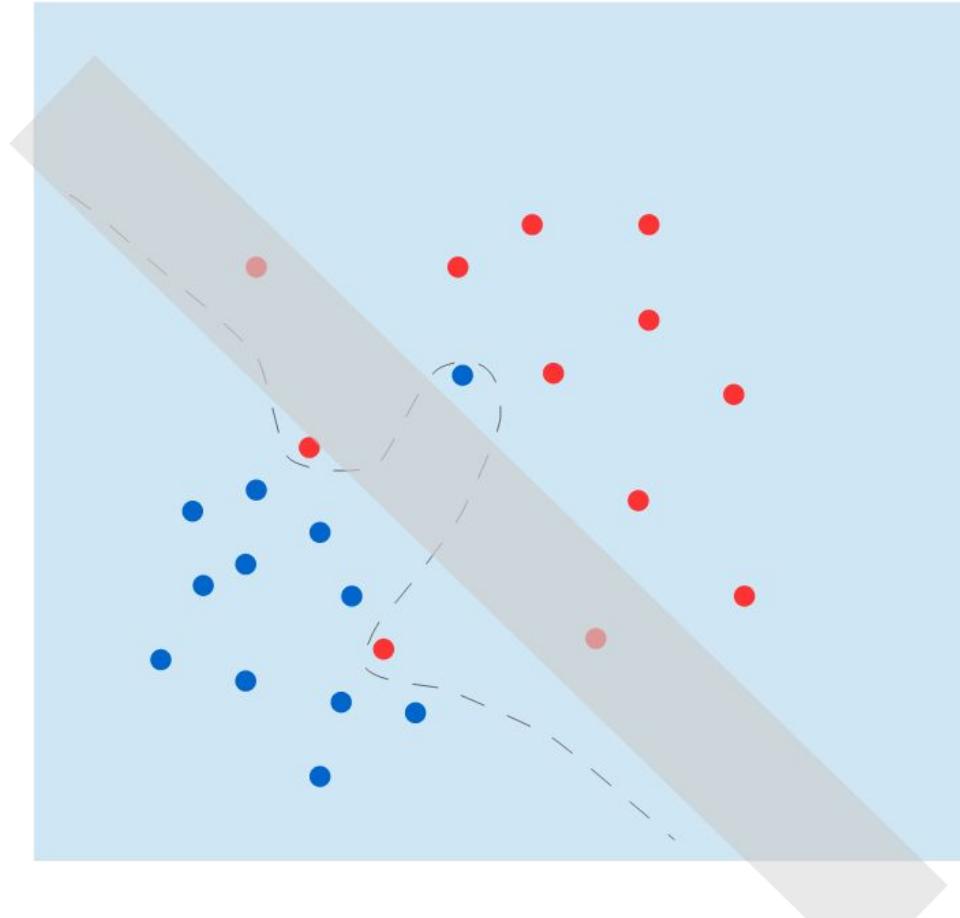
Преобразование признаков



- Красные точки – историческая застройка
- Синие точки – новые дома
- Нужно провести между ними дорогу



Преобразование признаков

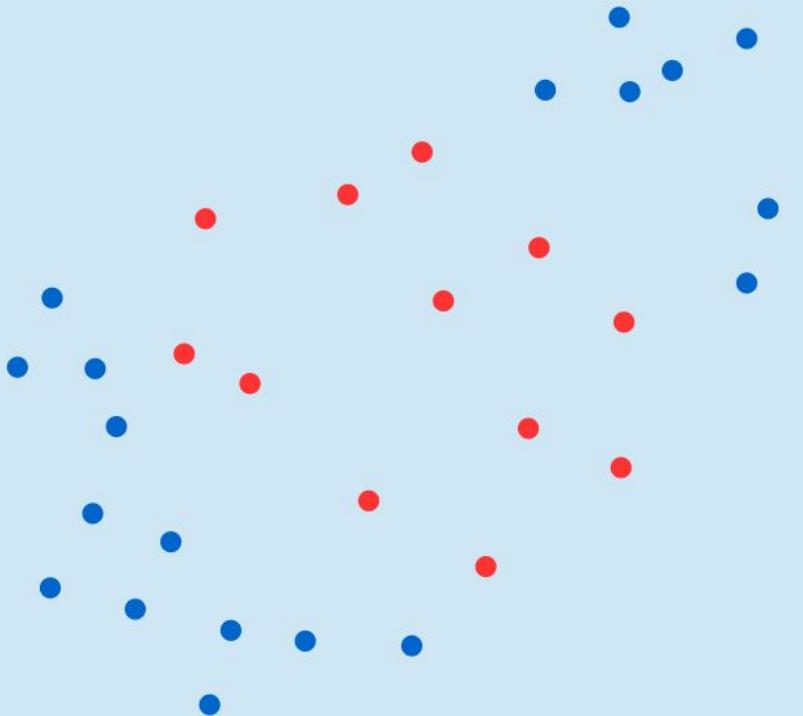


- Задача линейно неразделимая
- Но мы очень хотим чтобы дорога была прямой
- Придется согласиться с некоторой погрешностью



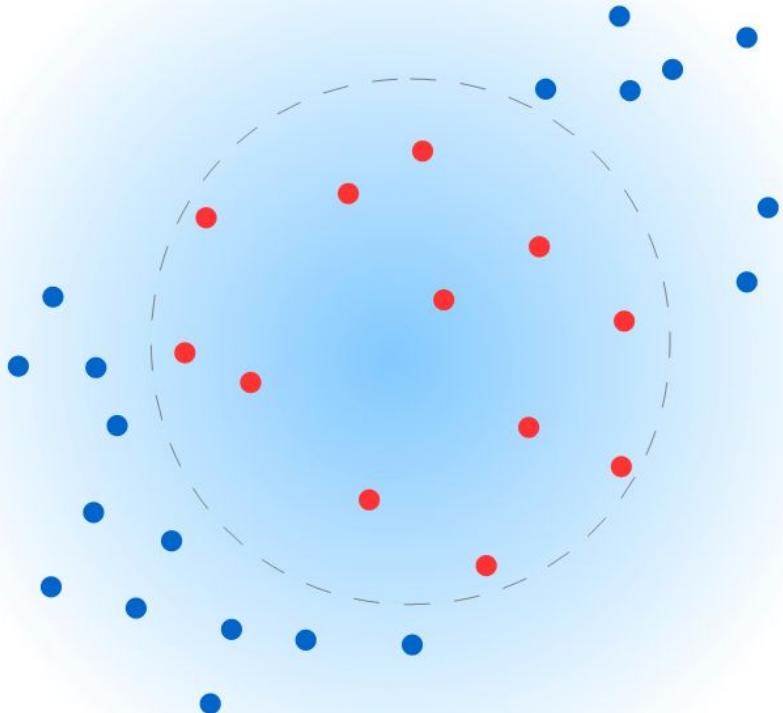
Преобразование признаков

- Почему все так сложно с этой дорогой?
- Ладно, по крайней мере пусть разделительная линия будет несложной





