Race classification

Lucas Ceschini

August 2024

1 Prompts

raw_race_labels are gender neutral and have a single prompt for each race: "a photo of a {race} person".

2 Prompt Comparison

Table 1: race classification comparison between prompt strategies.

Model	Prompt	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
ViT-B-16 openai	RAGP	**0.6558**	0.6026	**0.5472**	0.5354	0.7124	**0.8580**	**0.7236**	**0.6617**	**0.1276**
	RGP	0.6428	**0.6574**	0.4494	**0.6149**	0.7350	0.8220	0.6583	0.6369	0.2040
	RP	0.5886	0.5935	0.2710	0.5749	**0.7760**	0.7738	0.7038	0.5459	0.3346
ViT-B-32 openai	RAGP	**0.6472**	0.6652	**0.5583**	0.5656	0.7314	**0.8740**	0.6814	0.4533	0.1937
	RGP	0.6402	**0.6806**	0.5424	**0.6303**	0.7088	0.8445	0.6840	0.3722	0.2653
	RP	0.6146	0.6265	0.4446	0.4794	**0.7555**	0.8368	**0.7051**	**0.5095**	**0.1779**
ViT-L-14 openai	RAGP	**0.6724**	0.6677	**0.5098**	0.5761	0.7336	**0.8535**	0.7454	0.6915	**0.1727**
	RGP	0.6616	**0.6929**	0.4585	**0.6106**	0.7258	0.8413	0.7342	0.6427	0.2138
	RP	0.5922	0.6787	0.1065	0.5114	**0.7519**	0.8451	**0.746**	**0.7221**	0.5166
ViT-H-14 laion2b	RAGP	**0.6969**	**0.7503**	0.7583	**0.4972**	0.6827	**0.8734**	**0.7381**	0.5285	**0.1926**
	RGP	0.6920	0.7413	**0.8096**	0.4861	**0.6933**	0.8715	0.7005	0.4591	0.2211
	RP	0.6816	0.7277	0.7894	0.4030	0.6919	0.8432	0.6662	**0.6096**	0.2729
ViT-g-14 laion2b	RAGP	**0.6861**	**0.6458**	0.7122	0.4288	0.7484	**0.8907**	**0.7579**	0.6112	0.2562
	RGP	0.6824	0.6310	**0.7276**	**0.4763**	0.7689	0.8676	0.7164	0.5649	**0.2027**
	RP	0.6613	0.6000	0.6523	0.4011	**0.7908**	0.8316	0.7526	**0.6195**	0.2629

3 Model Scaling

Scaling models with laion-2b and raw_race_labels.

Table 2: Model scaling for raw_race_labels using laion2B datasource.

Model	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
ViT-B-16	0.6496	0.7142	0.5755	**0.6217**	0.6155	0.8271	0.7124	0.4648	**0.1825**
ViT-B-32	0.5977	**0.7310**	0.7472	0.4295	0.4594	0.7192	0.6095	0.3854	0.1976
ViT-L-14	0.6794	**0.7310**	0.7209	0.4775	0.6869	0.7808	**0.7770**	0.5509	0.1975
ViT-H-14	**0.6816**	0.7277	**0.7894**	0.4030	**0.6919**	**0.8432**	0.6662	0.6096	0.2729
ViT-g-14	0.6613	0.6000	0.6523	0.4011	0.7908	0.8316	0.7526	**0.6195**	0.2629

.

.

4 Data Scaling

Scaling datasources for ViT-B-16, ViT-B-32 and ViT-L-14 using raw_race_labels.

4.1 ViT-B-16

Table 3: Data scaling for ViT-B-16 using raw_race_labels

datasource	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
datacomp_l_s1b_b8k	0.3392	0.2729	0.2197	0.1134	0.1845	0.7886	0.4624	0.3821	0.2328
openai	0.5886	0.5935	0.2710	0.5749	**0.7760**	0.7738	0.7038	**0.5459**	0.3346
laion400m_e32	0.6190	0.6903	**0.6561**	0.5052	0.5244	0.7500	0.6623	0.5045	**0.1088**
datacomp_xl_s13b_b90k	0.5079	0.2813	0.6326	0.3703	0.5088	0.8194	0.5712	0.2862	0.2144
commonpool_l_s1b_b8k	0.3712	0.2503	0.5281	0.0074	0.3951	0.5925	0.6939	0.0265	0.3489
laion2b_s34b_b88k	**0.6496**	**0.7142**	0.5755	**0.6217**	0.6155	**0.8271**	**0.7124**	0.4648	0.1825

