A Large-Scale Examination of Inductive Biases Shaping High-Level Visual Representation in Brains and Machines

Supplementary Information

SI.1 Surveyed Models

The list of all surveyed models may be found in Table 1 below.

Table 1: List of All Models Tested

	Model ID	Architecture	Train Task	Train Data	cRSA	eRSA
1	CLiP-ResNet50	ResNet50	clip	openai400M	0.346	0.717
2	CLiP-ResNet101	ResNet101	clip	openai400M	0.353	0.715
3	RegNet-64Gf-SEER	RegNet-64Gf	seer	random1B	0.351	0.712
4	RegNet-128Gf-SEER	RegNet-128Gf	seer	random1B	0.331	0.711
5	CLiP-ViT-B/32	ViT-B/32	clip	openai400M	0.349	0.709
6	CLiP-ViT-B/32	ViT-B/16	clip	openai400M	0.350	0.708
7	ViT-B-SLIP	ViT-B	slip	YFCC15M	0.335	0.706
8	ConvNext-Large-IN21K	convnext_large	classification	imagenet21k	0.353	0.706
9	ResMLP-Big-24-IN21K	resmlp_big_24_224	classification	imagenet21k	0.395	0.705
10	RegNet-32Gf-SEER	RegNet-32Gf	seer	random1B	0.352	0.705
11	CLiP-ViT-L/14	ViT-L/14	clip	openai400M	0.314	0.704
12	ConvNext-Base-IN21K	convnext_base	classification	imagenet21k	0.330	0.702
13	ViT-L-SLIP	ViT-L	slip	YFCC15M	0.315	0.701
14	BiT-Expert-Abstraction	ResNet50-V2	bit_expert	big_transfer	0.350	0.701
15	Swin-B-P4-W7-IN21K	swin_base_patch4_window7_224	classification	imagenet21k	0.323	0.700
16	Swin-L-P4-W7-IN21K	swin_large_patch4_window7_224	classification	imagenet21k	0.324	0.700
17	BiT-Expert-Mammal	ResNet50-V2	bit_expert	big_transfer	0.350	0.700
18	ResNet50-SimCLR	ResNet50	selfsupervised	imagenet	0.362	0.699
19	BiT-Expert-Object	ResNet50-V2	bit_expert	big_transfer	0.348	0.699
20	RegNet-32Gf-SEER-INFT	RegNet-32Gf	seer	random1B	0.362	0.698
21	RegNet-64Gf-SEER-INFT	RegNet-64Gf	seer	random1B	0.333	0.697
22	BiT-Expert-Animal	ResNet50-V2	bit_expert	big_transfer	0.357	0.697
23	XCIT-N-12-P8	xcit_nano_12_p8_224	classification	imagenet	0.364	0.697
24	ResMLP-Big-24	resmlp_big_24_224	classification	imagenet	0.362	0.697
25	ViT-L-SLIP-CC12M	ViT-L	slip	YFCC15M	0.287	0.696
26	BiT-Expert-Bird	ResNet50-V2	bit_expert	big_transfer	0.349	0.696
27	YOLO-V5-S	yolov5s	yolo	coco,voc	0.356	0.695
28	YOLO-V5-M	yolov5m	yolo	coco,voc	0.370	0.695
29	YOLO-V5-L	yolov5l	yolo	coco,voc	0.350	0.694
30	BiT-Expert-Relation	ResNet50-V2	bit_expert	big_transfer	0.356	0.693
31	ResMLP-36	resmlp_36_224	classification	imagenet	0.346	0.692
32	NF-ResNet50	nf_resnet50	classification	imagenet	0.401	0.691
33	ViT-B-SimCLR	ViT-B	slip	YFCC15M	0.341	0.691
34	EfficientNet-B1	efficientnet_b1	classification	imagenet	0.422	0.691
35	BiT-Expert-Arthropod	ResNet50-V2	bit_expert	big_transfer	0.361	0.691
36	BiT-Expert-Flower	ResNet50-V2	bit_expert	big_transfer	0.334	0.691
37	BiT-Expert-Instrument	ResNet50-V2	bit_expert	big_transfer	0.353	0.691
38	BiT-Expert-Vehicle	ResNet50-V2	bit_expert	big_transfer	0.350	0.691
39	GMLP-S16	gmlp_s16_224	classification	imagenet	0.402	0.690
40	ViT-L-SimCLR	ViT-L	slip	YFCC15M	0.327	0.690
41	BiT-Expert-Food	ResNet50-V2	bit_expert	big_transfer	0.356	0.690
42	ResNet50-DeepClusterV2	ResNet50	selfsupervised	imagenet	0.330	0.690
43	RegNet-128Gf-SEER-INFT	RegNet-128Gf	seer	random1B	0.342	0.689
44	HardCoreNAS-F	hardcorenas_f	classification	imagenet	0.435	0.689
45	ViT-S-SLIP	ViT-S	slip	YFCC15M	0.331	0.689
46	ViT-S-SimCLR	ViT-S	slip	YFCC15M	0.339	0.688
47	LeViT128	levit_128	classification	imagenet	0.425	0.688
48	Dino-ResNet50	resnet50	selfsupervised	imagenet	0.358	0.687
49	ResMLP-24	resmlp_24_224	classification	imagenet	0.368	0.686
50	Swin-L-P4-W7	swin_large_patch4_window7_224	classification	imagenet	0.283	0.685
51	PoolFormer-S36	poolformer_s36	classification	imagenet	0.364	0.685
52	HardCoreNAS-A	hardcorenas_a	classification	imagenet	0.422	0.685
53	XCIT-N-12-P16	xcit_nano_12_p16_224	classification	imagenet	0.392	0.685
54	Dino-VIT-B16	vitb16	selfsupervised	imagenet	0.331	0.684
55	GMixer-24	gmixer_24_224	classification	imagenet	0.344	0.684
56	ResNet50-SwAV-BS4096	ResNet50	selfsupervised	imagenet	0.337	0.684
57	ECA-NFNeT-L0	eca_nfnet_l0	classification	imagenet	0.401	0.683
58	NF-Net-L0	nfnet_l0	classification	imagenet	0.417	0.683
59	EfficientNet-B3	efficientnet_b3	classification	imagenet	0.431	0.681
60	ResNet50	resnet50	classification	imagenet	0.379	0.680
61	Mixer-L16-IN22K	mixer_l16_224	classification	imagenet21k	0.234	0.680
62	SemNASNet100	semnasnet_100	classification	imagenet	0.405	0.679