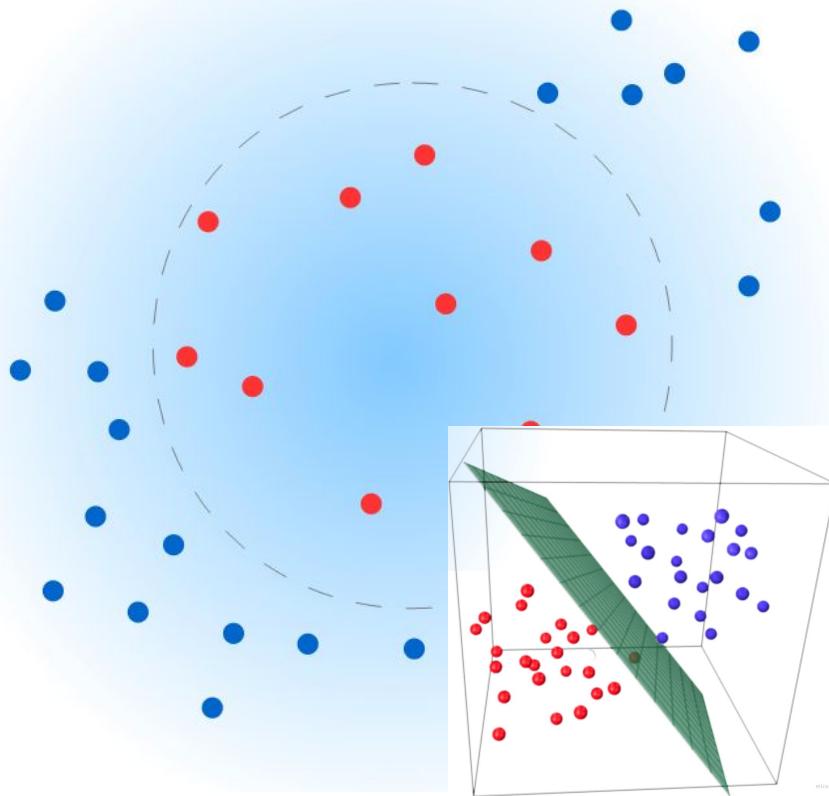
Преобразование признаков



- Почему все так сложно с этой дорогой?
- Ладно, по крайней мере пусть разделительная линия будет несложной
- Введем новое пространственное измерение



Преобразование признаков

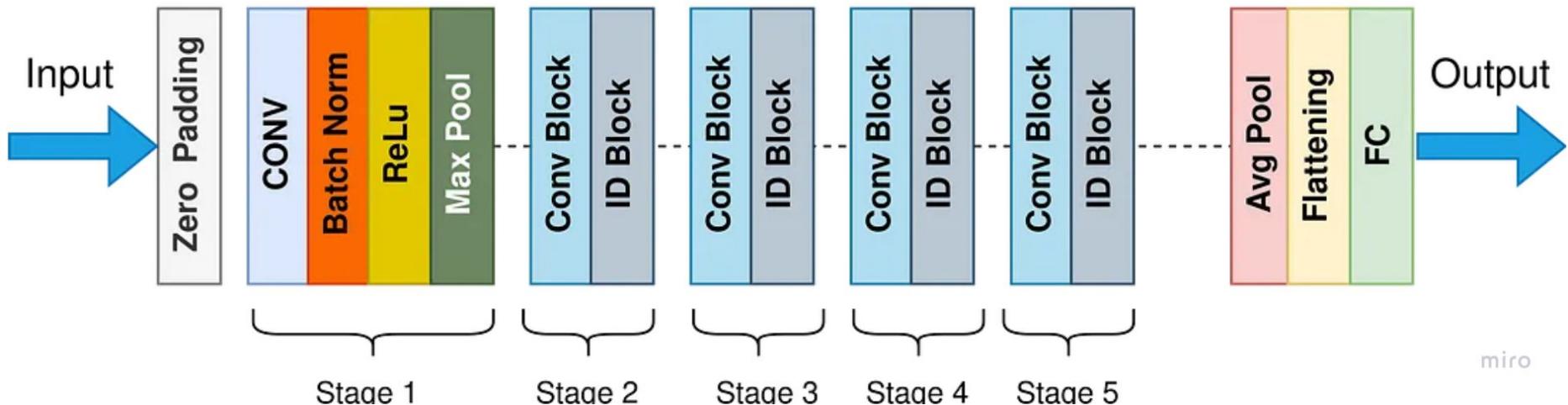


- Почему все так сложно с этой дорогой?
- Ладно, по крайней мере пусть разделительная линия будет несложной
- Введем новое пространственное измерение



