CLIP embedding analysis

Huawei, Sirius

Semantic extraction and synthesis

Introduction

- Multi-modality approaches are of big interest of the SOTA research.
- Dimensionality reduction is important task in ML/DL.
- Managed synthesis is key issue in generative models applications.

We designed and applied methods for dimensionality reduction, clustering and managed synthesis and checked their correctness using CLIP.

We compared methods we used both quantitatively and qualitatively on SOTA frameworks.

CLIP basics

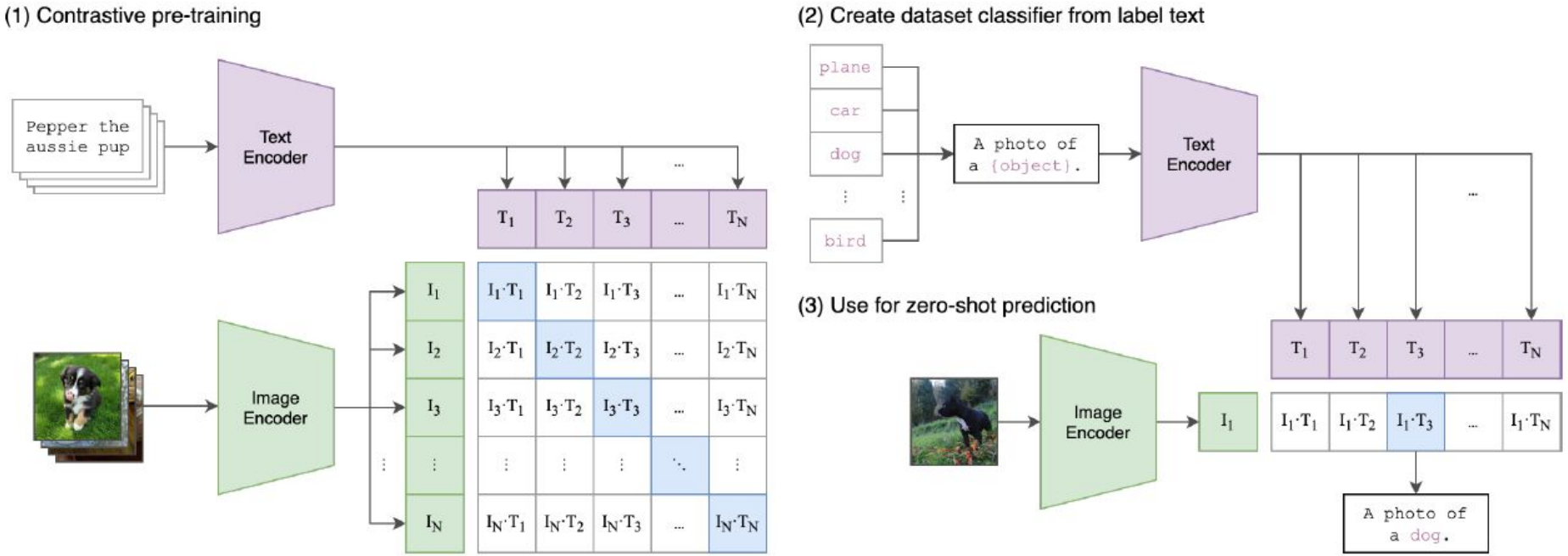


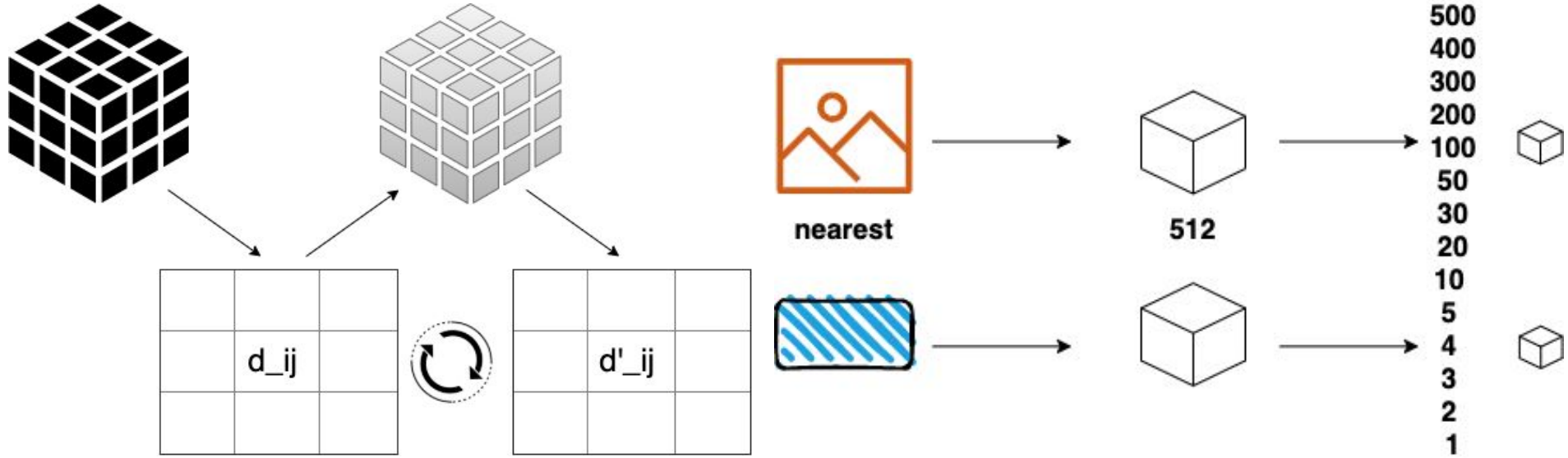
Figure 1: CLIP architecture

https://arxiv.org/abs/2103.00020

Dimensionality Reduction

Alexey Kolosov, Ekaterina Orlova

Task: Investigate dimensionality reduction methods and show that for the presented data there exist such embedding dimensionality D’ < D which doesn’t decrease embeddings correspondence quality (top-1 accuracy).