4.2 ViT-B-32

Table 4: Data scaling for ViT-B-32 using raw_race_labels

datasource	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
datacomp_m_s128m_b4k	0.2025	0.2039	0.1458	0.0099	0.0049	0.5611	0.3496	0.1423	0.1976
commonpool_m_s128m_b4k	0.1728	0.1026	0.2120	0.0043	0.0028	0.5116	0.2012	0.1489	**0.1663**
openai	**0.6146**	0.6265	0.4446	**0.4794**	**0.7555**	**0.8368**	**0.7051**	**0.5095**	0.1779
laion400m_e32	0.5538	0.5742	0.5640	0.3623	0.4601	0.8226	0.6755	0.3780	0.1858
datacomp_xl_s13b_b90k	0.4235	0.4381	0.4441	0.1744	0.3279	0.7526	0.5237	0.2663	0.2438
laion2b_s34b_b79k	0.5977	**0.7310**	**0.7472**	0.4295	0.4594	0.7192	0.6095	0.3854	0.1976

4.3 ViT-L-14

Table 5: Data scaling for ViT-L-14 using raw_race_labels

datasource	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
openai	0.5922	0.6787	0.1065	**0.5114**	**0.7519**	**0.8451**	0.7460	**0.7221**	0.5166
laion400m_e32	0.6512	**0.7794**	0.6887	0.5034	0.5032	0.7988	0.6821	0.5649	**0.1426**
datacomp_xl_s13b_b90k	0.6193	0.7123	**0.7703**	0.3561	0.5555	0.7635	0.7071	0.3722	0.2492
laion2b_s32b_b82k	**0.6794**	0.7310	0.7209	0.4775	0.6869	0.7808	**0.7770**	0.5509	0.1975
commonpool_xl_s13b_b90k	0.5774	0.7626	0.6173	0.3099	0.4883	0.8066	0.6003	0.4111	0.2610

.
.
.
.
.
.
.
.
.
.
.
.
.

5 Top K Aggregation

Best aggregation technique for each model using ours age_race_gender labels. Datasources ordered by ascending size.

5.1 ViT-B-16

Table 6: Race classification with ViT-B-16 with best aggregation method for each data source.

datasource	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
datacomp_l_s1b_b8k	Avg Sum	0.3416	0.1174	0.3386	0.1824	0.0580	**0.8856**	0.4340	0.3639	0.2820
openai	Top 04	**0.6640**	0.6400	0.5204	**0.6248**	**0.7208**	0.8638	0.6992	**0.6278**	**0.1506**
laion400m_e32	Top 02	0.6353	0.5606	**0.7453**	0.5792	0.6417	0.8252	0.6095	0.3970	0.2256
datacomp_xl_s13b_b90k	Top 02	0.5029	0.2335	0.7233	0.2933	0.4820	0.8740	0.5646	0.2192	0.2651
commonpool_l_s1b_b8k	Top 07	0.3907	0.3877	0.4681	0.0290	0.4841	0.7744	0.4347	0.0885	0.3519
laion2b_s34b_b88k	Top 02	0.6587	**0.6516**	0.6873	0.5434	0.6686	0.8451	**0.7078**	0.4599	0.1921

5.2 ViT-B-32

Table 7: Race classification with ViT-B-32 with best aggregation method for each data source.

datasource	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
datacomp_m_s128m_b4k	Avg Sum	0.2287	0.0839	0.4945	0.0339	0.0198	0.6767	0.0303	0.1340	0.1906
commonpool_m_s128m_b4k	Top 02	0.1657	0.0245	0.2604	0.0025	0.0042	0.5071	0.1385	0.1861	**0.1580**
openai	Top 07	**0.6487**	0.6555	0.5827	**0.6562**	**0.7004**	0.8824	**0.6491**	0.3821	0.2620
laion400m_e32	Top 04	0.5733	**0.7245**	0.6734	0.2545	0.3406	**0.9158**	0.6187	0.4094	0.3079
datacomp_xl_s13b_b90k	Top 01	0.4342	0.3884	0.7151	0.0739	0.2438	0.8695	0.4426	0.1439	0.3371
laion2b_s34b_b79k	Top 05	0.6232	0.6400	**0.7751**	0.3635	0.5880	0.8940	0.5686	**0.4491**	0.2477

5.3 ViT-L-14

Table 8: Race classification with ViT-L-14 with best aggregation method for each data source.