Table 1 List of All Models Tested (Continued)

	Model ID	Architecture	Train Task	Train Data	cRSA	eRSA
63	ResNet50-BarlowTwins-BS2048	ResNet50	selfsupervised	imagenet	0.367	0.679
64	Swin-B-P4-W7	swin_base_patch4_window7_224	classification	imagenet	0.296	0.679
65	Swin-T-P4-W7	swin_tiny_patch4_window7_224	classification	imagenet	0.315	0.678
66	ViT-B-P16-IN21K	vit_base_patch16_224	classification	imagenet21k	0.312	0.677
67	CSP-ResNet50	cspresnet50	classification	imagenet	0.400	0.677
68	MobileNet-V3-Large	mobilenetv3_large_100	classification	imagenet	0.412	0.676
69	MobileNet-V2	mobilenet_v2	classification	imagenet	0.358	0.676
70	Faster-RCNN-ResNet50-FPN	faster_rcnn_R_50_FPN_3x	detection	coco2017	0.370	0.676
71	GhostNet100	ghostnet_100	classification	imagenet	0.399	0.676
72	ConvNext-B	convnext_base	classification	imagenet	0.298	0.676
73	ConvMixer-768-32	convmixer_768_32	classification	imagenet	0.342	0.675
74	SEResNext50-32x4D	seresnext50_32x4d	classification	imagenet	0.384	0.675
75	ResMLP-12	resmlp_12_224	classification	imagenet	0.365	0.675
76	Inception-V3	inception_v3	classification	imagenet	0.417	0.675
77	RegNetX-64	regnetx_064	classification	imagenet	0.303	0.674
78	MNASNet1.0	mnasnet1_0	classification	imagenet	0.369	0.673
79	CoaT-Lite-Tiny	coat_lite_tiny	classification	imagenet	0.323	0.673
80	ResNet50-PIRL	ResNet50	selfsupervised	imagenet	0.355	0.672
81	Mask-RCNN-ResNet50-FPN	mask_rcnn_R_50_FPN_3x	segmentation	coco2017	0.363	0.671
82	GoogleNet	googlenet	classification	imagenet	0.377	0.671
83	ViT-S-CLIP	ViT-S	slip	YFCC15M	0.293	0.670
84	RegNetY-64	regnety_064	classification	imagenet	0.372	0.670
85	ViT-B-CLIP	ViT-B	slip	YFCC15M	0.294	0.668
86	DenseNet121	densenet121	classification	imagenet	0.369	0.668
87	JX-NesT-Tiny	jx_nest_tiny	classification	imagenet	0.319	0.668
88	MLP-Mixer-L16	mixer_l16_224	classification	imagenet	0.351	0.668
89	VGG16	vgg16	classification	imagenet	0.342	0.668
90	SKResNext50-32x4D	skresnext50_32x4d	classification	imagenet	0.355	0.666
91	RetinaNet-ResNet50-FPN	retinanet_R_50_FPN_3x	detection	coco2017	0.357	0.665
92	DLA34	dla34	classification	imagenet	0.359	0.665
93	ViT-B-R50-S16-IN21K	vit_base_r50_s16_224	classification	imagenet21k	0.334	0.664
94	ResNet101	resnet101	classification	imagenet	0.322	0.663
95	ViT-L-P16-IN21K	vit_large_patch16_224	classification	imagenet21k	0.306	0.663
96	MLP-Mixer-B16	mixer_b16_224	classification	imagenet	0.349	0.662
97	PiT-T-224	pit_ti_224	classification	imagenet	0.394	0.662
98	ResNet152	resnet152	classification	imagenet	0.349	0.662
99	XCeption	xception	classification	imagenet	0.332	0.661
100	Visformer	visformer_small	classification	imagenet	0.334	0.661
101	ViT-B-P32	vit_base_patch32_224	classification	imagenet	0.385	0.661