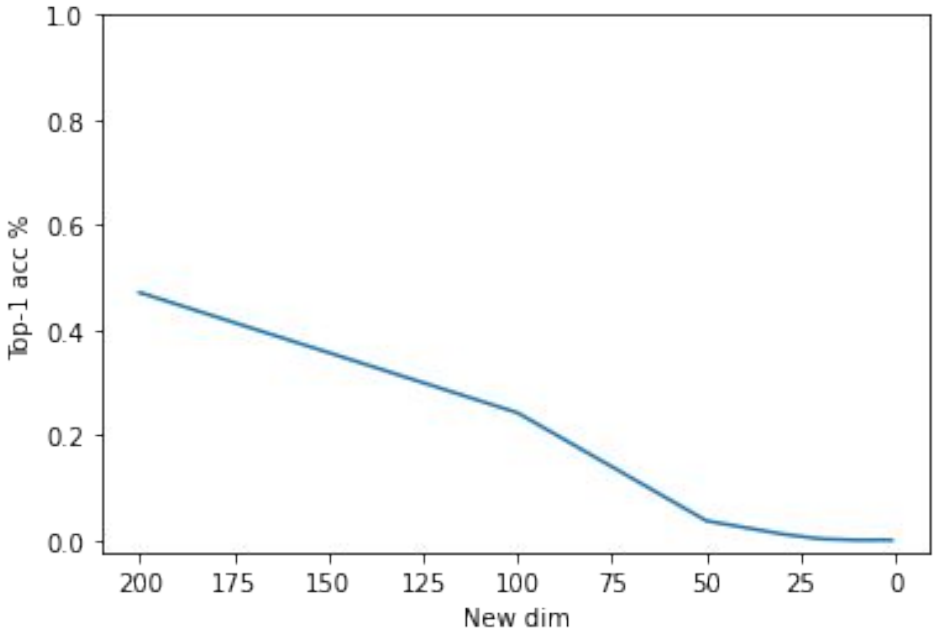
Neural MDS

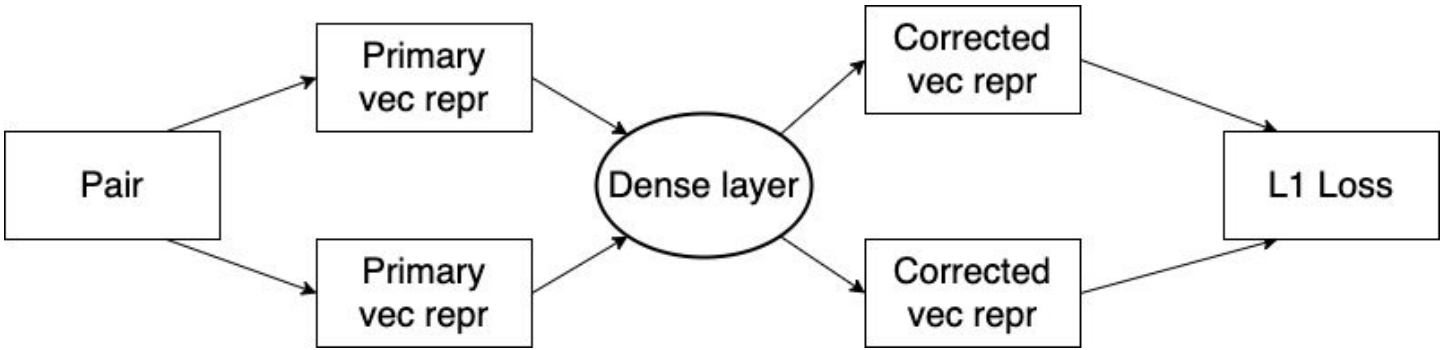
Alexey Kolosov

Problems statements

1. **COCO, isometric, train**
2. **COCO, isotonic, train**

$$d(i,j) = e(g(i), g(j))$$





Top-1 accuracy

val2017, isometric, train

5000 pairs, 200 epochs

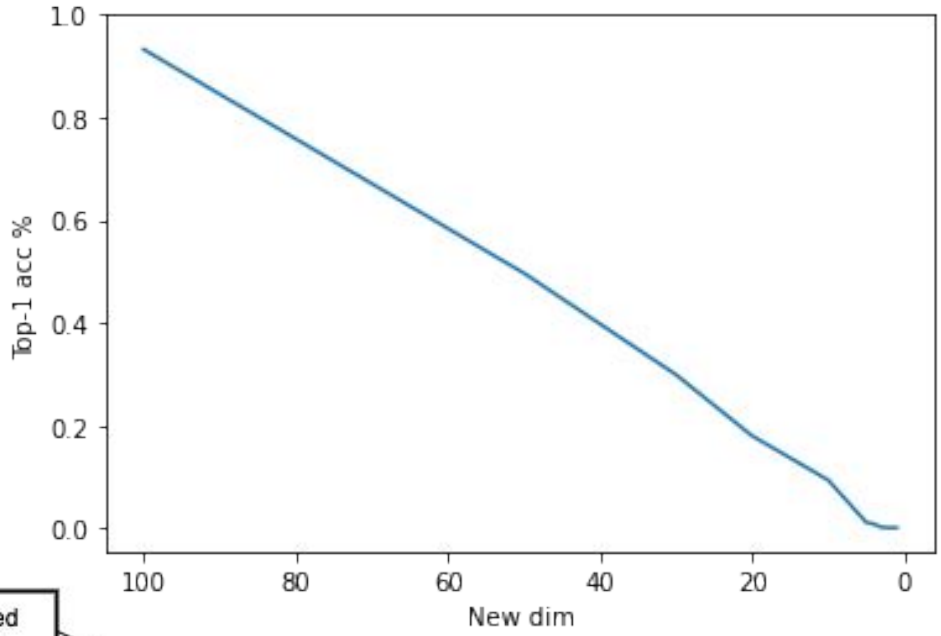