datasource	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
openai	Top 04	0.6732	0.6910	0.4935	**0.6303**	**0.7138**	0.8586	**0.7414**	**0.6460**	0.1886
laion400m_e32	Avg Sum	0.6665	0.6910	0.7050	0.5059	0.5781	**0.8985**	0.7058	0.5401	0.1547
datacomp_xl_s13b_b90k	Top 06	0.6204	0.7019	**0.8456**	0.1830	0.6000	0.8715	0.6649	0.3598	**0.4208**
laion2b_s32b_b82k	Avg Sum	**0.6930**	**0.7755**	0.7041	0.5903	0.6431	0.8374	0.7289	0.5335	0.1540
commonpool_xl_s13b_b90k	Top 01	0.5881	0.5819	0.7170	0.3401	0.6332	0.8265	0.5765	0.3615	0.2366

.
.
.
.
.
.
.
.
.

5.4 ViT-H-14 and ViT-g-14

Table 9: Race classification with ViT-H-14 and ViT-g-14 with best aggregation method for laion2b datasource.

Model	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
ViT-H-14	Avg Sum	0.7011	0.7697	0.7679	0.5391	0.6721	0.8785	0.7157	0.5029	0.1894
ViT-g-14	Top 02	0.6906	0.6452	0.7103	0.4633	0.7583	0.8869	0.7454	0.6195	0.2265

6 Aggregation Techniques

In the official openAI notebook called "Prompt_Engineering_for_ImageNet.ipynb" they show how to create zero-shot classifier weights by embedding, normalizing and then **averaging** the tokenized prompts before measuring cosine distances. In our average sum approach, we perform the averaging of the similarity scores. Below we will compare both aggregation techniques on all models and datasources.

6.1 ViT-B-16

Table 10: Aggregation comparison for ViT-B-16

datasource	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
datacomp_l_s1b_b8k	Avg Sum	**0.3416**	**0.1174**	**0.3386**	0.1824	**0.0580**	**0.8856**	**0.4340**	**0.3639**	**0.2820**
	openAI Aggregation	0.3393	0.1006	0.3137	**0.3210**	0.0417	0.8740	0.4169	0.2771	0.2933
openai	Avg Sum	0.6626	**0.6561**	0.5233	**0.6371**	0.7166	0.8638	0.6900	0.5889	0.1447
	openAI Aggregation	**0.6717**	0.6555	**0.5722**	0.5687	**0.7244**	**0.8702**	**0.7051**	**0.6435**	**0.1084**
laion400m_e32	Avg Sum	**0.6242**	0.4690	**0.6916**	0.6038	**0.7053**	**0.8560**	**0.5719**	**0.4069**	**0.2080**
	openAI Aggregation	0.6232	**0.5181**	0.6782	**0.6981**	0.6466	0.8368	0.5554	0.3457	0.2656
datacomp_xl_s13b_b90k	Avg Sum	**0.4995**	0.2032	**0.7348**	0.2957	**0.5039**	**0.8869**	**0.5336**	**0.2010**	**0.2789**
	openAI Aggregation	0.4984	**0.2148**	0.6911	**0.5163**	0.4064	0.8663	0.4941	0.1447	0.3315
commonpool_l_s1b_b8k	Avg Sum	0.3894	**0.3432**	**0.4691**	0.0314	**0.5067**	**0.7841**	**0.4314**	**0.0935**	0.3485
	openAI Aggregation	**0.3950**	0.3323	0.4211	**0.2230**	0.4813	0.7757	0.3865	0.0811	**0.3048**
laion2b_s34b_b88k	Avg Sum	**0.6578**	0.6632	**0.6921**	0.5576	**0.6629**	**0.8528**	**0.6880**	**0.4318**	**0.2180**
	openAI Aggregation	0.6508	**0.6665**	0.6566	**0.6932**	0.6219	0.8316	0.6636	0.3490	0.2913

.
.
.
.
.
.
.
.
.
.
.
.