102	ShuffleNet-V2-x1.0	shufflenet_v2_x1_0	classification	imagenet	0.357	0.660
103	TnT-P16-224	tnt_s_patch16_224	classification	imagenet	0.281	0.657
104	ViT-B-P16	vit_base_patch16_224	classification	imagenet	0.319	0.657
105	ViT-L-CLIP	ViT-L	slip	YFCC15M	0.284	0.655
106	CrossViT-B	crossvit_base_240	classification	imagenet	0.315	0.655
107	ViT-S-P32-IN21K	vit_small_patch32_224	classification	imagenet21k	0.287	0.654
108	ConvNext-L	convnext_large	classification	imagenet	0.243	0.654
109	ViT-B-P32-IN21K	vit_base_patch32_224	classification	imagenet21k	0.364	0.652
110	ResNet18	resnet18	classification	imagenet	0.325	0.648
111	ViT-L-P16	vit_large_patch16_224	classification	imagenet	0.320	0.648
112	ViT-S-P16	vit_small_patch16_224	classification	imagenet	0.311	0.647
113	ViT-S-P16-IN21K	vit_small_patch16_224	classification	imagenet21k	0.319	0.645
114	Mixer-B16-IN22K	mixer_b16_224	classification	imagenet21k	0.269	0.641
115	ViT-L-CLIP-CC12M	ViT-L	slip	YFCC15M	0.283	0.640
116	DeiT-B-P16-224	deit_base_patch16_224	classification	imagenet	0.309	0.640
117	ConViT-B	convit_base	classification	imagenet	0.345	0.639
118	ViT-S-P32	vit_small_patch32_224	classification	imagenet	0.289	0.638
119	PiT-B-224	pit_b_224	classification	imagenet	0.272	0.638
120	AlexNet	alexnet	classification	imagenet	0.338	0.633
121	ResNet50-ClusterFit-16K-RotNet	ResNet50	selfsupervised	imagenet	0.326	0.630
122	ConViT-T	convit_tiny	classification	imagenet	0.286	0.628
123	Keypoint-RCNN-ResNet50-FPN	keypoint_rcnn_R_50_FPN_3x	segmentation	coco2017	0.349	0.627
124	SqueezeNet1.0	squeezenet1_0	classification	imagenet	0.323	0.623
125	DPT-Hybrid	DPT_Hybrid	monoculardepth	MegaDepth+	0.265	0.620
126	MiDaS	MiDaS	monoculardepth	MegaDepth+	0.200	0.619
127	ResNet50-MoCoV2-BS256	ResNet50	selfsupervised	imagenet	0.266	0.610
128	ViT-T-P16	vit_tiny_patch16_224	classification	imagenet	0.293	0.590
129	ResNet50-JigSaw-P100	ResNet50	selfsupervised	imagenet	0.267	0.577
130	ResNet50-RotNet	ResNet50	selfsupervised	imagenet	0.238	0.572
131	ResNet50-JigSaw-Goyal19	ResNet50	selfsupervised	imagenet	0.257	0.565
132	AlexNet-GN-IPCLImageNet	alexnet_gn	ipcl	imagenet	0.272	0.562
133	AlexNet-GN-IPCLOpenImages	alexnet_gn	ipcl	openimages	0.249	0.517
134	AlexNet-GN-IPCLPlaces256	alexnet_gn	ipcl	places256	0.242	0.495
135	Object Classification	resnet50	taskonomy	taskonomy	0.189	0.436
136	Scene Classification	resnet50	taskonomy	taskonomy	0.147	0.426
137	3D Keypoints	resnet50	taskonomy	taskonomy	0.165	0.362
138	Euclidean Depth	resnet50	taskonomy	taskonomy	0.173	0.361
139	Reshading	resnet50	taskonomy	taskonomy	0.161	0.358
140	Occlusion Edges	resnet50	taskonomy	taskonomy	0.162	0.348
141	Surface Normals	resnet50	taskonomy	taskonomy	0.154	0.339