for 512 dim - 1639 pairs

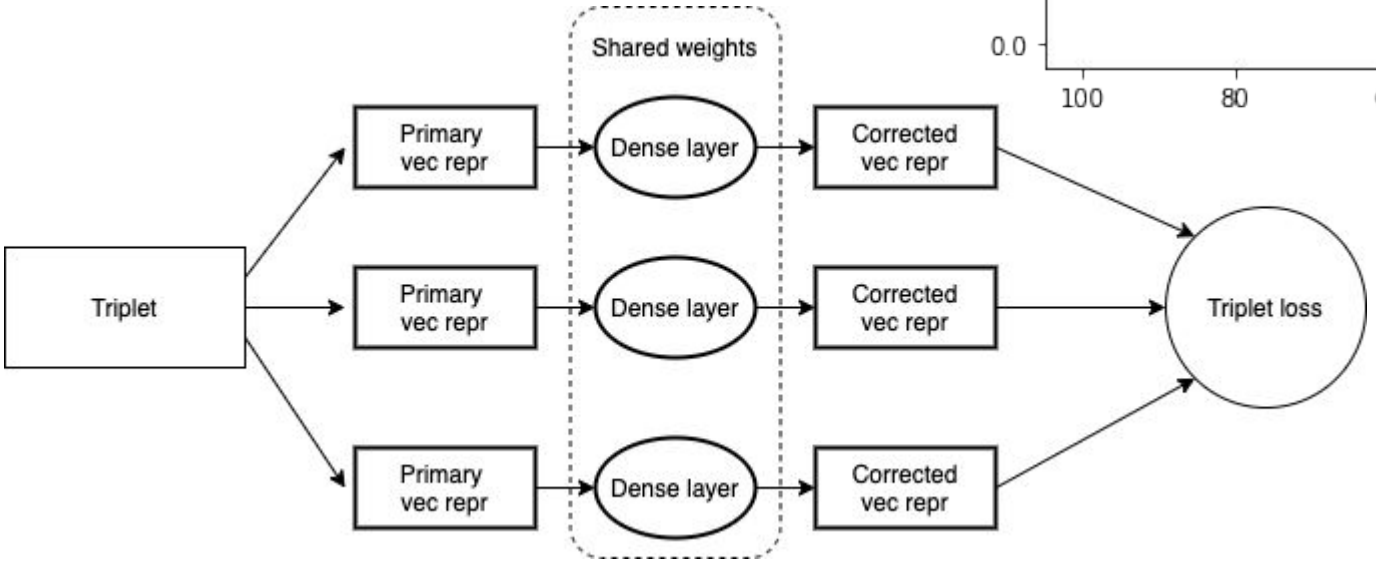
Neural MDS

Alexey Kolosov

Isotonic problem result

$$d(i,j) < d(k,l) \Rightarrow e(g(i),g(j)) < e(g(k),g(l))$$





Top-1 accuracy

COCO, isotonic, train

8000000 quads, 200 epochs

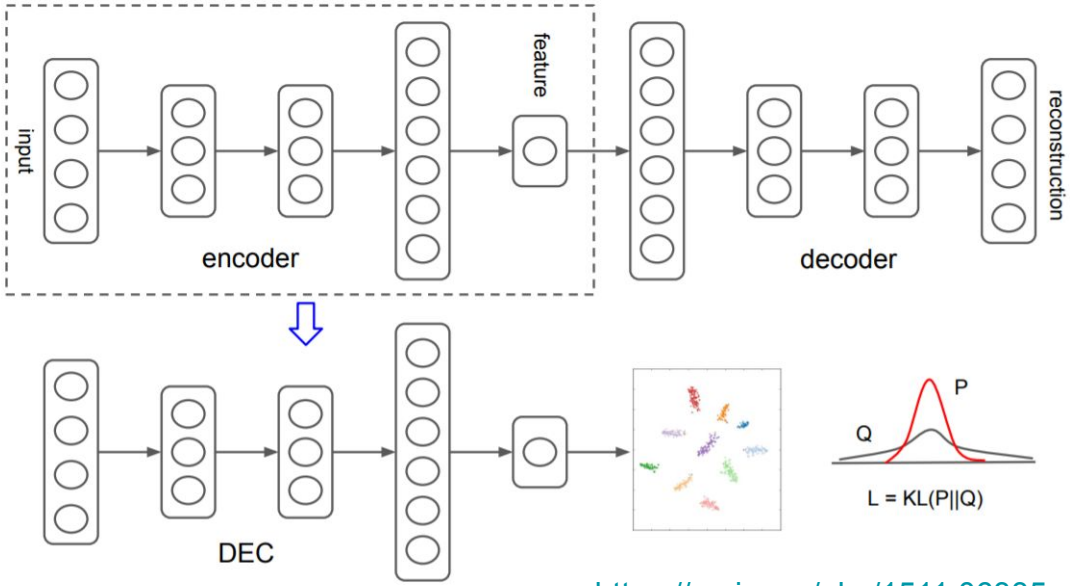
for 512 dim - 1639 pairs

