

Data	Data Composition							Question Analysis			Response Analysis			# of Rounds		
	General	OCR	Language	Counting	Code	Math	Science	Average	25%	75%	Average	25%	75%	Average	25%	75%
Cambrian-10M	24.1%	35.9%	17.2%	9.7%	8.7%	2.3%	2.1%	42.2	14.0	30.0	89.6	3.0	80.0	1.5	1.0	1.0
Cambrian-7M	33.3%	27.6%	23.8%	8.5%	0.8%	3.2%	2.9%	47.5	13.0	34.0	74.8	3.0	82.0	1.7	1.0	1.0

Table 1: **Data Analysis of Cambrian-10M and Cambrian-7M.** Here, we present data composition of Cambrian-10M and Cambrian-7M. We also use LLaMA-2 tokenizer to calculate number of tokens in questions and responses in the instruction tuning data.

Experiment	Backbone		Data		Adapter			Instruction Tuning			Compute		
	LLM	Vision	Adapter	Instruction Tuning	lr	wd	bs	lr	wd	bs	Device	Adapter Time	I.T. Time
Cambrian-1-8B	Llama-3-Ins-8B	SVA with 4 encoders*	2.5M	Cambrian-7M	1e-4	0	512	2e-5	0	512	-	TPU-V4-256	10.1h
Cambrian-1-13B	Vicuna-1.5-13B	SVA with 4 encoders*	2.5M	Cambrian-7M	1e-4	0	512	2e-5	0	512	-	TPU-V4-256	18.8h
Cambrian-1-34B	Hermes-2-Yi-34B	SVA with 4 encoders*	2.5M	Cambrian-7M	1e-4	0	512	2e-5	0	1024	-	TPU-V4-512	24.5h

Table 2: **Updated Implementation details and hyperparameters for all experiments.** *4 encoders are: OpenAI CLIP ViT-L/14@336, SigLIP ViT-SO400M/14@384, DINOv2 ViT-L/14@518, OpenCLIP ConvNeXt-XXL@1024

	Average	General	GQA	MME-P	MMB	SEED-I	Knowledge	SQA-I	MMMU	MathVisia	A12D	OCR & Chart	ChartQA	OCRBench	TextVQA	DocVQA	Vision-Centric	MMVP	V*Star	RealWorldQA	CV-Bench
All Model Average	2.79	2.50	2.13	1.43	2.40	4.13	0.17	-0.98	0.40	1.16	0.91	4.42	7.55	7.16	1.43	2.90	4.18	12.82	3.58	3.88	2.32
Language-Supervised Average	3.47	2.25	1.39	1.82	2.97	2.90	0.48	-0.91	0.05	0.92	2.16	9.67	19.32	13.99	1.92	8.59	3.05	6.21	0.63	4.74	1.56
OpenAI CLIP ViT-L/14@336	4.62	2.16	2.33	0.03	1.92	4.64	0.81	0.22	-0.29	-2.71	3.90	14.43	26.00	30.10	1.22	9.69	3.68	-9.28	4.25	7.14	6.50
SigLIP SO400M/14@384	2.89	2.77	1.38	2.30	4.54	2.91	-0.70	-2.29	-1.11	-0.34	1.22	8.42	17.18	9.68	2.36	7.04	1.60	3.83	2.18	4.60	-3.10
SigLIP ViT-L/16@384	3.28	2.77	1.41	0.34	7.02	2.84	0.89	-0.95	0.27	0.00	3.82	5.54	15.20	6.57	0.45	2.07	4.44	16.68	2.39	3.67	0.66
OpenCLIP ConvNeXt-L@512	7.51	3.76	2.10	3.56	4.62	4.78	-1.21	-1.09	-2.49	0.68	-1.53	26.10	51.28	26.43	6.01	36.54	6.87	12.22	5.96	8.30	3.49
EVA-CLIP-02 ViT-L/14@336	3.44	2.79	2.48	2.94	2.39	3.29	0.49	-1.85	-1.38	3.87	2.74	6.83	18.42	10.33	2.67	-0.18	4.98	28.84	-6.02	5.85	-0.20
DFN-CLIP ViT-L/14@224	-0.90	-0.74	-1.33	1.77	-2.70	-1.07	2.61	0.48	5.30	4.01	2.82	-3.33	-12.18	0.86	-1.17	-3.58	-3.26	-15.00	-4.99	-1.15	2.00
Self-Supervised Average	3.37	4.21	3.44	2.91	3.51	7.11	0.12	-1.30	0.38	1.99	0.70	-0.65	-3.79	-0.65	1.31	-3.54	7.97	28.63	10.82	5.13	2.84
DINOv2 ViT-L/14@336	4.88	6.22	3.69	5.93	4.11	11.11	0.29	-0.83	0.88	1.09	0.87	-2.30	-6.33	3.03	-0.55	-4.83	11.47	21.89	23.99	9.48	0.20
MAE ViT-L/16@224	3.25	4.29	2.32	4.30	4.91	5.84	0.03	-2.10	-1.66	5.04	1.13	-1.42	-2.76	-6.06	0.31	-4.55	8.12	16.69	16.42	4.11	3.66
MoCo v3 ViT-B/16@224	1.36	2.25	2.04	4.08	0.19	2.15	-0.66	-0.97	0.30	0.00	-1.20	-1.12	-6.07	6.67	0.75	-3.23	3.91	50.00	-1.31	-1.91	2.51
I-JEPA ViT-H/16@224	3.97	4.10	5.70	-2.66	4.83	9.34	0.81	-1.29	2.01	1.85	2.01	2.24	0.00	-6.25	4.72	-1.55	8.37	25.94	4.16	8.83	5.00
Other Average	-0.39	-0.19	1.71	-2.70	-1.52	1.84	-0.65	-0.53	1.49	0.19	-2.45	-1.17	-5.06	2.30	0.16	-1.28	0.01	1.00	-2.05	-1.20	3.53
MiDaS 3.0 ViT-L/16@384	-2.39	-3.42	0.95	-5.91	-7.73	-1.55	-1.20	-1.37	3.86	1.87	-5.49	-2.66	-6.56	-9.68	-0.37	-3.69	-2.25	-21.57	-1.43	0.85	0.30
SupViT ViT-L/16@224	1.60	3.05	2.47	0.52	4.69	5.23	-0.09	0.32	-0.87	-1.49	0.59	0.31	-3.57	14.29	0.69	1.12	2.26	23.57	-2.67	-3.26	6.77

Figure 1: Percentage(%) Change in Benchmark Performance (Frozen → Unfrozen)