Table 1 List of All Models Tested (Continued)

	Model ID	Architecture	Train Task	Train Data	cRSA	eRSA
142	Z-Buffer Depth	resnet50	taskonomy	taskonomy	0.132	0.331
143	Semantic Segmentation	resnet50	taskonomy	taskonomy	0.144	0.324
144	Curvatures	resnet50	taskonomy	taskonomy	0.132	0.320
145	Unsupervised 2.5D Segmentation	resnet50	taskonomy	taskonomy	0.171	0.313
146	Room Layout	resnet50	taskonomy	taskonomy	0.137	0.309
147	AlexNet-GN-IPCLVGGFace2	alexnet_gn	ipcl	vggface2	0.101	0.290
148	NF-ResNet50	nf_resnet50	random_weights		0.083	0.280
149	Camera Pose (Fixated)	resnet50	taskonomy	taskonomy	0.099	0.236
150	Point Matching	resnet50	taskonomy	taskonomy	0.093	0.216
151	Egomotion	resnet50	taskonomy	taskonomy	0.087	0.197
152	Inpainting	resnet50	taskonomy	taskonomy	0.095	0.195
153	Vanishing Point	resnet50	taskonomy	taskonomy	0.088	0.184
154	JX-NesT-Tiny	jx_nest_tiny	random_weights		0.053	0.175
155	Unsupervised 2D Segmentation	resnet50	taskonomy	taskonomy	0.081	0.170
156	Texture Edges	resnet50	taskonomy	taskonomy	0.063	0.166
157	GoogleNet	googlenet	random_weights		0.056	0.161
158	Jigsaw	resnet50	taskonomy	taskonomy	0.062	0.155
159	Camera Pose (Nonfixated)	resnet50	taskonomy	taskonomy	0.095	0.153
160	XCeption	xception	random_weights		0.063	0.150
161	Denoising	resnet50	taskonomy	taskonomy	0.061	0.142
162	HardCoreNAS-A	hardcorenas_a	random_weights		0.028	0.140
163	XCIT-N-12-P8	xcit_nano_12_p8_224	random_weights		0.038	0.138
164	SKResNext50-32x4D	skresnext50_32x4d	random_weights		0.044	0.135
165	AlexNet	alexnet	random_weights		0.039	0.133
166	DenseNet121	densenet121	random_weights		0.057	0.132
167	CSP-ResNet50	cspresnet50	random_weights		0.041	0.131
168	Random Weights	resnet50	taskonomy		0.037	0.125
169	Visformer	visformer_small	random_weights		0.053	0.123
170	ViT-L-P16	vit_large_patch16_224	random_weights		0.052	0.123
171	VGG16	vgg16	random_weights		0.040	0.122
172	SqueezeNet1.0	squeezenet1_0	random_weights		0.071	0.121
173	DLA34	dla34	random_weights		0.018	0.118
174	2D Keypoints	resnet50	taskonomy	taskonomy	0.060	0.118
175	ConViT-B	convit_base	random_weights		0.053	0.118
176	ViT-B-P32	vit_base_patch32_224	random_weights		0.052	0.116
177	MNASNet1.0	mnasnet1_0	random_weights		0.065	0.115
178	Swin-L-P4-W7	swin_large_patch4_window7_224	random_weights		0.064	0.115
179	RegNetY-64	regnety_064	random_weights		0.052	0.115
180	SEResNext50-32x4D	seresnext50_32x4d	random_weights		0.049	0.115
181	HardCoreNAS-F	hardcorenas_f	random_weights		0.024	0.114
182	GhostNet100	ghostnet_100	random_weights		0.037	0.114
183	ViT-B-P16	vit_base_patch16_224	random_weights		0.051	0.113
184	MLP-Mixer-L16	mixer_l16_224	random_weights		0.072	0.113
185	ResNet18	resnet18	random_weights		0.044	0.113
186	DeiT-B-P16-224	deit_base_patch16_224	random_weights		0.055	0.112
187	GMLP-S16	gmlp_s16_224	random_weights		0.065	0.111
188	Swin-B-P4-W7	swin_base_patch4_window7_224	random_weights		0.071	0.111
189	Swin-T-P4-W7	swin_tiny_patch4_window7_224	random_weights		0.074	0.110
190	Inception-V3	inception_v3	random_weights		0.016	0.109
191	ECA-NFNeT-L0	eca_nfnet_l0	random_weights		0.067	0.109
192	ConvNext-L	convnext_large	random_weights		0.074	0.107
193	XCIT-N-12-P16	xcit_nano_12_p16_224	random_weights		0.026	0.106
194	ViT-S-P16	vit_small_patch16_224	random_weights		0.069	0.106
195	ViT-S-P32	vit_small_patch32_224	random_weights		0.068	0.106
196	ResMLP-12	resmlp_12_224	random_weights		0.055	0.103
197	Autoencoder	resnet50	taskonomy	taskonomy	0.077	0.103
198	MLP-Mixer-B16	mixer_b16_224	random_weights		0.069	0.103
199	EfficientNet-B3	efficientnet_b3	random_weights		0.078	0.102
200	ConvNext-B	convnext_base	random_weights		0.072	0.102
201	ResMLP-36	resmlp_36_224	random_weights		0.059	0.101
202	ResMLP-24	resmlp_24_224	random_weights		0.059	0.101
203	CrossViT-B	crossvit_base_240	random_weights		0.053	0.100
204	MobileNet-V2	mobilenet_v2	random_weights		0.043	0.100
205	ResMLP-Big-24	resmlp_big_24_224	random_weights		0.058	0.098
206	PiT-B-224	pit_b_224	random_weights		0.068	0.098
207	ViT-T-P16	vit_tiny_patch16_224	random_weights		0.061	0.097
208	CoaT-Lite-Tiny	coat_lite_tiny	random_weights		0.067	0.096
209	ConViT-T	convit_tiny	random_weights		0.069	0.096
210	LeViT128	levit_128	random_weights		0.032	0.095
211	PoolFormer-S36	poolformer_s36	random_weights		0.057	0.094
212	NF-Net-L0	nfnet_l0	random_weights		0.080	0.094
213	PiT-T-224	pit_ti_224	random_weights		0.046	0.093
214	ResNet50	resnet50	random_weights		0.039	0.093
215	SemNASNet100	semnasnet_100	random_weights		0.032	0.090
216	TnT-P16-224	tnt_s_patch16_224	random_weights		0.074	0.089
217	ResNet101	resnet101	random_weights		0.033	0.086
218	EfficientNet-B1	efficientnet_b1	random_weights		0.049	0.085
219	ConvMixer-768-32	convmixer_768_32	random_weights		0.017	0.082
220	GMixer-24	gmixer_24_224	random_weights		0.073	0.078