Embeddings clusterization

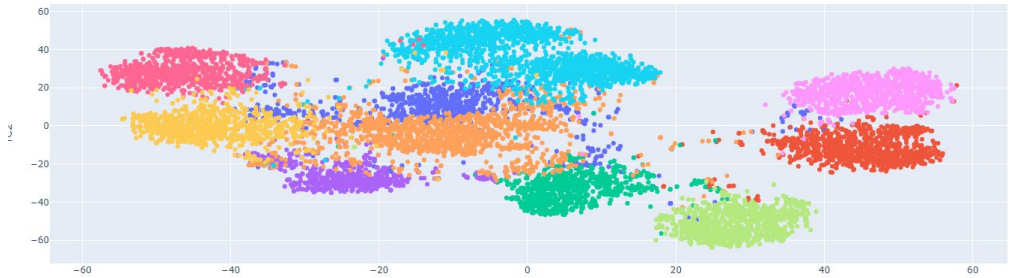
Abdullaeva Uma, Anna Dmitrienko, Sergey Skorik, Anna Rudenko

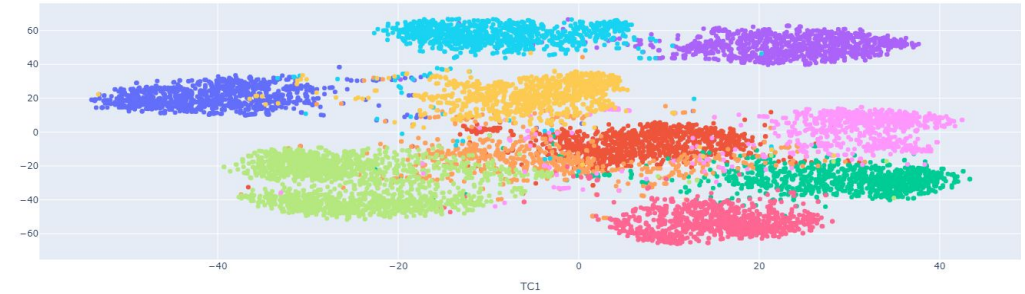
Task: Investigate the clusterization methods and show that there exist clusters in the embeddings data. Perform the visualization of these data clusters.