Эмбеддинги многослойных нейросетей

Схема нейросети ResNet50

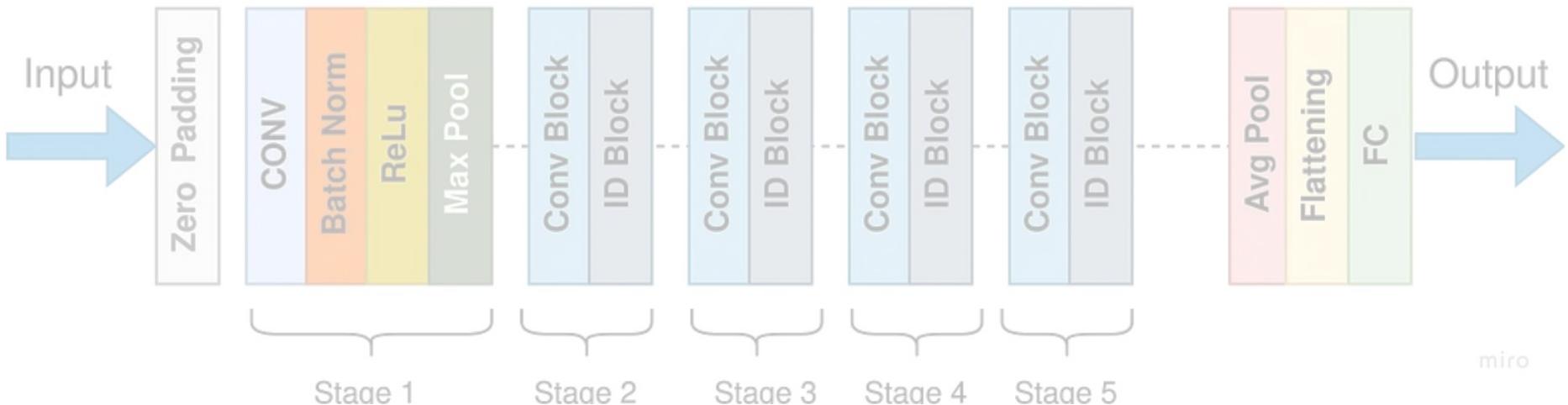


miro



Эмбеддинги многослойных нейросетей

Схема нейросети ResNet50

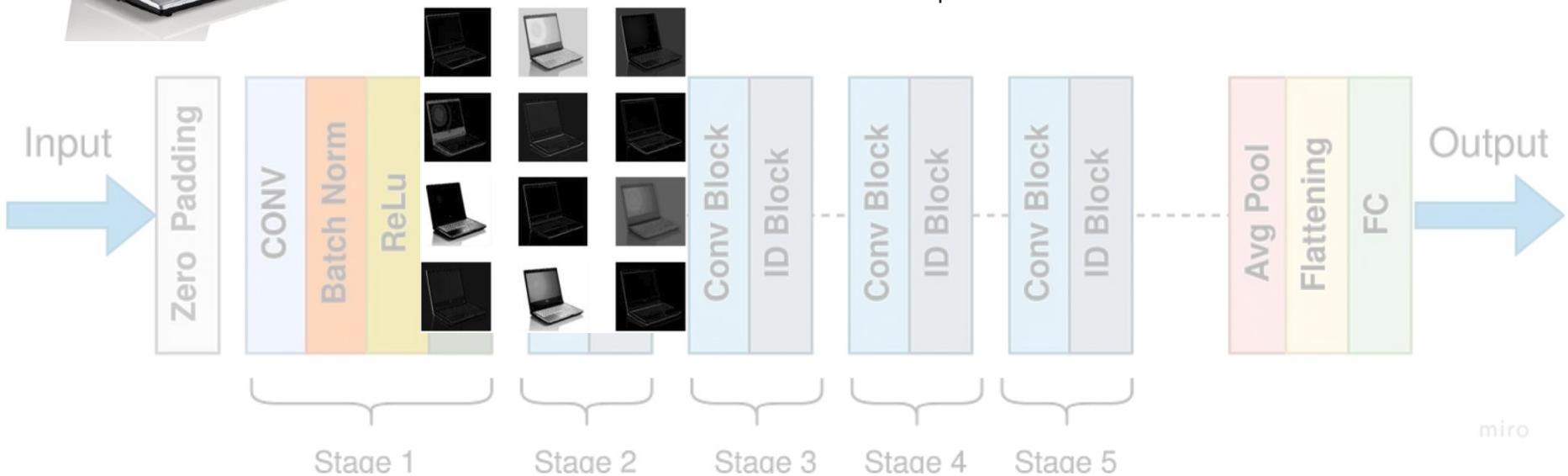


miro



Эмбеддинги многослойных нейросетей

Схема нейросети ResNet50

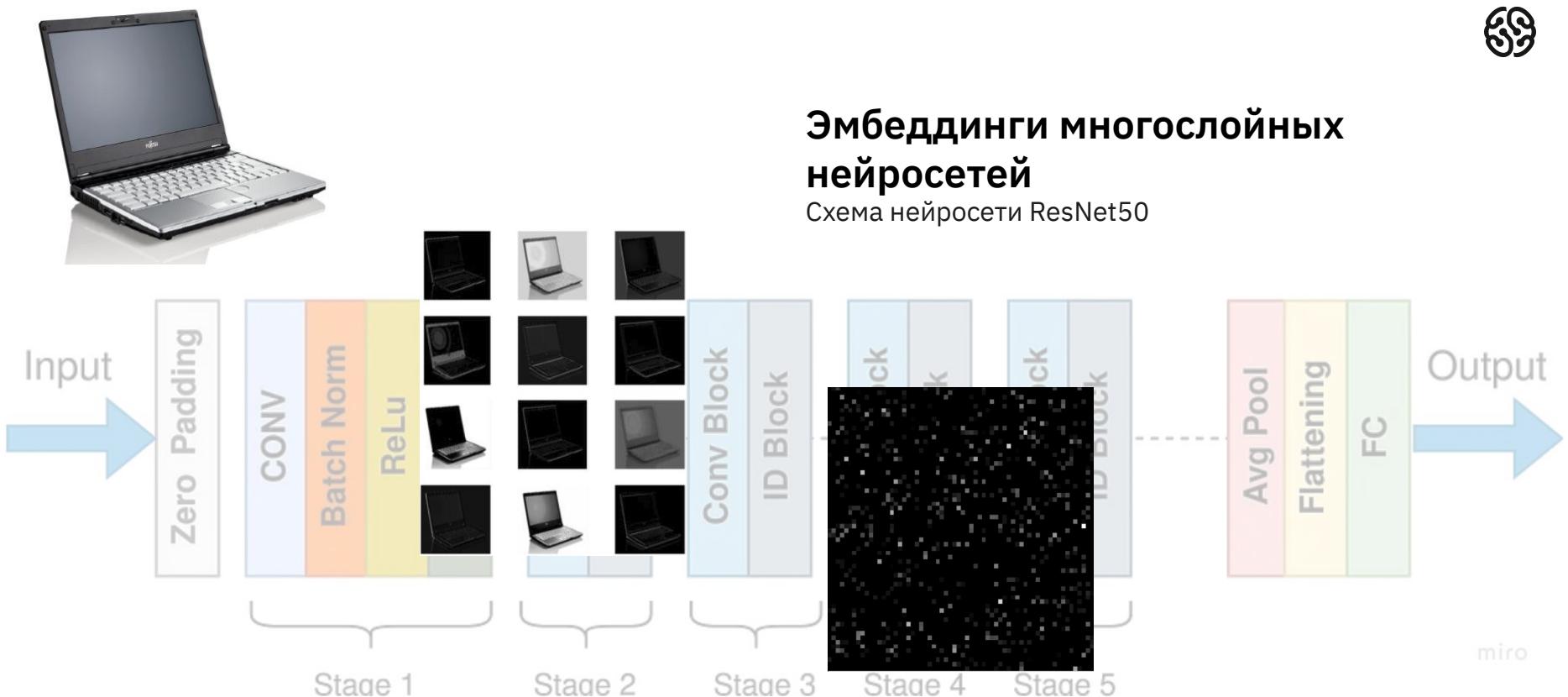


miro



Эмбеддинги многослойных нейросетей

Схема нейросети ResNet50

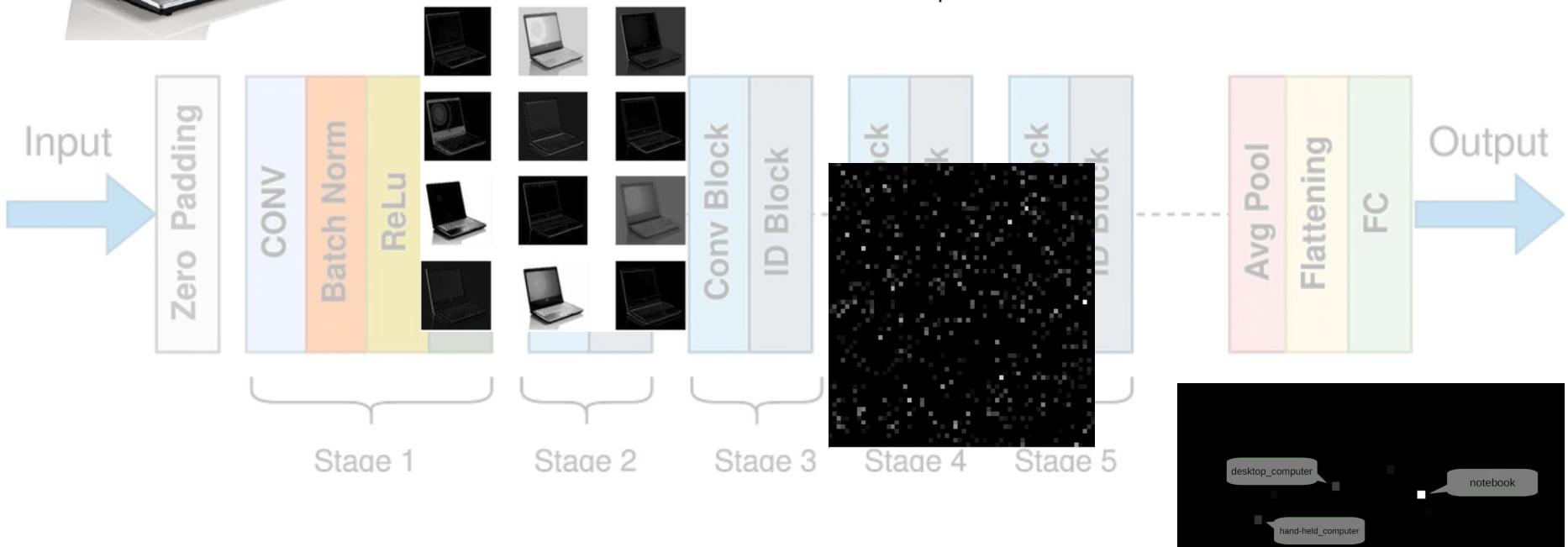


miro



Эмбеддинги многослойных нейросетей

Схема нейросети ResNet50

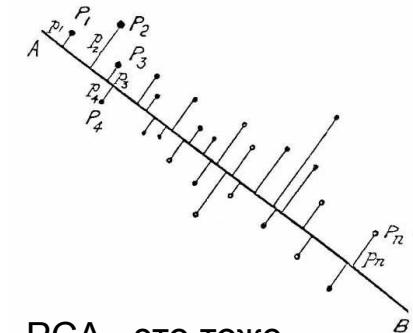
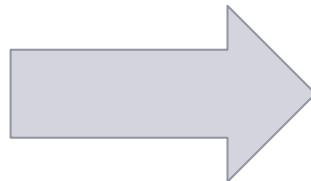