Table 1 List of All Models Tested (Continued)

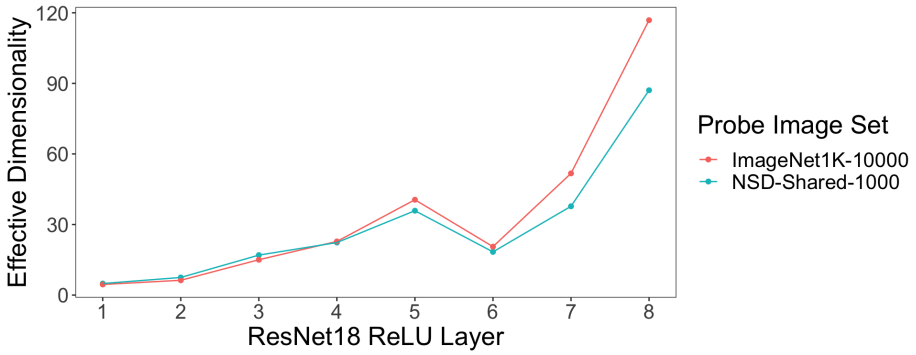
	Model ID	Architecture	Train Task	Train Data	cRSA	eRSA
221	ResNet152	resnet152	random_weights		0.035	0.075
222	RegNetX-64	regnetx_064	random_weights		0.039	0.075
223	MobileNet-V3-Large	mobilenetv3_large_100	random_weights		0.009	0.057
224	ShuffleNet-V2-x1.0	shufflenet_v2_x1_0	random_weights		0.041	0.048

SI.2 Effective Dimensionality

Concurrent work by Elmoznino and Bonner [66] found that model feature spaces with greater effective dimensionality (ED) were better predictors of high-level visual cortex, including in prediction of occipitotemporal responses in the Natural Scenes Dataset. In our analysis, we do not find this trend. What reasons might account for this disjunct between findings?

