

DataComp

Доклад подготовил Воробьев Дмитрий

- DataComp – в первую очередь **фреймворк** для тестирования, а не научное исследование. Призван упростить и стандартизировать исследования.
- Мотивация: редко тестируют датасет для обучения, чаще архитектуру моделей.
- DataComp про тестирование мультимодальных наборов данных из пар **изображение-текст** с архитектурой **CLIP**.
- Пользователи фреймворка тестируют обучающие данные при **фиксированных** параметрах архитектуры.

- Не всегда понятен вклад новых данных в итоговый результат.
- Основное внимание уделяется архитектуре сети, но датасет также может быть важен.

Zero-shot performance of CLIP models trained on different datasets

Dataset	Dataset size	# samples seen	Architecture	Train compute (MACs)	ImageNet accuracy
OpenAI's WIT [111]	0.4B	13B	ViT-L/14	1.1×10^{21}	75.5
LAION-400M [128, 28]	0.4B	13B	ViT-L/14	1.1×10^{21}	72.8
LAION-2B [129, 28]	2.3B	13B	ViT-L/14	1.1×10^{21}	73.1
LAION-2B [129, 28]	2.3B	34B	ViT-H/14	6.5×10^{21}	78.0
LAION-2B [129, 28]	2.3B	34B	ViT-g/14	9.9×10^{21}	78.5
DATAComp-1B (ours)	1.4B	13B	ViT-L/14	1.1×10^{21}	79.2

Основной вклад авторов

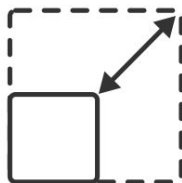
4

- **DataComp benchmark** – система бенчмаркинга. Лидерборд доступен на сайте: datacomp.ai
- **CommonPool** – набор данных из 12.8 млрд пар изображение-текст, собранный из интернета.
- **DataComp-1B** – мультимодальный датасет, собранный с помощью реализованных авторами алгоритмов фильтрации из набора CommonPool.

Benchmark workflow

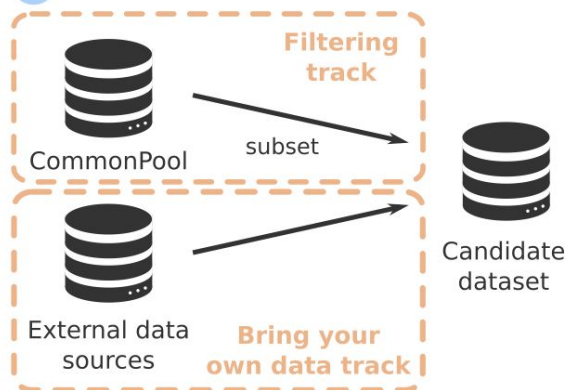
5

A Choose scale



Choose scale:
small, medium,
large or xlarge

B Select data



C Train



Train a CLIP model
with a fixed architecture
and hyper-parameters

D Evaluate

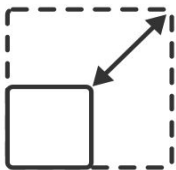


Evaluate the model
on 38 zero-shot
downstream tasks

Используемая архитектура

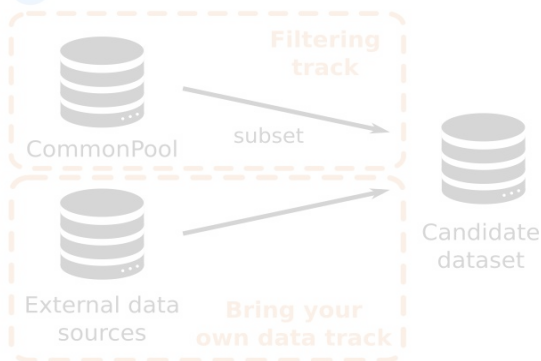
6

A Choose scale



Choose scale:
small, medium,
large or xlarge

B Select data



C Train



Train a CLIP model
with a fixed architecture
and hyper-parameters

D Evaluate



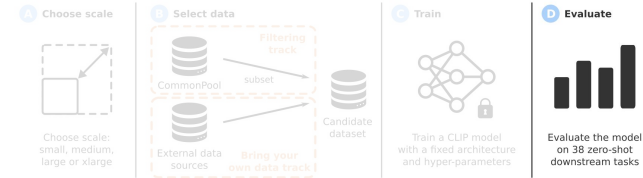
Evaluate the model
on 38 zero-shot
downstream tasks