Deep embedded clustering (DEC) model



https://arxiv.org/abs/1511.06335

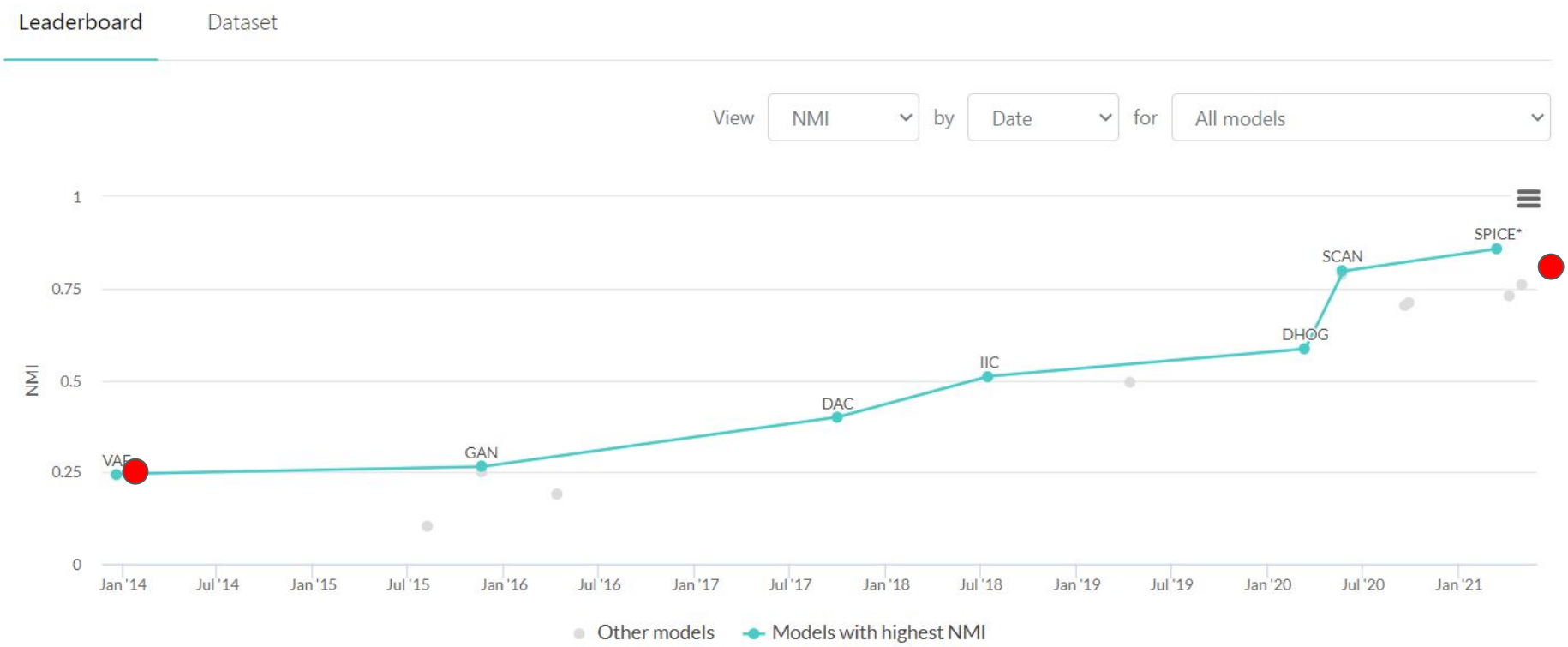




	AMI	ARI	FMI	NMI
K-Means	0.72	0.63	0.67	0.72
Auto - encoder	0.74	0.67	0.70	0.76
DEC	**0.82**	**0.77**	**0.79**	**0.82**

Embeddings clusterization

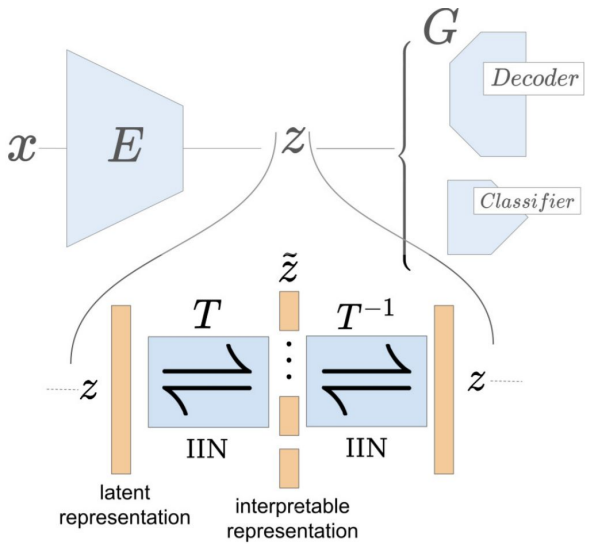
Abdullaeva Uma, Anna Dmitrienko, Sergey Skorik, Anna Rudenko