Что такое эмбеддинг изображения



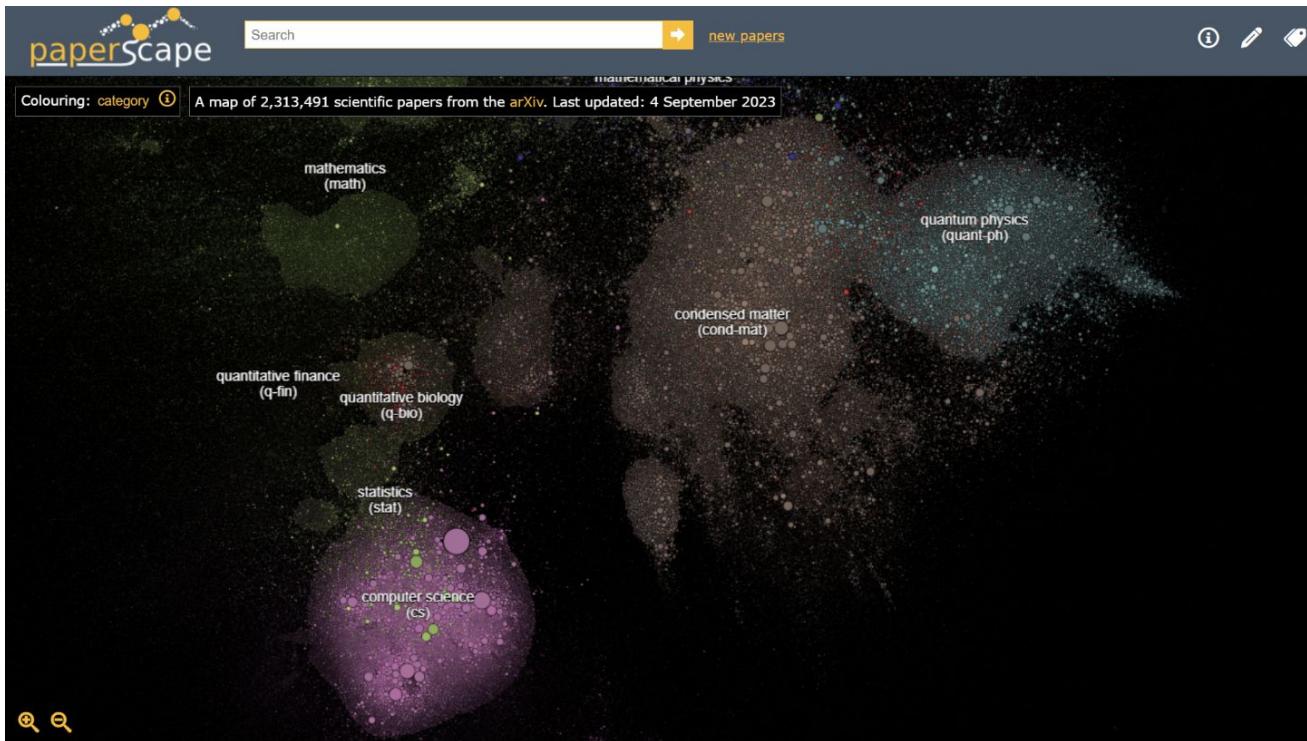
это эмбеддинг



PCA - это тоже
эмбеддинг

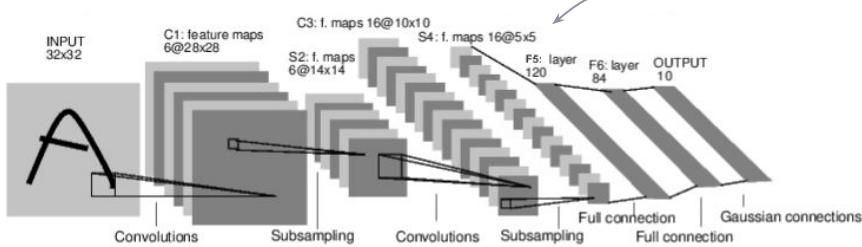


Что такое эмбеддинг изображения

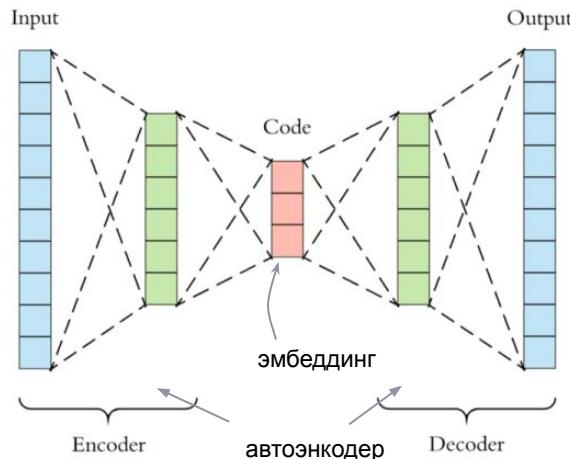




нейросеть AlexNet (~2012 г.)



сколько параметров можно улучшить?



Как создаются нейросети с хорошими эмбеддингами



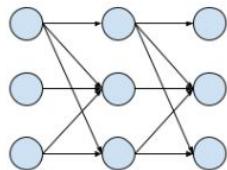
моделирование маскированных слов языка (MLM)



пример задачи, допускающей самообучение
(self-supervised learning)

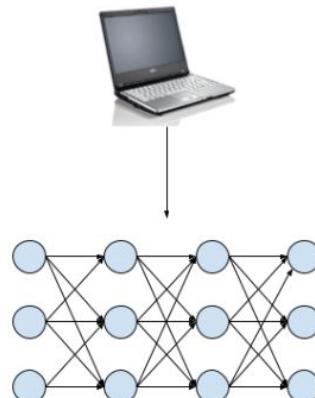


В каких задачах применяются эмбеддинги картинок



Классификация

Машинное обучение



Создание признаков + классификация

Глубокое обучение



Пиксели как признаки?

Neural Network	Depth	Size	Parameters (Millions)	Image Input Size
squeezeNet	10	5 MB	1.24	227-by-227
googlenet		255 255 255 255 255 255 255 255 255 255 255 255 255 255 20 0 255 255 255 255 255 255 255 255 255 255 75 75 255 255 255 255 255 255 255 255 255 75 95 95 75 255 255 255 255 255 255 255 255 96 127 145 175 255 255 255 255 255 255 255 255 127 145 175 175 255 255 255 255 255 255 255 255 127 145 200 200 175 175 95 255 255 255 255 255 127 145 200 200 175 175 95 47 255 255 255 255 127 145 145 175 127 127 95 47 255 255 255 255 74 127 127 127 95 95 95 47 255 255 255 255 255 74 74 74 74 74 74 255	7.0	224-by-224
inceptionv3			23.9	299-by-299
densenet201			20.0	224-by-224
mobilenetv2			3.5	224-by-224
resnet18			11.7	224-by-224
resnet50			25.6	224-by-224
resnet101			44.6	224-by-224
xception			22.9	299-by-299
inceptionresne			55.9	299-by-299
shufflenet			1.4	224-by-224
nasnetmobile	*	20 MB	5.3	224-by-224
nasnetlarge	*	332 MB	88.9	331-by-331
darknet19	19	78 MB	20.8	256-by-256
darknet53	53	155 MB	41.6	256-by-256
efficientnetb0	82	20 MB	5.3	224-by-224
alexnet	8	227 MB	61.0	227-by-227
vgg16	16	515 MB	138	224-by-224
vgg19	19	535 MB	144	224-by-224