Scale	Model	Train compute (MACs)	Pool size and # samples seen
small	ViT-B/32	9.5×10^{16}	12.8M
medium	ViT-B/32	9.5×10^{17}	128M
large	ViT-B/16	2.6×10^{19}	1.28B
xlarge	ViT-L/14	1.1×10^{21}	12.8B

Evaluate

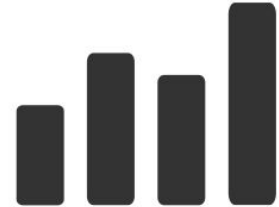
7

тестируется CLIP на задачи классификации и поиска (retrieval) по принципу zero-shot



- 6 ImageNet-ов в т. ч. с разными распределениями:
 - ImageNet-Sketch, ImageNet-V2, ImageNet-A, ImageNet-O, ImageNet-R, ObjectNet
- 13 датасетов из VTAB (Visual Task Adaptation Benchmark)
- 3 датасета из WILDS (A Benchmark of in-the-Wild Distribution Shifts)
- 22 задачи из статьи “Learning Transferable Visual Models From Natural Language Supervision”
- Retrieval datasets: Flickr30k, MSCOCO, WinoGavil

D Evaluate

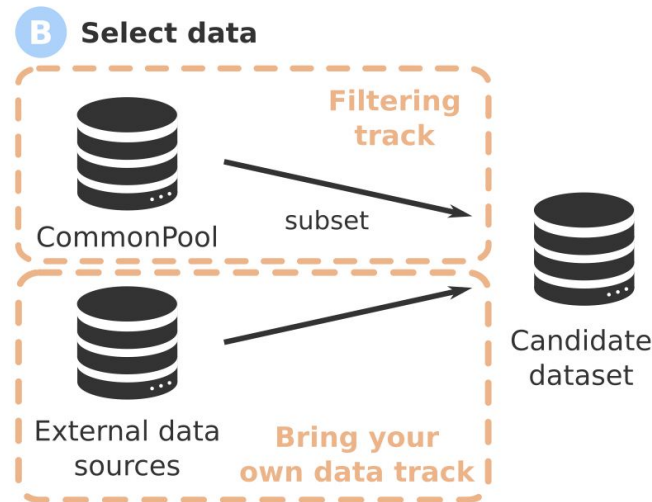
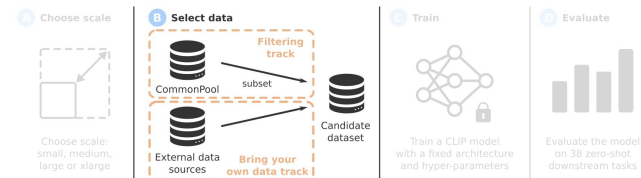


Evaluate the model on 38 zero-shot downstream tasks

Подготовка данных

8

- Filtering
 - Фильтрация предоставляемых авторами данных, собранных в CommonPool.
 - Пользователю необходимо реализовать собственный алгоритм фильтрации и/или использовать предоставляемые авторами.
- BYOD (Bring your own data)
 - Добавление собственного датасета пользователя.



Filtering baselines

9

- No filtering.
- Random subsets.
- Basic filtering:
 - фильтрация по длине текста (более двух слов и пяти символов)
 - фильтрация по размеру изображения (меньшее измерение более 200 пикселей и соотношение сторон менее 3)
 - фильтрация по языку (с использованием fasttext и/или cld3).