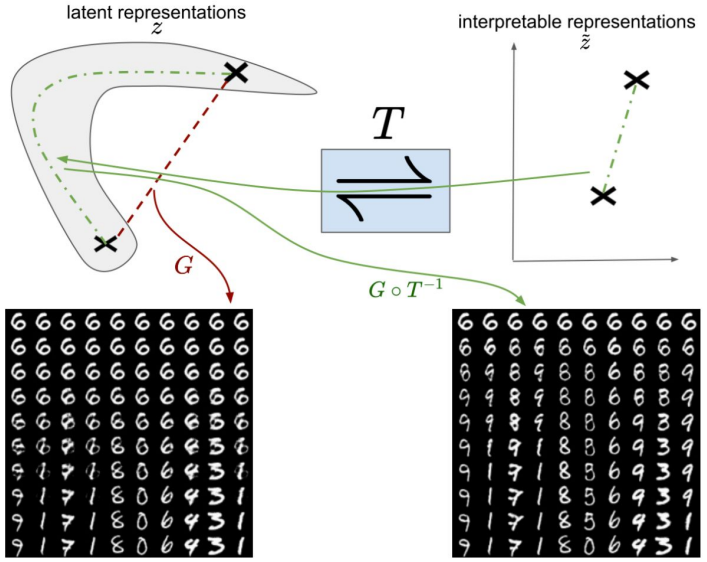


Disentanglement

Ekaterina Orlova, Anna Dmitrienko, Sergey Skorik, Anna Rudenko, Abdullaeva Uma

Task: Propose a method for disentangling NN's internal representation, so that the untangled representation has an understandable interpretation and independent components.