miro



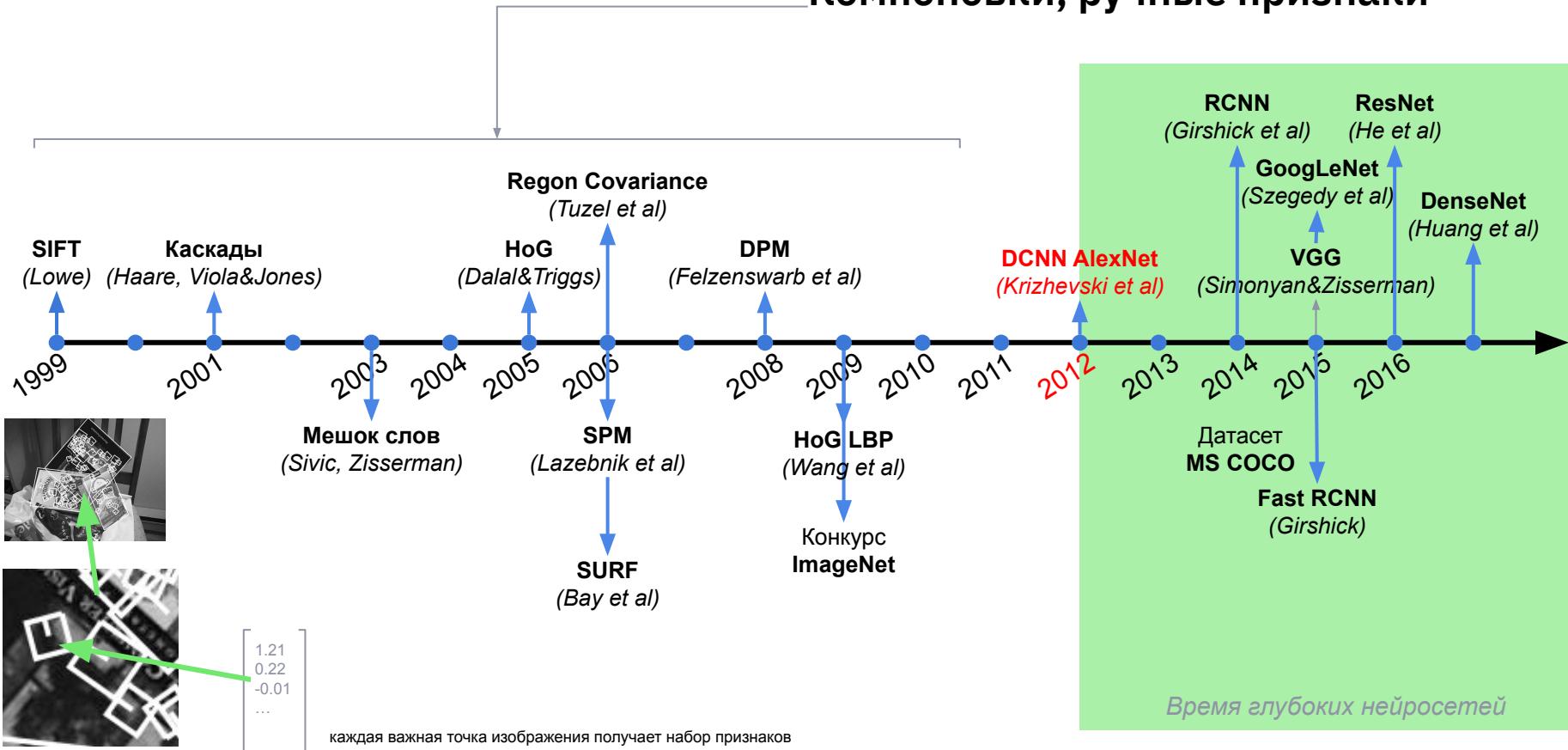
Пиксели как признаки?

Neural Network	Depth	Size	Parameters (Millions)	Image Input Size
squeezeNet	18	5.2 MB	1.24	227-by-227
googlenet	22	27 MB	7.0	224-by-224
inceptionv3	48	89 MB	23.9	299-by-299
densenet201	201	77 MB	20.0	224-by-224
mobilenetv2	53	13 MB	3.5	224-by-224
resnet18	18	44 MB	11.7	224-by-224
resnet50	50	96 MB	25.6	224-by-224
resnet101	101	167 MB	44.6	224-by-224
xception	71	85 MB	22.9	299-by-299
inceptionresnetv2	164	209 MB	55.9	299-by-299
shufflenet	50	5.4 MB	1.4	224-by-224
nasnetmobile	*	20 MB	5.3	224-by-224
nasnetlarge	*	332 MB	88.9	331-by-331
darknet19	19	78 MB	20.8	256-by-256
darknet53	53	155 MB	41.6	256-by-256
efficientnetb0	82	20 MB	5.3	224-by-224
alexnet	8	227 MB	61.0	227-by-227
vgg16	16	515 MB	138	224-by-224
vgg19	19	535 MB	144	224-by-224

miro

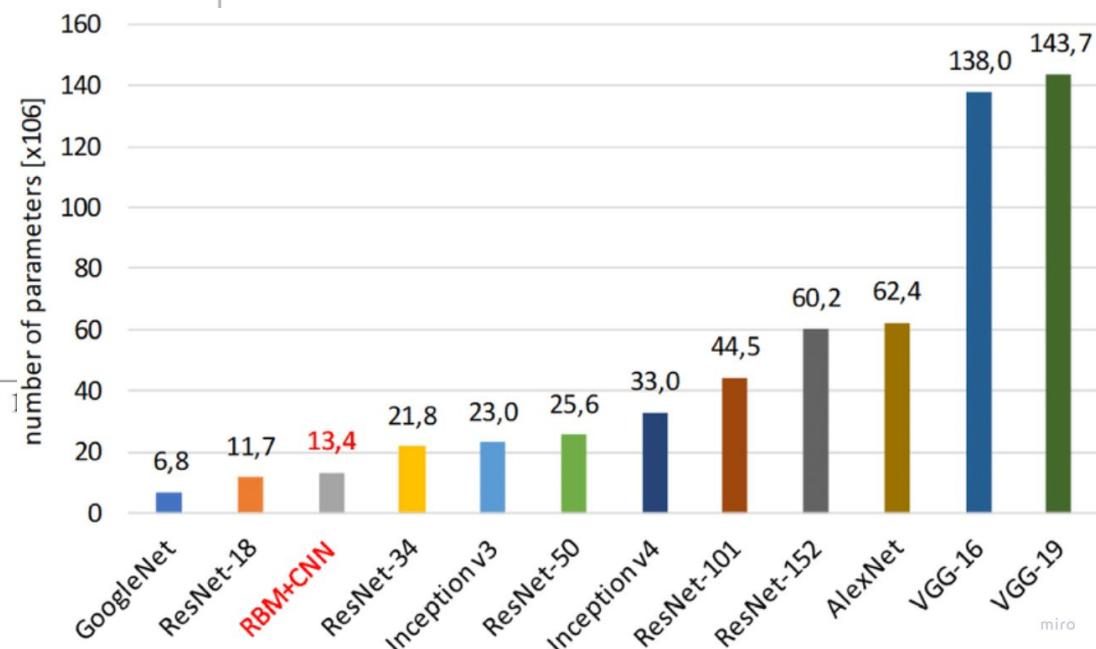
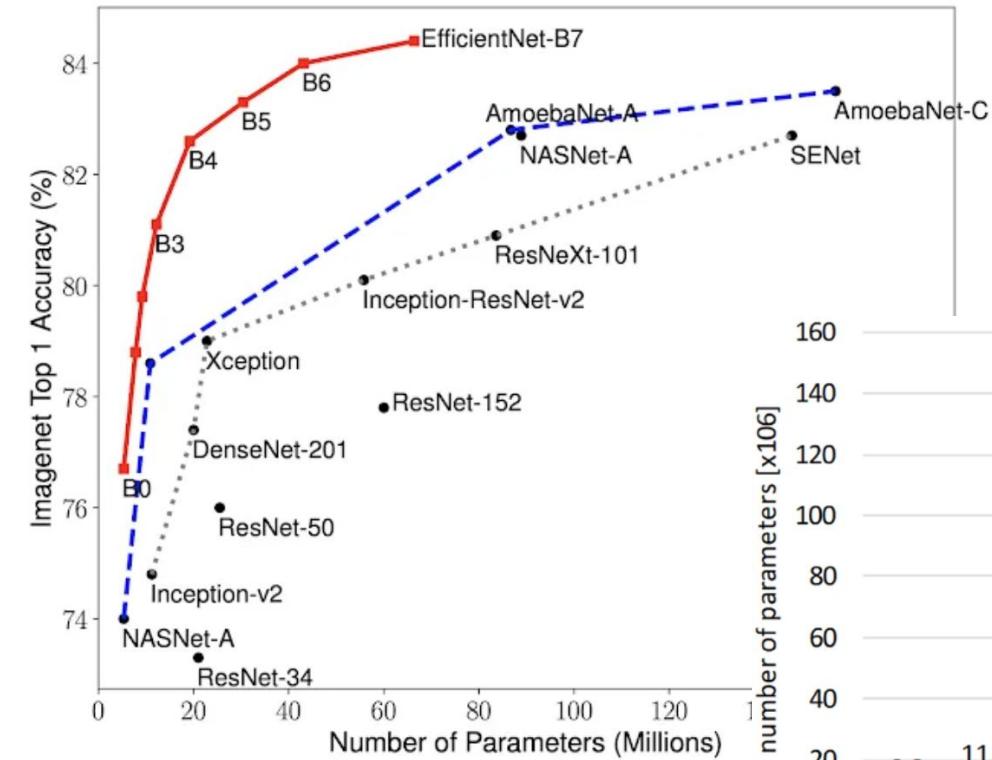