First, we considered model sets. At the time of this preprint, the models tested by Elmoznino and Bonner [66] include only ResNet18 and ResNet50 models, trained via self- or category-supervision on ImageNet1K; the ResNet50 models of Taskonomy; and untrained ResNet18 and ResNet50 models. When we subsetted our models to include only the best layers of the Taskonomy Resnet50s (N = 24), we found that ED was indeed a significant positive predictor of OTC predictivity ($r_{Spearman}$ = 0.41 [0.015, 0.73], p = 0.026 in cRSA; 0.435 [0.162, 0.730], p = 0.039 in veRSA). However, as is perhaps already evident from the large confidence intervals, we found this effect to be somewhat brittle, quickly breaking with the addition of the 12 less-performant models that ranked similarly to those of Taskonomy (and below the breakpoint suggested by the segmented regression analysis; $r_{Spearman}$ = 0.041 [-0.357, 0.392], p = 0.341 in cRSA; 0.0532 [-0.324, 0.393], p = 0.37 in veRSA).

Second, we tested the impact of the image set on computation of ED. Elmoznino and Bonner [66] estimated ED over 10,000 ImageNet1K validation images, whereas we estimated ED over the 'shared1000' COCO images. To check for differences in ED estimates computed over the different image sets, we directly replicated the approach of Elmoznino and Bonner [66]. This involved considering only the ReLU stages of each convolutional block, and performing a global-pooling operation over features of each model layer, prior to computing ED. While we did not include this pooling step in our main analyses, we did so in this supplementary analysis to ensure that ED levels across the two image sets could be compared with all other analytical choices held constant. The results of this analysis can be seen in Supplementary Figure 1. We observe similar ED estimates when probing these two different image sets, with slightly lower ED estimates in the later layers when using the 'shared1000' probe set.



Supplementary Figure 1: **Replication of Elmoznino and Bonner [66]'s Effective Dimensionality (ED)**. We use the ReLU outputs of ResNet18's residual blocks to determine whether the 'shared1000' COCO images were a sufficiently diverse image set for calculating a meaningful measure of ED, compared with 10,000 ImageNet1K validation used in previous analyses. While the estimated ED yielded by the 10,000 ImageNet1K images is visibly higher in later layers, the correlation between the ED across layers (which undergirds the primary statistics we use to assess the relationship of ED to brain predictivity in our main analysis) is effectively 1:1 ($r_{Pearson}$ = 0.99 [0.98, 0.99]). This shows that the 'shared1000' COCO images are a sufficiently diverse image set for calculating ED.

Broadly, there are several key analytical differences in our test of ED and brain predictivity that likely underlie the divergence between our findings and those of Elmoznino and Bonner [66]. First, Elmoznino and

Bonner [66] jointly analyzed ED from multiple layers of individual models, as well as from untrained models. We argue that this choice introduces significant covariation between high-level feature quality and effective dimensionality. If the importance of ED for high brain predictivity was truly a general principle, then our stronger test examining the ED variation across models from only the most brain-aligned layer should also yield correlations with brain predictivity (but this was not the case). Second, Elmoznino and Bonner [66] applied a global max-pooling to the convolutional feature maps before computing ED, noting their primary interest in the variance of image features, rather than the variance in those properties across space. However, in our analysis pipeline, such a global max-pooling operation is frequently infeasible (e.g, for non-convolutional models). And, more generally, estimating ED over the exact same feature space used to fit the brain responses seems preferable from a theoretical standpoint.