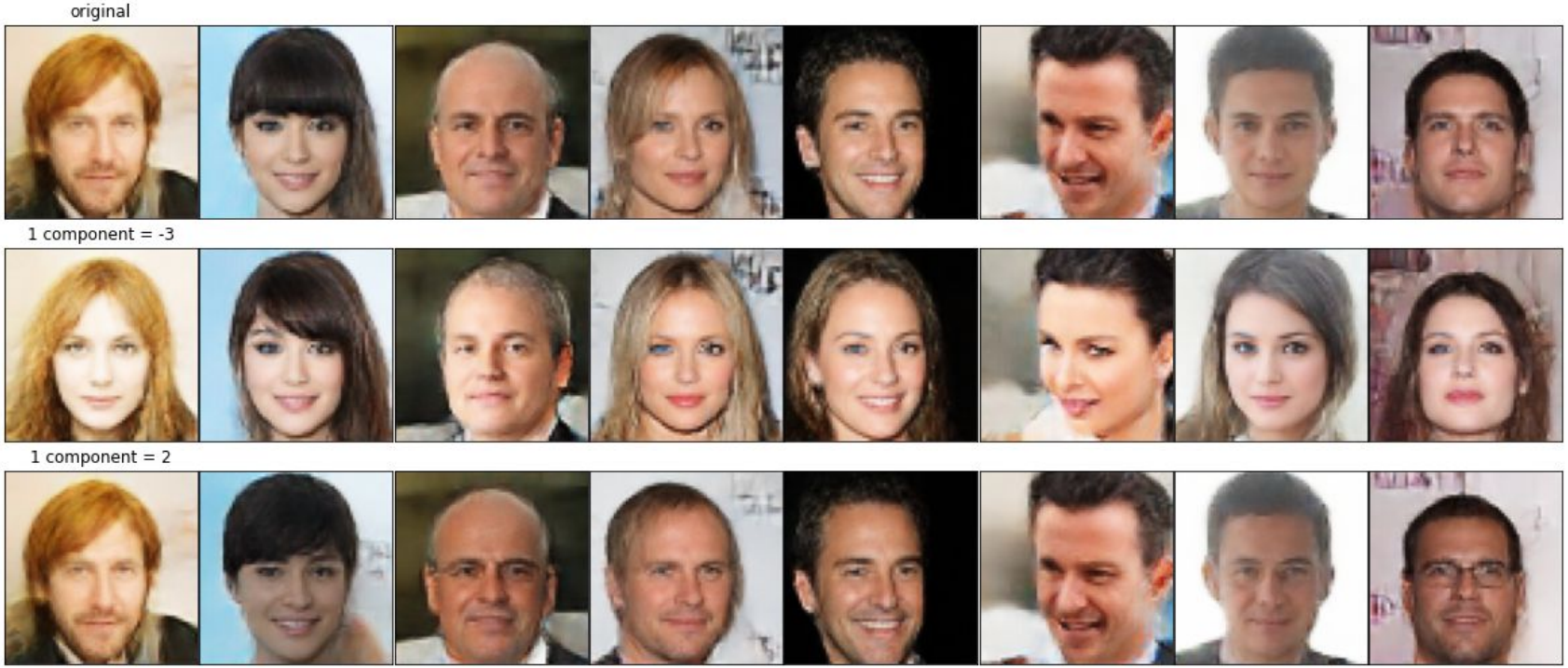


https://arxiv.org/pdf/2004.13166.pdf

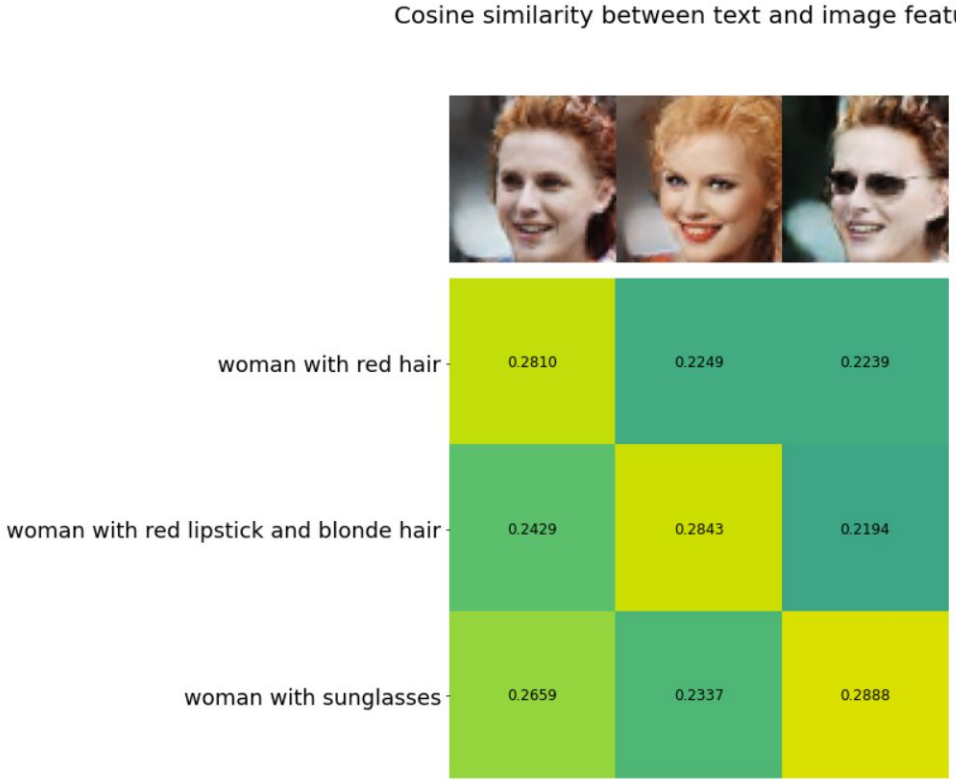
CelebA - Glasses

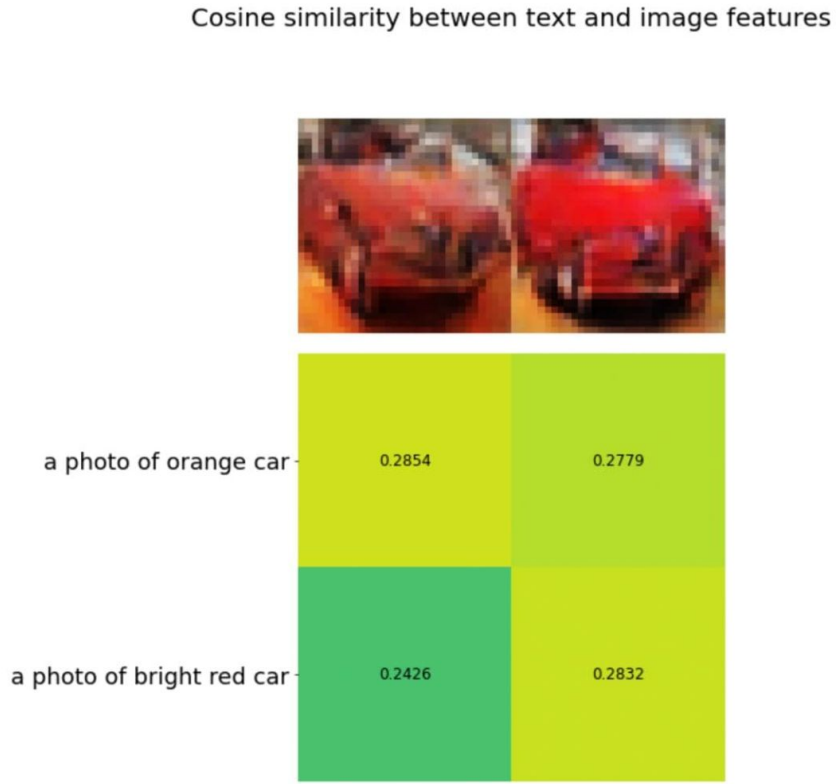
CelebA - Skin

CelebA - Sex



CLIP


Cosine similarity between text and image features
woman with red hair
woman with red lipstick and blonde hair
woman with sunglasses
0.2810
0.2249
0.2239
0.2429
0.2843
0.2194
0.2659
0.2337
0.2888


Cosine similarity between text and image features
a photo of orange car
a photo of bright red car
0.2854
0.2779
0.2426
0.2832

Conclusions and plans

Proposed new method for dimensionality reduction of CLIP embeddings.

Proposed and evaluated new method for clusterization, close to SOTA

Latent representations of various VAE were interpreted and evaluated by CLIP

Future:

Automatic data augmentation with text descriptions

Managed image synthesis

https://github.com/nihao88/SemanticSirius2021