6.2 ViT-B-32

Table 11: Aggregation comparison for ViT-B-32

datasource	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
datacomp_m_s128m_b4k	Avg Sum	**0.2287**	**0.0839**	**0.4945**	0.0339	**0.0198**	**0.6767**	**0.0303**	**0.1340**	0.1906
	openAI Aggregation	0.1849	0.0052	0.1554	**0.8983**	0.0042	0.1414	0.0020	0.0050	**0.1711**
commonpool_m_s128m_b4k	Avg Sum	0.1599	**0.0252**	**0.2398**	0.0037	**0.0028**	**0.4505**	**0.1438**	**0.2349**	**0.1544**
	openAI Aggregation	**0.1676**	0.0052	0.1300	**0.4898**	**0.0028**	0.3143	0.0963	0.1017	0.1601
openai	Avg Sum	0.6460	0.6458	0.5856	**0.6679**	0.6982	0.8837	0.6326	0.3706	0.2700
	openAI Aggregation	**0.6560**	**0.6587**	**0.6321**	0.5792	**0.7053**	**0.8920**	**0.6537**	**0.4384**	**0.2129**
laion400m_e32	Avg Sum	**0.5722**	**0.7535**	**0.6609**	0.2323	**0.3413**	**0.9177**	**0.5937**	**0.4417**	0.3307
	openAI Aggregation	0.5669	0.7181	0.5842	**0.5625**	0.2975	0.8933	0.5389	0.2796	**0.2738**
datacomp_xl_s13b_b90k	Avg Sum	0.4277	**0.3310**	**0.7626**	0.0511	**0.2141**	**0.9004**	**0.4235**	**0.1266**	0.3502
	openAI Aggregation	**0.4372**	0.2935	0.7165	**0.3210**	0.1668	0.8811	0.3892	0.1009	**0.3090**
laion2b_s34b_b79k	Avg Sum	0.6200	**0.6381**	**0.7803**	0.3752	**0.5760**	**0.8991**	**0.5455**	**0.4351**	**0.2318**
	openAI Aggregation	**0.6230**	0.6303	0.7549	**0.5564**	0.5364	0.8772	0.5310	0.3648	0.2425

6.3 ViT-L-14

Table 12: Aggregation comparison for ViT-L-14

datasource	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
openai	Avg Sum	0.6682	**0.7123**	**0.4873**	0.6543	0.6926	**0.8663**	0.7322	0.5790	**0.1876**
	openAI Aggregation	**0.6689**	0.7019	0.4820	**0.6691**	**0.7046**	0.8535	**0.7361**	**0.5848**	0.1940
laion400m_e32	Avg Sum	**0.6665**	**0.6910**	**0.7050**	0.5059	**0.5781**	**0.8985**	**0.7058**	**0.5401**	**0.1547**
	openAI Aggregation	0.6614	0.6903	0.6954	**0.6205**	0.5152	0.8843	0.6893	0.4698	0.1823
datacomp_xl_s13b_b90k	Avg Sum	0.6202	**0.7058**	**0.8499**	0.1731	**0.5922**	**0.8683**	**0.6636**	**0.3739**	0.4307
	openAI Aggregation	**0.6305**	0.6948	0.8129	**0.4042**	0.5625	0.8631	0.6405	0.3052	**0.3067**
laion2b_s32b_b82k	Avg Sum	0.6930	0.7755	0.7041	0.5903	**0.6431**	**0.8374**	0.7289	**0.5335**	**0.1540**
	openAI Aggregation	**0.6934**	**0.7845**	**0.7132**	**0.5933**	0.6261	0.8355	**0.7335**	0.5219	0.1650
commonpool_xl_s13b_b90k	Avg Sum	0.5846	**0.5297**	**0.7472**	0.2286	**0.6912**	**0.8695**	0.5310	**0.4285**	0.3465
	openAI Aggregation	**0.5952**	**0.5297**	0.7281	**0.4085**	0.6615	0.8528	**0.5435**	0.3565	**0.2264**

6.4 ViT-H-14

Table 13: Aggregation comparison for ViT-H-14

datasource	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
laion2b_s32b_b79k	Avg Sum	0.7011	0.7697	**0.7679**	0.5391	**0.6721**	**0.8785**	**0.7157**	**0.5029**	**0.1894**
laion2b_s32b_b79k	openAI Aggregation	**0.7019**	**0.7774**	0.7530	**0.6291**	0.6530	0.8650	0.7032	0.4607	0.2309

.
.
.
.
.
.
.
.
.
.
.

6.5 ViT-g-14

Table 14: Aggregation comparison for ViT-g-14

datasource	Mode	accuracy	East Asian	White	Latino_Hispanic	Southeast Asian	Black	Indian	Middle Eastern	Race Gap
laion2b_s34b_b88k	Avg Sum	0.6890	0.6858	0.6719	0.4781	**0.7336**	**0.8875**	**0.7296**	**0.6468**	0.2124
laion2b_s34b_b88k	openAI Aggregation	**0.6904**	**0.6877**	**0.6868**	**0.5243**	0.7194	0.8850	0.7223	0.5988	**0.1649**