Компоновки, ручные признаки





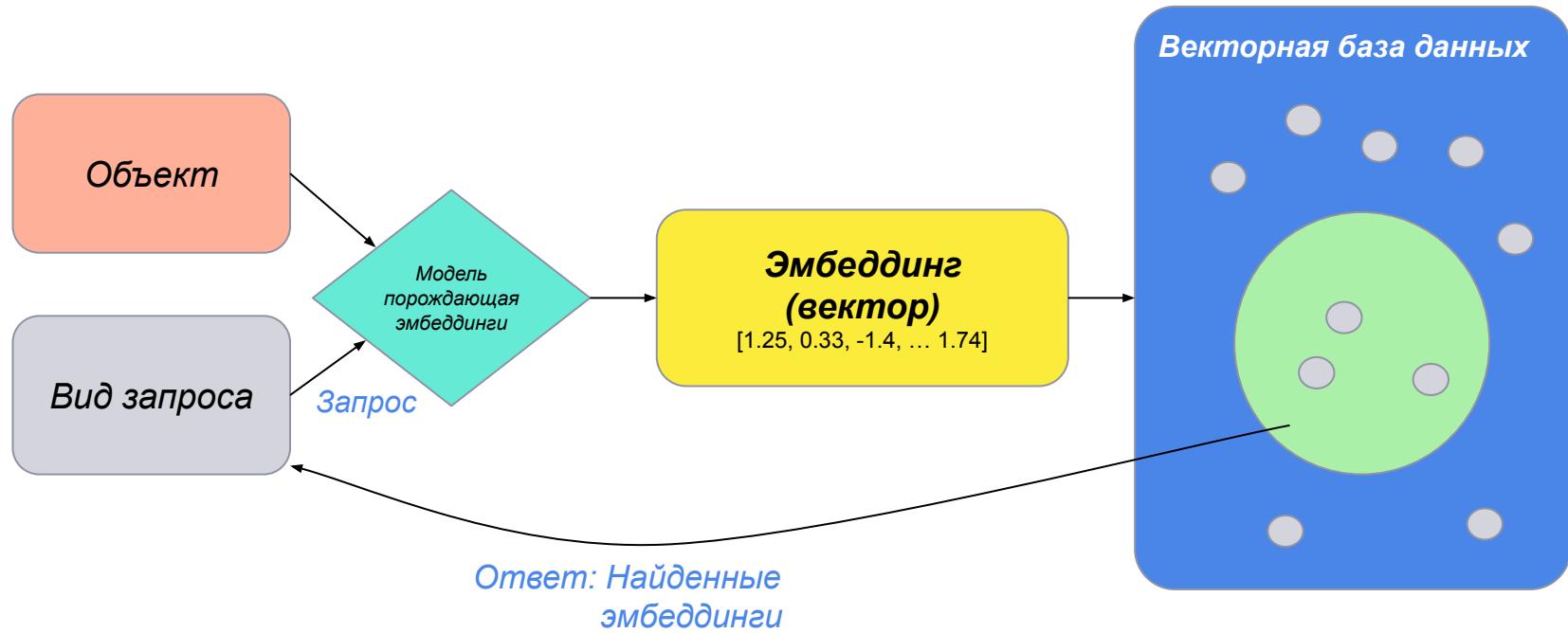
Эмбеддинги ResNet



miro

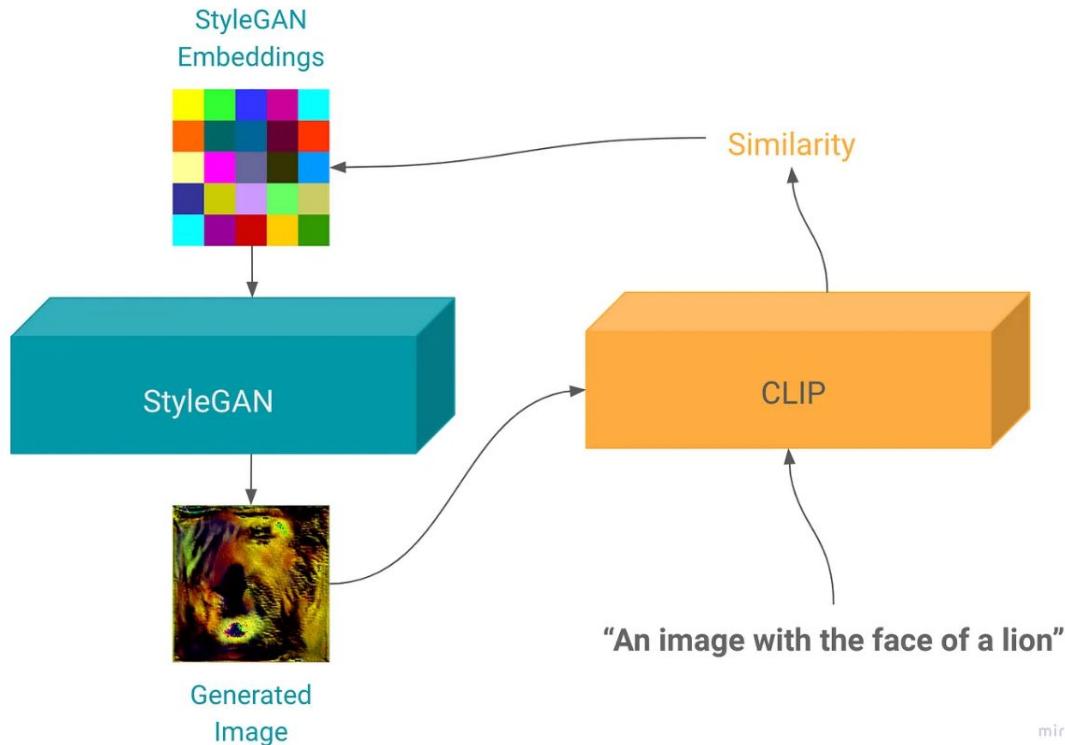


Базы данных эмбеддингов



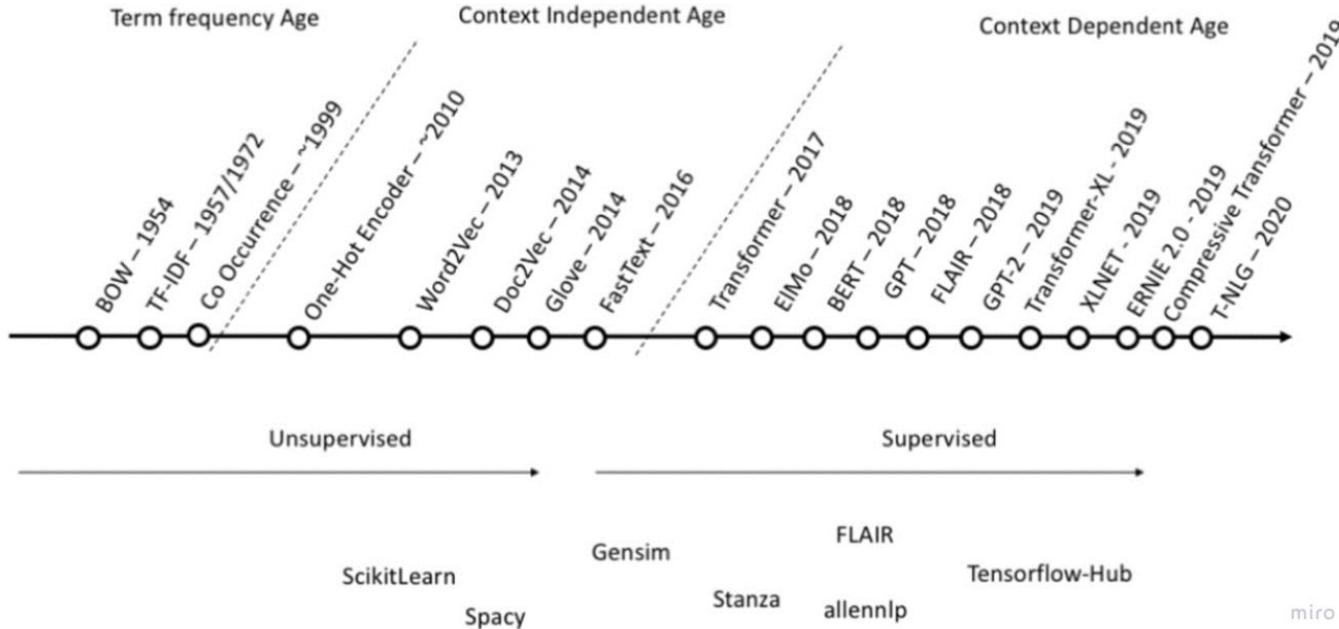


Генерация из эмбеддингов



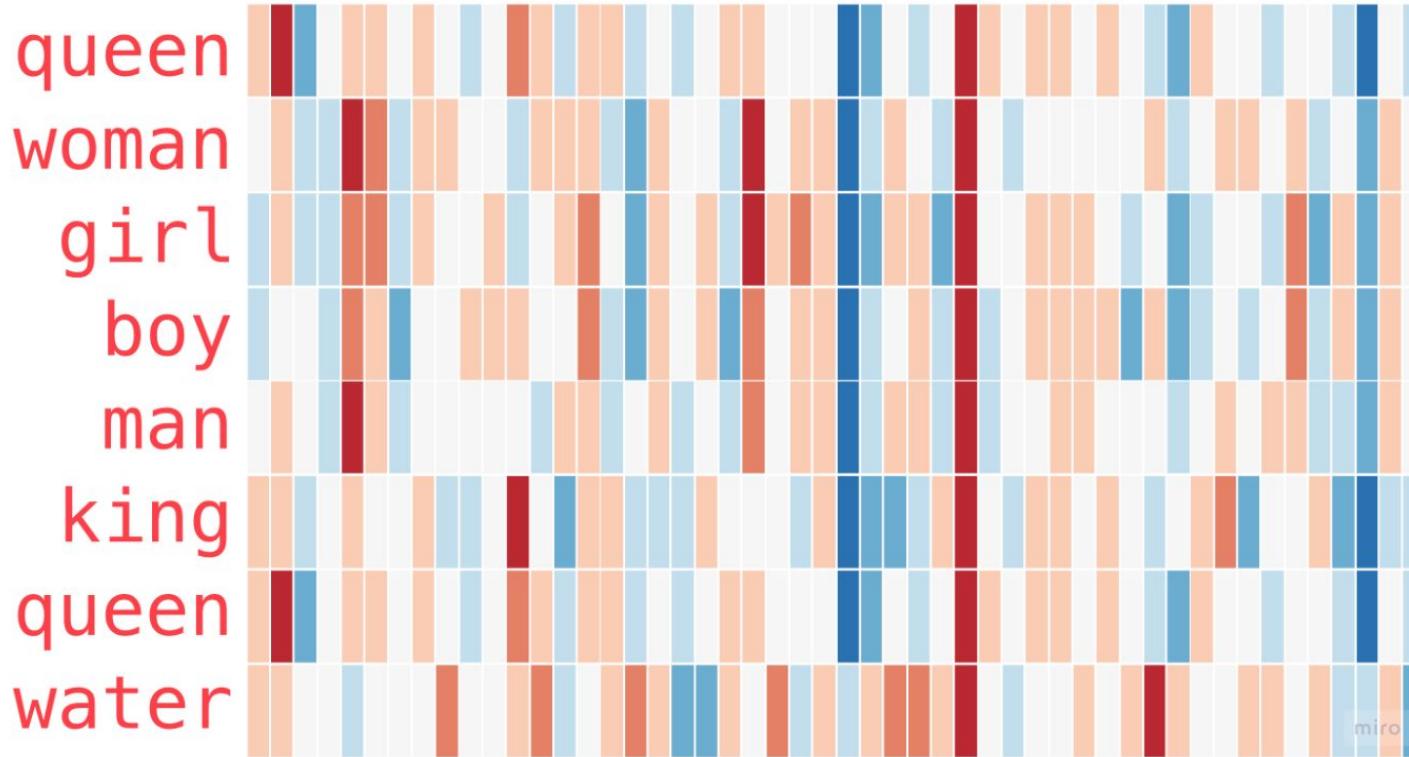


Словарь, one hot, частота слов, TFIDF и автоэнкодеры





word2vec и GloVe