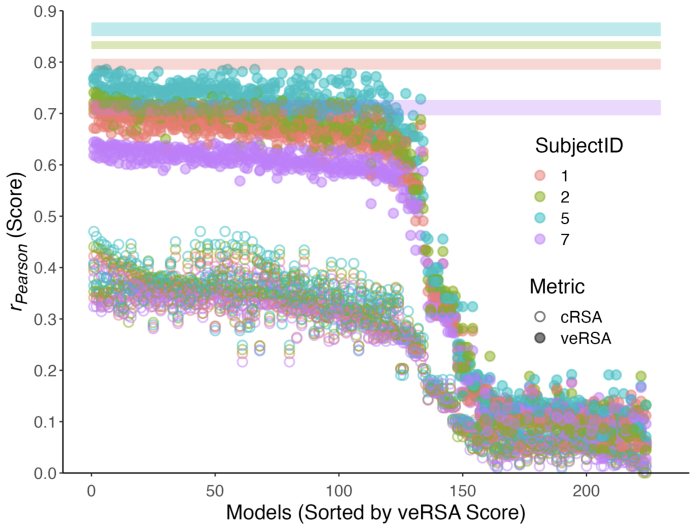
SI.3 Stability of Rankings across Subjects

In our main analysis, all predictivity scores we report are the average across 4 subjects. That even the smallest differences in predictivity between models is significant across many of our controlled model comparisons already suggests that patterns in these scores are meaningfully stable and consistent across subjects. We further quantity this stability in two ways.

The first is with a rank-order correlation of model rankings between individual pairs of subjects (N = 6 unique pairwise comparisons). Across *all models* – including randomly-initialized models – this correlation is $r_{Spearman}$ = 0.98 [0.975, 0.985] for cRSA and 0.957 [0.945, 0.967]. Across only the top 125-ranking models, this correlation remains remarkably high, at $r_{Spearman}$ = 0.923 [0.907, 0.938] for cRSA and $r_{Spearman}$ = 0.835 [0.797, 0.873] for eRSA. A permutation test in which we scramble the model scores across subjects and recompute this same rank order correlation suggests these values are *extremely* unlikely to occur by chance (mean permuted $r_{Spearman}$ = 0.0052 [-0.0455, 0.0736] in cRSA and 0.008 [-0.0239, 0.0278] in eRSA; not a single permutation achieves $r_{Spearman} > 0.1$).

What we are testing here, in effect, is the logic of brain-predictivity leaderboards. This analysis suggests that – while small – the differences between consecutively ranked models in our analyses are statistically meaningful. This does not spare inferences derived from leaderboards from the critique that differences between consecutively ranked models are rarely attributable to unique or controlled sources of variation, but it does suggest that small score differences reflect more than random statistical variation.

We visualize the degree of inter-subject variability in model rankings in Supplementary Figure 2 below.

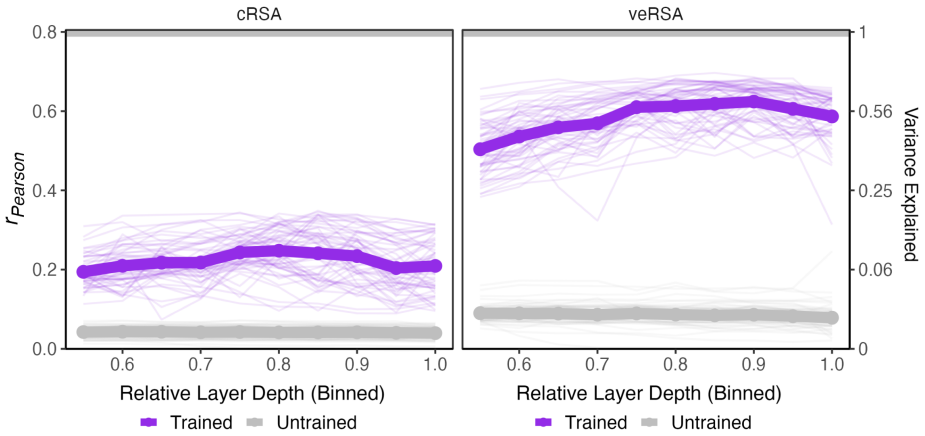


Supplementary Figure 2: **Overall Model Variation for Individual Subjects**. Brain predictivity is plotted for all models (N = 224), sorted by the group-average veRSA score. Each point is the score from the most brain-predictive layer (selected by cross-validation) of a single model, plotted for both cRSA (open) and veRSA (filled) metrics. Colors refer to the 4 NSD subjects. Shaded regions refer to the subject-specific noise ceilings.