- CLIP score
 - cosine similarity между предобученными эмбедами выше порога.
 - обучаются эмбединги также на ViT-B/32 или ViT-L/14
- LAION
 - CLIP score + cld3
- Text-based
 - Текстовые описания “похожи” на метки ImageNet
- Image-based
 - Изображения “похожи” на изображения ImageNet

Filtering baselines

Table 3: Zero-shot performance for select baselines in the *filtering* track. On all scales, filtering strategies lead to better performance than using the entire, unfiltered pool. The intersection between imaged-based and CLIP score strategies performs well on most tasks and scales. For all metrics, higher is better (see Appendix O for details). \cap denotes the intersection of filtering strategies.

Scale	Filtering strategy	Dataset size	Samples seen	ImageNet	ImageNet dist. shifts	VTAB	Retrieval	Average over 38 datasets
small	No filtering	12.8M	12.8M	0.025	0.033	0.145	0.114	0.132
	Basic filtering	3M	12.8M	0.038	0.043	0.150	0.118	0.142
	Text-based	3.2M	12.8M	0.046	0.052	0.169	<u>0.125</u>	0.157
	Image-based	3M	12.8M	0.043	0.047	0.178	0.121	0.159
	LAION-2B filtering	1.3M	12.8M	0.031	0.040	0.136	0.092	0.133
	CLIP score (L/14 30%)	3.8M	12.8M	<u>0.051</u>	<u>0.055</u>	<u>0.190</u>	0.119	0.173
	Image-based \cap CLIP score (L/14 30%)	1.4M	12.8M	0.039	0.045	0.162	0.094	0.144
medium	No filtering	128M	128M	0.176	0.152	0.259	0.219	0.258
	Basic filtering	30M	128M	0.226	0.193	0.284	0.251	0.285
	Text-based	31M	128M	0.255	0.215	0.328	0.249	0.307
	Image-based	29M	128M	0.268	0.213	0.319	<u>0.256</u>	0.312
	LAION-2B filtering	13M	128M	0.230	0.198	0.307	0.233	0.292
	CLIP score (L/14 30%)	38M	128M	0.273	0.230	0.338	0.251	<u>0.328</u>
	Image-based \cap CLIP score (L/14 30%)	14M	128M	<u>0.297</u>	<u>0.239</u>	<u>0.346</u>	0.231	<u>0.328</u>
large	No filtering	1.28B	1.28B	0.459	0.378	0.426	0.419	0.437
	Basic filtering	298M	1.28B	0.516	0.423	0.446	0.480	0.458
	Text-based	317M	1.28B	0.561	0.465	0.465	0.352	0.466
	Image-based	293M	1.28B	0.572	0.454	0.483	0.479	0.476
	LAION-2B filtering	130M	1.28B	0.553	0.453	0.510	0.495	0.501
	CLIP score (L/14 30%)	384M	1.28B	0.578	0.474	0.538	0.466	0.529
	Image-based \cap CLIP score (L/14 30%)	140M	1.28B	<u>0.631</u>	<u>0.508</u>	<u>0.546</u>	<u>0.498</u>	<u>0.537</u>
xlarge	No filtering	12.8B	12.8B	0.723	0.612	0.611	0.569	0.621
	LAION-2B filtering	1.3B	12.8B	0.755	0.637	0.624	<u>0.620</u>	0.636
	CLIP score (L/14 30%)	3.8B	12.8B	0.764	0.655	0.643	0.588	0.650
	Image-based \cap CLIP score (L/14 30%)	1.4B	12.8B	<u>0.792</u>	<u>0.679</u>	<u>0.652</u>	0.608	<u>0.663</u>

Спасибо за внимание!

Полезные ссылки:

статья: <https://arxiv.org/abs/2304.14108>

сайт: <https://www.datacomp.ai>

github: <https://github.com/mlfoundations/datacomp>

лидерборд: <https://www.datacomp.ai/leaderboard.html>