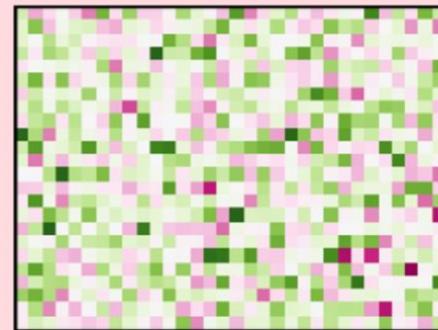
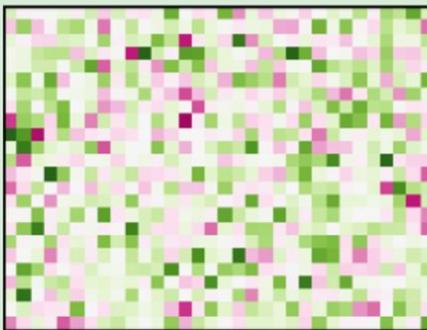


Эмбеддинги BERT

сочный плод яблони, который употребляется в **пищу** в свежем виде, служит сырьём в кулинарии и для приготовления напитков

колёсное **транспортное** средство (или спортивный снаряд), приводимое в движение мускульной силой человека через ножные **педали** или (крайне редко) через ручные рычаги