SI.4 Brain Predictivity as a Function of Layer Depth

In our main analyses, we compare models based on only their most brain-predictive layers, identified via cross-validation. We also examined whether these peak layers tended to arise at a particular level of computational depth. In Supplementary Figure 3 below, we summarize prediction scores as a function of relative layer depth for all models in the controlled architecture comparison (CNNs versus Transformers, with constant task and input diet; see Results Figure 2). The x-axis ranges from 0.5 to 1.0 relative layer depth (where 0 is the input layer, and 1 is the output layer), reflecting our focus on testing the correspondence of the later model layers to OTC for computational efficiency.

Supplementary Figure 3 reveals that the later layers of trained models show gradually increasing brain predictivity with layer depth, peaking at a relative depth of around 0.9 on average, followed by slight decrease in mean predictivity toward the output layers. This pattern is more pronounced in veRSA compared to cRSA. No such trend is evident in untrained models, regardless of the metric used. These comparisons further demonstrate the gap in predictive capacity of trained versus untrained models across the late-stage model hierarchy.



Supplementary Figure 3: **Layer Summaries of Brain Prediction Scores**. OTC prediction scores are plotted against relative layer depth for the 64 DNN architectures included in our survey of ImageNet-trained models (including the architectures in Figure 2), for cRSA (left) and veRSA (right). Scores for each distinct model are represented by thin lines, with the average trend shown by a thick line (purple for trained models and gray for untrained). The x-axis reflects relative layer positions within each network, where 0.5 reflects the middle layer and 1 reflects the final layer. Data points are binned by 0.1 intervals of relative depth, with OTC-prediction scores averaged within each bin.

SI.5 Modeling Results in Early Visual Cortex

While the focus of our main analyses was the predictivity of a unified OTC ROI, we designed our pipeline to generate predictions for a number of additional ROIs – including early visual cortex (EVC: V1-V4).

This ROI encapsulates the ventral and dorsal aspects of areas V1, V2, and V3, as well as area hV4. To define the EVC ROI for each subject, we again first isolated voxels within the "nsdgeneral" ROI, and then selected for analyses any voxels that both fell within one of the early visual regions listed above, and that exceeded the NCSNR threshold of 0.2. This procedure yielded a total of 4,657 voxels for subject 01, 3,757 voxels for subject 02, 3,661 voxels for subject 05, and 3,251 voxels for subject 07.

A first question we might ask, then, is whether the better models of OTC in general are better models of EVC. To do this, we can correlate the predictivity of each model we test in OTC with that same model's predictivity in a macro-scale EVC ROI. (Note that, as in our analyses of OTC, we select the most EVC-predictive layer from each model using the training set of 500 images, and report the score of this layer on the 500 held-out test images). Across *all models*, this correlation is high: $r_{Spearman}$ = 0.809 [0.801, 0.822] in cRSA and 0.835 [0.816, 0.853] in eRSA. Across only the 125 highest ranking models of OTC, this correlation is markedly lower: 0.212 [0.166, 0.261] in cRSA and 0.287 [0.209, 0.364] in eRSA. In other words, it seems, poor models of OTC (e.g. the Taskonomy-trained and randomly-initialized) models are also poor models of

EVC; excluding these poor-performing models, however, better models of OTC are not necessarily better models of EVC.

This rank-order correlation across many models, more generally, does not necessarily capture the subtleties and trends we saw in our opportunistic experiments and controlled model comparisons, which we can directly repeat in EVC. While a comprehensive recap of each opportunistic experiment applied to EVC is beyond the scope of this analysis, what we can say is that many, but not all, of the trends we observe in OTC are recapitulated in EVC. For example, there is, once again, only a negligible difference in the average predictivity of CNNs versus transformers (cRSA β = -0.029, p = 7.08e-10; veRSA β = -0.007, p = 2.61e- 3). Perhaps the most notable divergence between EVC and OTC in terms of our opportunistic experiments is the minimal difference of the veRSA metric in EVC between the self-supervised (IPCL) model trained on objects (ImageNet) versus faces (VGGFace2) (EVC β = -0.0394, p = 8.62e- 4, OTC β = -0.2719, p = 1.03e-9). Unlike in OTC, then, face-trained models (with reweighting) perform on par with object-trained models. This means that even the depleted set of natural image statistics available in a visual diet of faces alone may still be sufficient to capture core aspects of early visual cortical representation.

SI.6 Compute Required

We used a single machine with 8 Nvidia RTX 3090 GPUs, 755gb of RAM, and 96 CPUs. GPUs were used only for extracting model activations, and could (without major slowdown) be removed from the analytic pipeline. Dimensionality reduction and regression computations were CPU and RAM intensive. Replicating all of our results would take approximately three weeks on a similar machine.