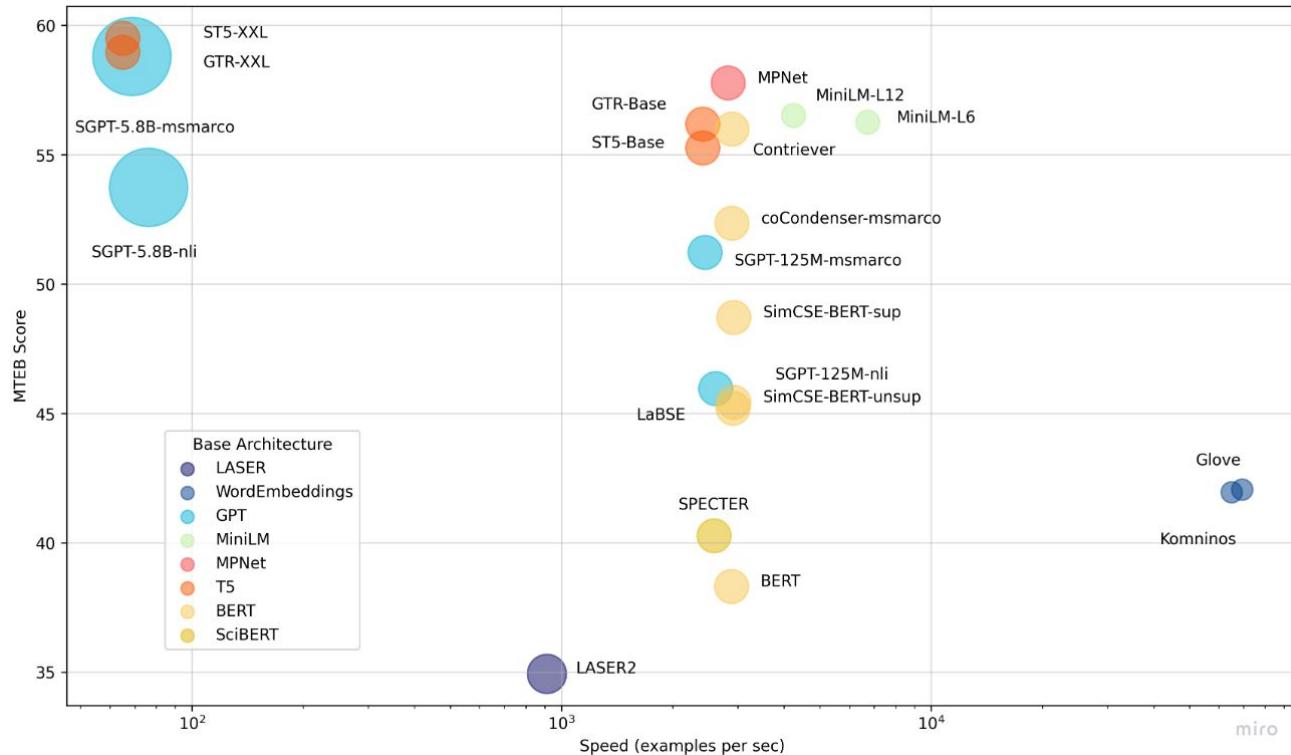
Эмбеддинги



miro

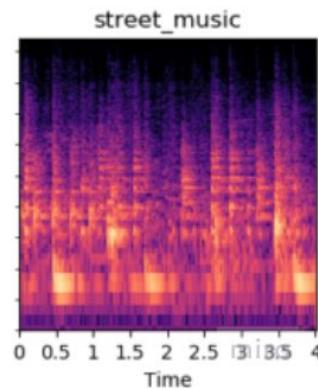
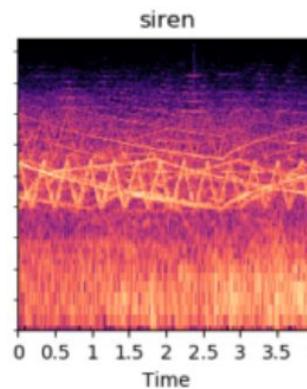
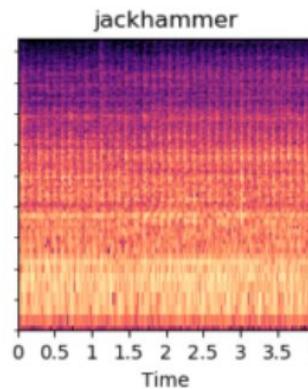
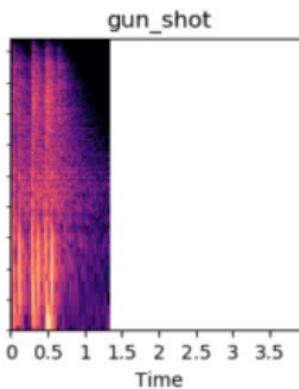
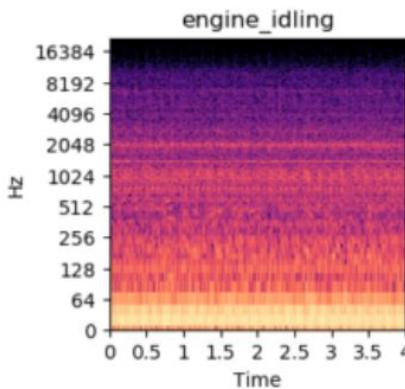
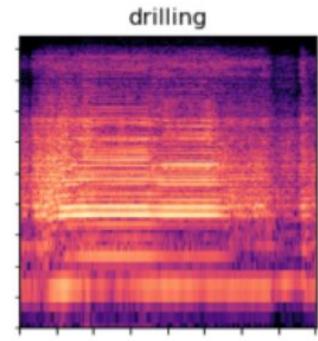
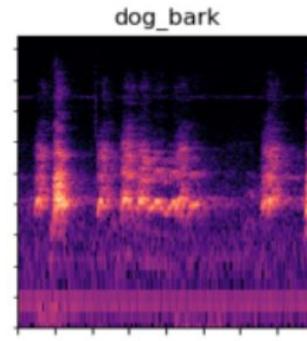
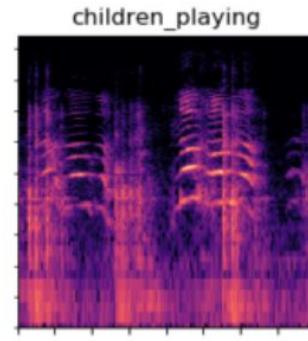
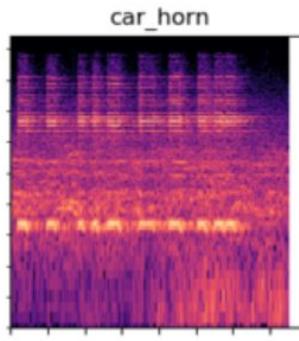
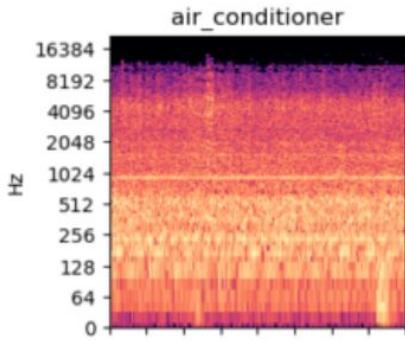


Эмбеддинги GPT и время API



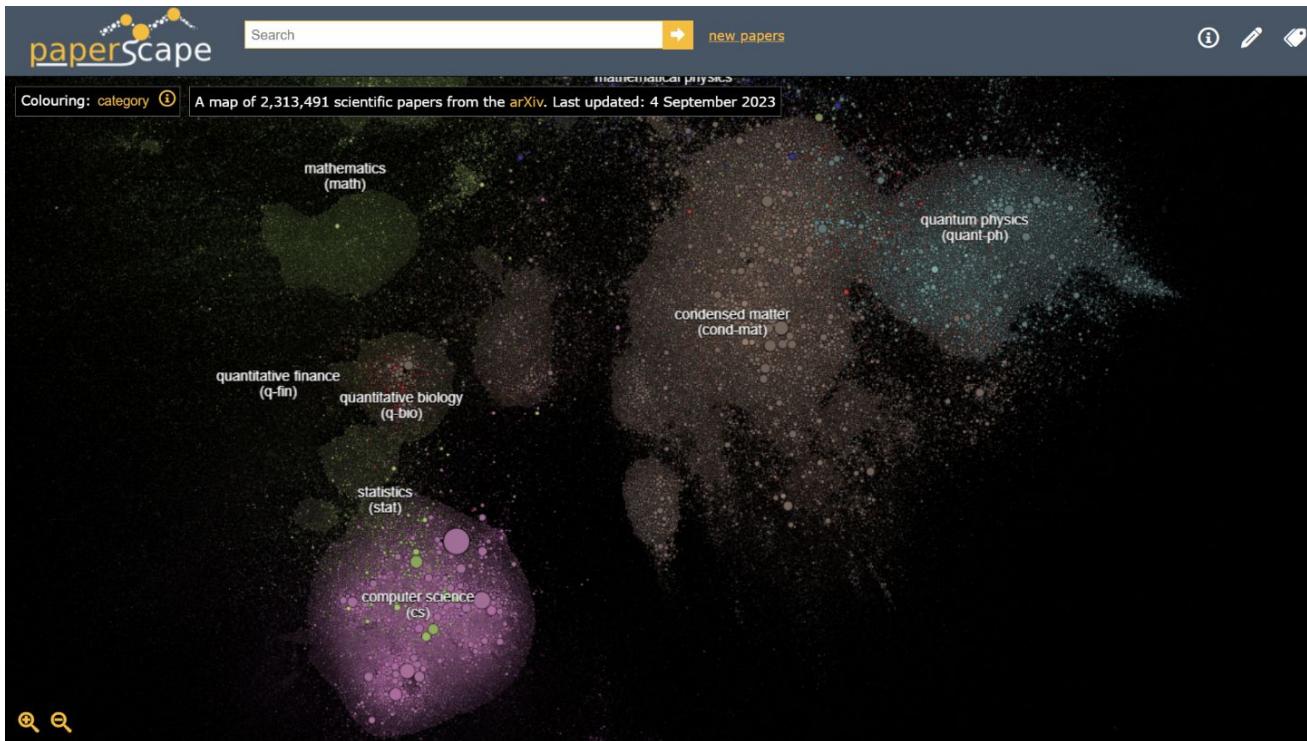


Спектры, вейвлеты, MFCC и нейронки





Снижение размерности – слишком локальные признаки





Спасибо за внимание

